The New Encyclopædia Britannica

Volume 29

MACROPÆDIA

Knowledge in Depth

FOUNDED 1768
15 TH EDITION

Encyclopædia Britannica, Inc.
Robert P. Gwinn, Chairman, Board of Directors
Peter B. Norton, President
Philip W. Goetz, Editor in Chief

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

THE UNIVERSITY OF CHICAGO

"Let knowledge grow from more to more
and thus be human life enriched."

The *Encyclopædia Britannica* is published with the editorial advice of the faculties of the University of Chicago.

Additional advice is given by committees of members drawn from the faculties of the Australian National University, the universities of British Columbia (Can.), Cambridge (Eng.), Copenhagen (Den.), Edinburgh (Scot.), Florence (Italy), Leiden (Neth.), London (Eng.), Marburg (Ger.), Montreal (Can.), Oxford (Eng.), the Ruhr (Ger.), Sussex (Eng.), Toronto (Can.), Victoria (Can.), and Waterloo (Can.); the Complutensian University of Madrid (Spain); the Max Planck Institute for Biophysical Chemistry (Ger.); the New University of Lisbon (Port.); the School of Higher Studies in Social Sciences (Fr.); Simon Fraser University (Can.); and York University (Can.).

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

Eleventh Edition
© 1911
By Encyclopædia Britannica, Inc.

Twelfth Edition
© 1922
By Encyclopædia Britannica, Inc.

Thirteenth Edition
© 1926
By Encyclopædia Britannica, Inc.

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

Fifteenth Edition
© 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991
By Encyclopædia Britannica, Inc.

© 1991
By Encyclopædia Britannica, Inc.

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

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

Printed in U.S.A.

Library of Congress Catalog Card Number: 89-81675
International Standard Book Number: 0-85229-529-4

CONTENTS

1 United Kingdom
140 United Nations
149 United States of America
458 Uruguay
465 Velázquez
468 Venezuela
480 Venice
488 Victoria and the Victorian Age
493 Vienna
499 Virgil
502 Viruses
512 Volcanism
524 Voltaire
529 The Technology of War
628 The Theory and Conduct of War
696 Warsaw
699 George Washington
707 Washington, D.C.
718 Wellington
721 The West Indies
790 Western Africa
916 Work and Employment
961 The World Wars
1025 Writing
1077 Yiddish Literature
1083 Yugoslav Literature
1086 Yugoslavia
1116 Zoroastrianism and Parsiism

United Kingdom

The United Kingdom (in full United Kingdom of Great Britain and Northern Ireland) is located off the northwestern coast of mainland Europe and consists of four geographic and historical parts—England, Scotland, Wales, and Northern Ireland. The United Kingdom makes up most of the British Isles—the geographic term for the group of islands that also includes the republic of Ireland. Together, England, Wales, and Scotland constitute the larger of the two principal islands, while Northern Ireland and the republic of Ireland constitute the second largest. To the southwest of England are the Isles of Scilly, and to the south is the Isle of Wight. The Orkney and Shetland islands lie to the north of Scotland, and the Hebrides lie to the west. Off northwestern Wales is the island of Anglesey.

Apart from the land border with the Irish republic, the United Kingdom is surrounded by sea. To the south of England, and between the United Kingdom and France, is the English Channel. The North Sea lies to the east. To the west of Wales and northern England is the Irish Sea, while western Scotland and the northwestern coast of Northern Ireland face the Atlantic Ocean.

The area of the United Kingdom is 94,251 square miles (244,110 square kilometres). At its widest the United Kingdom is 300 miles across. From the top of Scotland to the southern coast of England it is about 600 miles. No part is more than 75 miles from the sea. The national capital is London, situated on the River Thames in the southeastern corner of England.

The names United Kingdom, Great Britain, and England are often confused, even by U.K. inhabitants. England is just one country within the kingdom. Great Britain comprises England, Wales, and Scotland, while the United Kingdom also includes Northern Ireland (although the name Britain is sometimes used to refer to the United Kingdom as a whole). Wales and England were unified politically, administratively, and legally by the acts of union of 1536 and 1542. In 1707 Scotland joined England and Wales in forming a single parliament for Great Britain, although the three countries had previously shared a monarch.

The United Kingdom is characterized by a long history and by political and cultural links with other areas of the world, the latter mostly a legacy of its large former empire. In modern times the United Kingdom is perhaps best seen as a middle-sized, middle-ranking industrial country.

The political system of the United Kingdom has provided stability and consistency since the 19th century, albeit through a structure that has evolved rather than been designed. It is a unitary system centred on London, with some responsibilities devolved to local governments. The national government is a parliamentary democracy dominated by the monarchy, which links the executive, legislature, judiciary, armed forces, and Church of England. Although in practice almost all responsibilities are deferred, the monarch and the royal family are a source of unity and national spirit. In Parliament the House of Lords still consists mainly of hereditary or appointed peers, while members of the House of Commons are elected by a simple "first-past-the-post" system.

This stability of institutions contrasts with and complements the striking heterogeneous social character of the United Kingdom. There is a vocal nationalist spirit in both Wales and Scotland, while Northern Ireland is plagued by division between the Protestant and Roman Catholic communities. Northern Ireland's troubles have led to terrorist actions, most notably by the Provisional Irish Republican Army (IRA).

On the world stage, the United Kingdom is part of the European Communities, while retaining links with parts of its former empire through the Commonwealth. It also benefits from historical and cultural links with the United States and is a member of the North Atlantic Treaty Organization (NATO). This diversity of interests, together with a network of international relations, means that the United Kingdom is not easily identified as belonging to a specific geographic block among the Western industrialized nations.

Economically the United Kingdom has benefited since the 1970s from production of oil from deposits in the North Sea. London has become a leading world financial centre, the time zone allowing it to bridge the gap between trading in Tokyo and New York City. The United Kingdom's traditional strength in manufacturing, however, has been eroded, with employment in manufacturing falling in absolute terms. This has undoubtedly contributed to the stark differences in the social and economic composition of the industrialized north and the more service-oriented, prosperous south, creating a north–south divide.

Socially the United Kingdom suffers pockets of poverty, with some inner-city areas among the worst in Europe. The growth in ethnic minorities from former colonies has added to social tensions, occasionally fueling violence. In contrast, home ownership is widespread and, while the state supports an educational system, public schools—which despite their title are part of the private sector—thrive. With some exceptions, notably curbs on public servants, freedom of expression is not restricted, and the United Kingdom is renowned for the strength of its arts.

(R.C.At.)

This article begins with a discussion of the physical and human geography of the United Kingdom, followed by a history of England and Great Britain from prehistoric to present times. The final section of the article discusses the physical and human geography of England, Scotland, Wales, and Northern Ireland, the last three accompanied by histories.

The article is divided into the following sections:

Physical and human geography 2
- The land 2
  - Relief
  - Drainage
  - Soils
  - Climate
  - Plant and animal life
  - Settlement patterns
- The people 12
  - Ethnic composition
  - Linguistic composition
  - Religions
  - Demographic trends
- The economy 14
  - Resources
  - Agriculture, forestry, and fishing
  - Industry
  - Finance
  - Trade
  - Transportation
- Administration and social conditions 16
  - Government
  - Politics
  - Justice
  - Police services
  - The armed forces
  - Education
  - Health and welfare
- Cultural life 20
  - General considerations
  - Mass media

History of England and Great Britain 21
- Ancient Britain 21
  - Pre-Roman Britain

Roman Britain
Anglo-Saxon England 25
The invaders and their early settlements
The heptarchy
The period of the Scandinavian invasions
The achievement of political unity
The Anglo-Danish state
The Normans (1066–1154) 31
William I (1066–87)
The sons of William I
The period of anarchy (1135–54)
England in the Norman period
The early Plantagenets 33
Henry II (1154–89)
Richard I (1189–99)
John (1199–1216)
The 13th century 35
Henry III (1216–72)
Edward I (1272–1307)
Social, economic, and cultural change
The 14th century 38
Edward II (1307–27)
Edward III (1327–77)
Richard II (1377–99)
Economic crisis and cultural change
Lancaster and York 41
Henry IV (1399–1413)
Henry V (1413–22)
Henry VI (1422–61 and 1470–71)
Edward IV (1461–70 and 1471–83)
Richard III (1483–85)
England in the 15th century
England under the Tudors 44
Henry VII (1485–1509)
Henry VIII (1509–47)
Edward VI (1547–53)
Mary I (1553–58)
Elizabeth I (1558–1603)
The early Stuarts and the Commonwealth 53
England in 1603
James I (1603–25)
Charles I (1625–49)
The later Stuarts 61
Charles II (1660–85)
James II (1685–88)
William III (1689–1702) and Mary II (1689–94)
Anne (1702–14)
18th-century Britain, 1714–1815 65
The state of Britain in 1714
Britain from 1715 to 1742
Britain from 1742 to 1754
British society by the mid-18th century
Britain from 1754 to 1783
Britain from 1783 to 1815
Great Britain, 1815–1914 76
Britain after the Napoleonic Wars
The politics of reform
Social cleavage and social control in the early Victorian years
Mid-Victorian society and culture
Mid-Victorian politics
Gladstone and Disraeli
Late Victorian politics
Edwardian and prewar Britain
Britain from 1914 to the present 88
World War I
Between the wars
World War II
Britain since 1945
England 95
Physical and human geography 95
The land
The people
The economy
Administration and social conditions
Cultural life
History 102
Scotland 103
Physical and human geography 103
The land
The people
The economy
Administration and social conditions
Cultural life
History 108
Ancient times
The unification of the kingdom
The wars of independence
Scotland in the 15th century
Scotland in the 16th and early 17th centuries
The age of revolution (1625–89)
The era of Union
19th-century Scotland
Modern Scotland
Wales 118
Physical and human geography 118
The land
The people
The economy
Administration and social conditions
Cultural life
History 122
Wales before the Norman Conquest
Wales in the Middle Ages
Wales from the 16th to 20th century
Northern Ireland 128
Physical and human geography 128
The land
The people
The economy
Administration and social conditions
Cultural life
History 132
Early Ulster
Early modern Ulster
The 18th and 19th centuries
Northern Ireland since 1922
Bibliography 135

PHYSICAL AND HUMAN GEOGRAPHY

The land

Great Britain, the island comprising England, Scotland, and Wales, forms, together with numerous smaller islands, an archipelago that is as irregular in shape as it is diverse in its natural heritage. This latter circumstance stems largely from the nature and disposition of the underlying rocks, which are westward extensions of European structures, with the shallow waters of the Strait of Dover and the North Sea concealing former land links. Northern Ireland—which politically completes the United Kingdom—is a westward extension of the rock structures of Scotland. These common rock structures are breached by the narrow North Channel.

On a global scale, this natural endowment covers a small area—approximating that of Oregon, in the United States, or the African nation of Guinea—and its internal diversity, accompanied by rapid changes of often beautiful scenery, may perhaps convey to visitors from larger countries a striking sense of compactness and consolidation. The peoples who, over the centuries, have made their way to, and hewed an existence from, this Atlantic extremity of Eurasia have put their own imprint on the environment, with the ancient and distinctive palimpsest of their field patterns and settlements complementing the natural diversity.

RELIEF

The traditional division of Great Britain is into a Highland and Lowland zone. A line running from the mouth of the River Exe, in the southwest, to that of the Tees, in the northeast, is a crude expression of this division. The course of the 700-foot (213-metre) contour, or of the boundary separating the older rocks of the north and west from the younger southeastern strata, provides more appropriate indications of the extent of the Highlands. The Tees–Exe Line

The Highland zone. The creation of the Highlands was a long process, yet altitudes, compared with European equivalents, are low, with the highest summit, Ben Nevis, only 4,406 feet (1,343 metres) above sea level. In addition,

the really mountainous areas above 2,000 feet often lie in smooth profiles against the changing skies, reminders of the effects of former periods of erosion.

Scotland's three main topographic regions follow the northeast-to-southwest trend of the ancient underlying rocks. The northern Highlands and the Southern Uplands are separated by the intervening rift valley, or subsided structural block, of the Central Lowlands. The core of the Highlands is the elevated, worn-down surface of the Grampian Mountains, 1,000–3,500 feet above sea level, with the Cairngorm Mountains rising to over 4,000 feet. This majestic mountain landscape is furrowed by numerous wide valleys, or straths, and occasional large areas of lowland, often fringed with long lines of sand dunes, add variety to the east. The Buchan peninsula, the Moray Firth estuarine flats, and the plain of Caithness—all low-lying areas—contrast sharply with the mountain scenery and

Landscapes of the Scottish Highlands

show more mellow outlines than do the glacier-scoured landscapes of the west, where northeasterly-facing hollows, or corries, separated by knife-edge ridges and deep glens, sculpt the surfaces left by erosion. The many freshwater lochs further enhance a landscape of wild beauty. The linear Glen Mor, where the Caledonian Canal now threads a chain of lakes, is the result of a vast structural sideways tear in the whole mass of the North West Highlands. To the northwest of Glen Mor stretches most of the land given over to agricultural small holdings, or crofts; settlement is intermittent and mostly coastal, a pattern clearly reflecting the pronounced dissection of a highland massif that has been scored and plucked by the Ice Age glaciers. Many sea-drowned, glacier-widened river valleys (fjords) penetrate deeply into the mountains, the outliers of which rise from the sea in stately, elongated peninsulas or emerge in hundreds of offshore islands.

In comparison with the northern Highlands of Scotland, the Southern Uplands present a more subdued relief, the land nowhere rising above 2,800 feet. The main hill masses are the Cheviots, which rise to 2,676 feet, while Mount Merrick and Broad Law reach just above the 2,700-foot contour line. Broad plateau surfaces, separated by numerous dales, are again characteristic of these uplands, and in the west most of the rivers flow across the prevailing northeast–southwest trend, following the general slope of the plateau, toward the Solway Firth. Bold masses of granite and the rugged imprint of former glaciers occasionally impart a mountainous quality to the scenery. In the east the valley network of the Tweed and its many tributaries forms a broad lowland expanse between the Lammermuir and Cheviot hills.

The Central Lowlands are bounded by great regular structural faults. The northern boundary with the Highlands is a wall-like feature, but the boundary with the Southern Uplands exhibits a linear topographic form only near the coast. This vast trench is by no means a continuous plain, for high ground—often formed of sturdy, resistant masses of volcanic rock—meets the eye in all directions, rising above the low-lying areas that flank the rivers and the deeply penetrating estuaries of the Firth of Clyde and the Firth of Forth.

Northern Ireland's extensions of the Scottish Highlands

In Northern Ireland, structural extensions of Scottish Highland geomorphology reappear in the generally rugged mountain scenery and in the peat-covered summits of the 2,241-foot Sperrin Mountains. The uplands in Northern Ireland (the former counties Down and Armagh) are the western continuation of the Southern Uplands but rise over 500 feet only in limited areas; the one important exception is the Mourne Mountains, a lovely cluster of granite summits the loftiest of which, Slieve Donard, rises to 2,789 feet within two miles of the sea. Compared with that of the Scottish Central Lowlands, Northern Ireland's structure has been complicated by the outpouring of basaltic lavas to form a huge plateau, much of which is occupied by the shallow Lough Neagh, the largest freshwater lake in the British Isles.

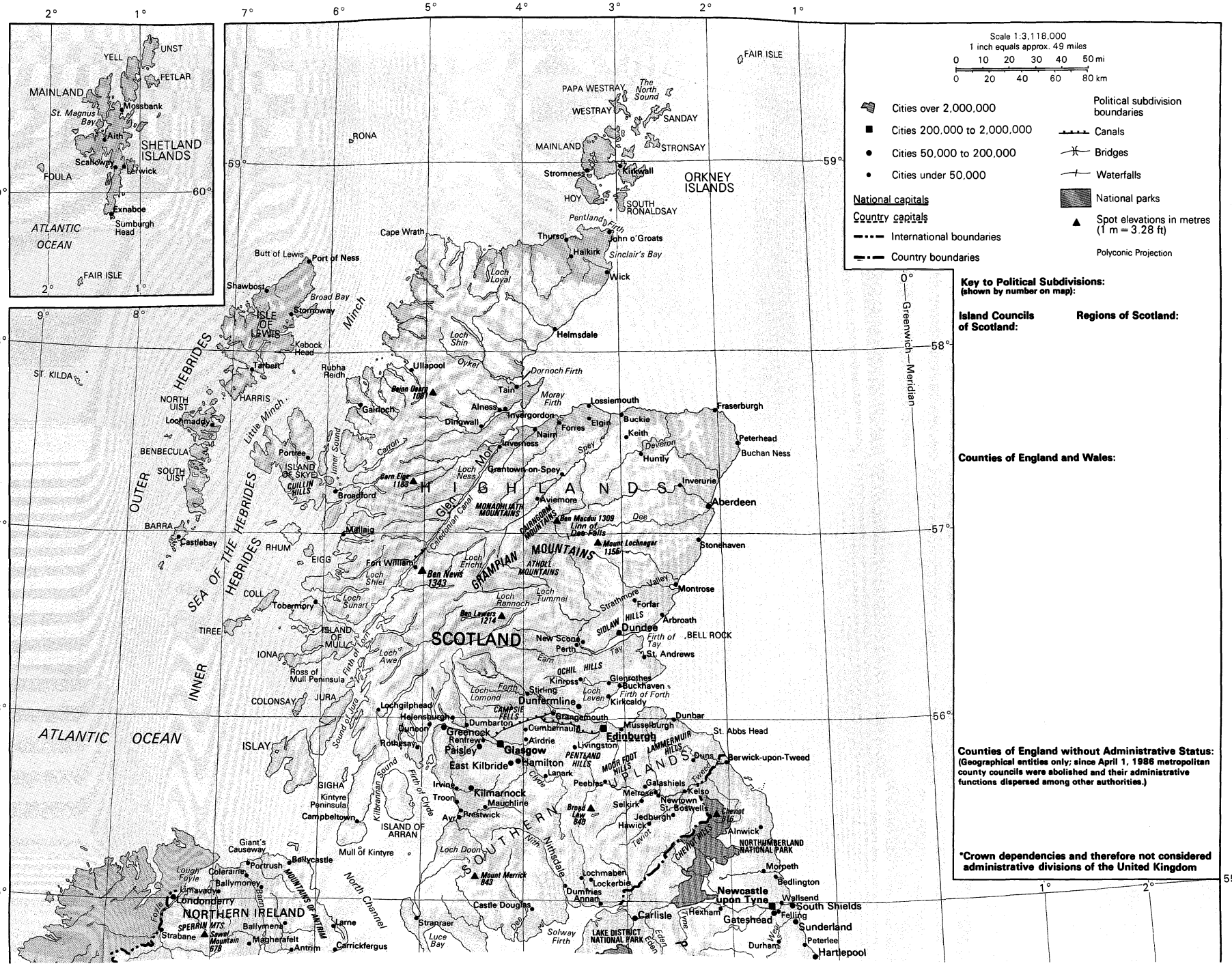
The Highland zone of England and of Wales consists, from north to south, of four broad upland masses: the Pennines, the Cumbrian Mountains, the Cambrian Mountains, and the South West peninsula. The Pennines are usually considered to end in the north along the River Tyne gap, but the surface features of several hills in Northumberland are in many ways similar to those of the northern Pennines. The general surface of the asymmetrically arched backbone (anticline) of the Pennines is remarkably smooth, because many of the valleys, though deep, occupy such a small portion within the total area that the windswept moorland between them appears almost featureless. This is particularly true of the landscape around Alston, in Cumberland, which—cut off by faults on its north, west, and south sides—stands out as an almost rectangular block of high moorland plateau with isolated peaks (known to geographers as monadnocks) rising up above it. Farther south the Pennine plateau is cut into by deep and scenic dales, their craggy sides formed of Millstone Grit, beneath which flow streams stepped by waterfalls. The most southerly part of the Pennines is a grassy upland, in places over 2,000 feet above sea level, but it is characterized by the dry valleys, steep-sided gorges, and underground streams and caverns of a limestone-drainage system rather than the bleak moorland that might be expected at this altitude. At lower levels the larger dales are more richly wooded, the trees standing out against a background of rugged cliffs of white-gray rocks. On both Pennine flanks, older rocks disappear beneath younger layers, and the uplands merge into flanking coastal lowlands.

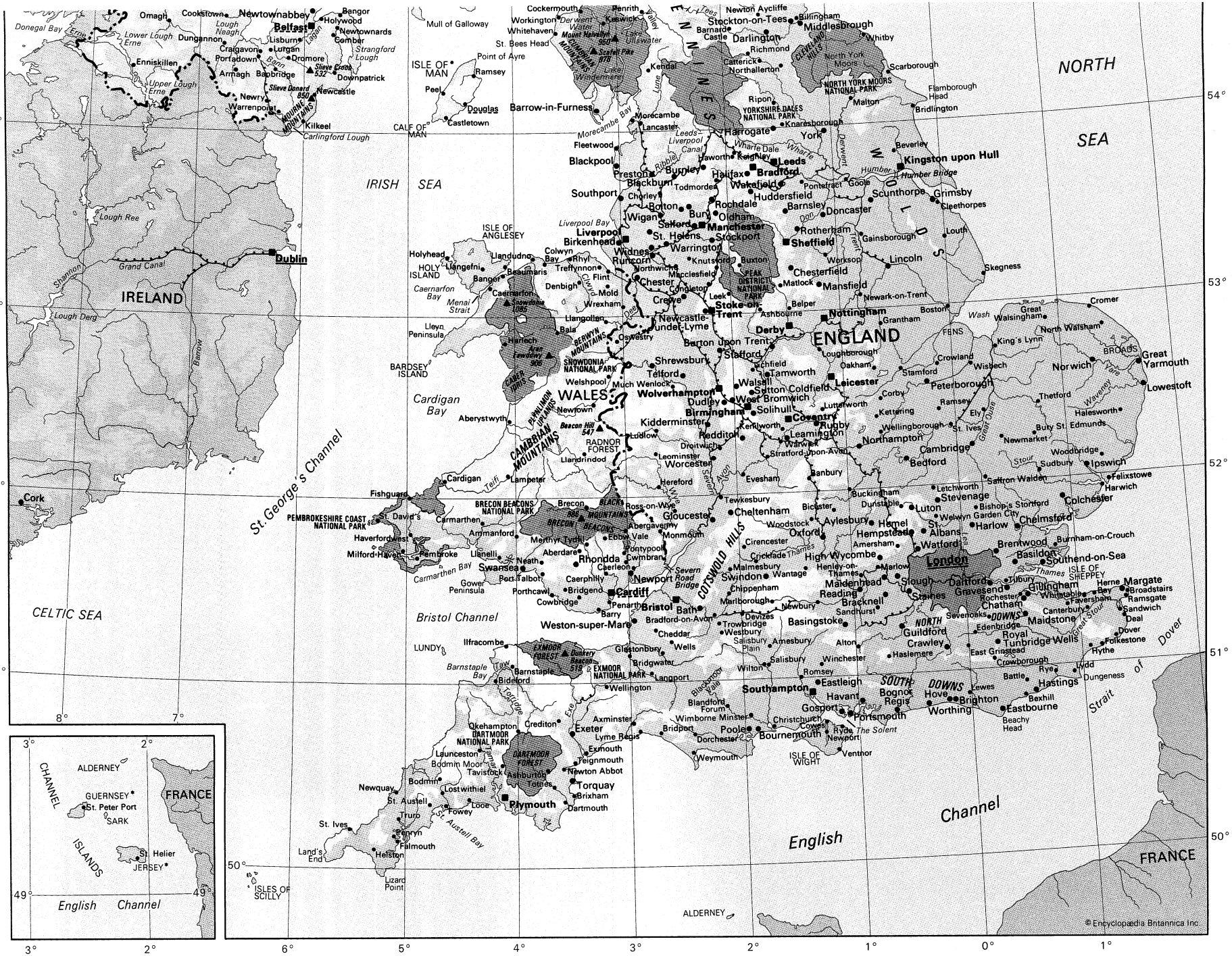
The Cumbrian Mountains, which include the famous Lake District celebrated in poetry by William Wordsworth and the other "Lake poets," constitute an isolated, compact mountain group to the west of the Pennines. The tough slate rocks of the northern portion have been cut into many deep gorges, separated by narrow ridges and sharp peaks. Greater expanses of level upland, formed from thick beds of lava and the ash thrown out by ancient volcanoes, are found to the south. Although Scafell Pike, at 3,210 feet, and Helvellyn, at 3,116 feet, are high for Britain, the volcanic belt is largely an irregular upland traversed by deep, narrow valleys. Nine rivers flowing out in all directions from the centre of this uplifted dome form a classic radial drainage pattern. The valleys, often occupied by long, narrow lakes, have been widened to a U shape by glacial action, which has also etched corries from the mountainsides and deposited the debris in moraines. Glacial action also created a number of "hanging valleys" by truncating former tributary valleys.

The Cambrian Mountains, which form the core of the principality of Wales, are clearly defined by the sea except on the eastern side, where a sharp break of slope often marks the transition to the English plain. Cycles of erosion have several times worn down the ancient and austere surfaces; many topographic features may be attributed to glacial processes. Some of the most striking scenery owes much to former volcanism. The mountain areas above 2,000 feet are most extensive in North Wales, in Snowdonia and its southward extensions, Cader Idris and the Berwyn mass. With the exception of Plynlimon and the Radnor Forest, central Wales lacks similar high areas, but the monadnocks of South Wales—notably the Black Mountains and the Brecon (Brecknock) Beacons—again stand out in solitary splendour above the upland surfaces. Three of these are distinguishable: a high plateau of 1,700 to 1,800 feet; a middle peneplain, or worn-down surface, of 1,200 to 1,600 feet; and a low peneplain of 700 to 1,100 feet. These smooth, rounded, grass-covered moorlands present a remarkably even skyline. Below 700 feet lies a further series of former wave-cut surfaces. Several valleys radiate from the highland core to the coastal regions. In the west these lowlands have provided a haven for traditional Welsh culture, but the deeply penetrating eastern valleys have channeled Anglicizing influences into the highland. A more extensive lowland—physically and structurally an extension of the English plain—borders the Bristol Channel in the southeast. The irregularities of the 600-mile Welsh coast exhibit differing adjustments to the pounding attack of the sea.

The South West peninsula

The South West—England's largest peninsula—has six locally conspicuous uplands: Exmoor, where Dunkery Beacon reaches 1,704 feet; the wild, granite uplands of Dartmoor (High Willhays; 2,038 feet); Bodmin Moor; St.


Scale 1:3,118,000
1 inch equals approx. 49 miles
0 10 20 30 40 50 mi
0 20 40 60 80 km
Cities over 2,000,000
Cities 200,000 to 2,000,000
Cities 50,000 to 200,000
Cities under 50,000
National capitals
Country capitals
International boundaries
Country boundaries
Political subdivision boundaries
Canals
Bridges
Waterfalls
National parks
Spot elevations in metres (1 m = 3.28 ft)
Polyconic Projection
Key to Political Subdivisions:
(shown by number on map):
Island Councils of Scotland:
Regions of Scotland:
Counties of England and Wales:
Counties of England without Administrative Status:
(Geographical entities only; since April 1, 1986 metropolitan county councils were abolished and their administrative functions dispersed among other authorities.)
*Crown dependencies and therefore not considered administrative divisions of the United Kingdom
SCOTLAND
NORTHERN IRELAND
HIGHLANDS
SOUTHERN UPLANDS
GRAMPIAN MOUNTAINS
ATLANTIC OCEAN
ORKNEY ISLANDS
SHETLAND ISLANDS
OUTER HEBRIDES
INNER HEBRIDES
SEA OF THE HEBRIDES
Greenwich Meridian


IRELAND
ENGLAND
WALES
SCOTLAND
NORTH SEA
IRISH SEA
Cardigan Bay
St. George's Channel
CELTIC SEA
Bristol Channel
English Channel
Dublin
Belfast
Cork
Liverpool
Manchester
Leeds
Sheffield
Birmingham
Nottingham
Leicester
Coventry
Derby
Stoke-on-Trent
Wolverhampton
Kingston upon Hull
Newcastle upon Tyne
Middlesbrough
Sunderland
Darlington
Harrogate
York
Bradford
Huddersfield
Wakefield
Barnsley
Doncaster
Rotherham
Scunthorpe
Grimsby
Lincoln
Skegness
Mansfield
Chesterfield
Derby
Nottingham
Boston
Great Yarmouth
Lowestoft
Norwich
Cromer
Peterborough
Cambridge
Ipswich
Colchester
Harwich
Felixstowe
Northampton
Leamington
Warwick
Coventry
Rugby
Stratford-upon-Avon
Banbury
Oxford
Aylesbury
Hemel Hempstead
St. Albans
Watford
London
Luton
Stevenage
Harlow
Chelmsford
Southend-on-Sea
Basildon
Brentwood
Burnham-on-Crouch
High Wycombe
Slough
Reading
Maidenhead
Bracknell
Basingstoke
Newbury
Swindon
Wantage
Henley-on-Thames
Marlow
Dartford
Gravesend
Tilbury
Gillingham
Chatham
Rochester
Maidstone
Canterbury
Margate
Ramsgate
Broadstairs
Deal
Sandwich
Dover
Folkestone
Ashford
Tunbridge Wells
Crawley
Guildford
Haslemere
Aldershot
Farnborough
Winchester
Salisbury
Amesbury
Andover
Southampton
Eastleigh
Havant
Portsmouth
Gosport
Fareham
Hove
Brighton
Worthing
Bognor Regis
Eastbourne
Hastings
Bexhill
Lewes
Battle
Rye
Dungeness
Romney
Folkestone
Dover
Strait of Dover
Channel
Isle of Wight
Newport
Ventnor
Isle of Sheppey
Isle of Anglesey
Isle of Man
Douglas
Ramsay
Peel
Castletown
Calf of Man
Isles of Scilly
St. Mary's
Land's End
Lizard Point
Plymouth
Torquay
Exeter
Bournemouth
Poole
Weymouth
Dorchester
Taunton
Bridgwater
Wells
Bath
Bristol
Cardiff
Swansea
Newport
Gloucester
Cheltenham
Worcester
Hereford
Shrewsbury
Telford
Wolverhampton
Stoke-on-Trent
Chester
Wrexham
Liverpool
Birkenhead
Preston
Blackpool
Lancaster
Carlisle
Workington
Whitehaven
Penrith
Keswick
Barrow-in-Furness
Morecambe
Middlesbrough
Stockton-on-Tees
Darlington
Richmond
Scarborough
Whitby
Malton
Ripon
Northallerton
Catterick
Kendal
Windermere
Ullswater
Derwent Water
Mount Helvellyn 950
Scafell Pike 978
Cleveland Hills
North York Moors
Yorkshire Dales National Park
North York Moors National Park
Peak District National Park
Lake District National Park
Snowdonia National Park
Brecon Beacons National Park
Pembrokeshire Coast National Park
Exmoor National Park
Dartmoor National Park
Dartmoor Forest
Cotswold Hills
Radnor Forest
Cambrian Mountains
Berwyn Mountains
Black Mountains
Brecon Beacons
Fens
Broads
Wash
Humber
Humber Bridge
Severn Road Bridge
Menai Strait
Carmarthen Bay
Swansea Bay
St. Austell Bay
Barnstaple Bay
Liverpool Bay
Morecambe Bay
Solway Firth
Lough Neagh
Lough Erne
Lough Ree
Lough Derg
Carlingford Lough
Strangford Lough
Donegal Bay
Grand Canal
Shannon
Erne
Barrow
Bann
Foyle
Tamar
Exe
Dart
Tweed
Clyde
Mersey
Ribble
Wharfe
Don
Trent
Severn
Thames
Great Ouse
Stour
Wye
Usk
Taff
Teifi
Tawe
Dee
Conwy
Aln
Tees
Swale
Ure
Nidd
Aire
Calder
Derwent
Ouse
Yare
Waveney
Lune
Ribble
Irwell
Weaver
Avon
Itchen
Test
Frome
Piddle
Arun
Adur
Ouse
Cuckmere
Rother
Medway
Darent
Lee
Colne
Stour
Orwell
Deben
Alde
Blackwater
Crouch
Roach
Chelmer
Brent
Lea
Wey
Kennet
Loddon
Cherwell
Windrush
Leach
Coln
Evenlode
Isis
Thame
Ouse
Nene
Welland
Witham
Glen
Bain
Ancholme
Hull
Derwent
Foss
Ouse
Wharfe
Aire
Calder
Don
Rother
Idle
Torne
Trent
Soar
Anker
Tame
Stour
Severn
Avon
Teme
Lugg
Wye
Usk
Taff
Ely
Tawe
Loughborough
Oakham
Stamford
Grantham
Newark-on-Trent
Worksop
Retford
Doncaster
Pontefract
Goole
Selby
Beverley
Driffield
Bridlington
Flamborough Head
Filey
Scarborough
Whitby
Saltburn
Redcar
Middlesbrough
Billingham
Stockton-on-Tees
Newton Aycliffe
Barnard Castle
Penrith
Cockermouth
Workington
Whitehaven
Keswick
Ambleside
Kendal
Lancaster
Morecambe
Fleetwood
Blackpool
Preston
Blackburn
Burnley
Todmorden
Halifax
Bradford
Keighley
Haworth
Leeds
Wakefield
Pontefract
Goole
Scunthorpe
Grimsby
Cleethorpes
Barnsley
Doncaster
Rotherham
Sheffield
Gainsborough
Louth
Lincoln
Skegness
Mansfield
Chesterfield
Matlock
Buxton
Worksop
Retford
Newark-on-Trent
Grantham
Boston
Spalding
Peterborough
Stamford
Oakham
Loughborough
Derby
Nottingham
Ashbourne
Belper
Leek
Stoke-on-Trent
Newcastle-under-Lyme
Crewe
Nantwich
Northwich
Macclesfield
Knutsford
Warrington
Widnes
Runcorn
Liverpool
Birkenhead
St. Helens
Wigan
Bolton
Chorley
Southport
Ormskirk
Skelmersdale
Rochdale
Oldham
Bury
Salford
Manchester
Stockport
Todmorden
Holyhead
Llangefni
Beaumaris
Bangor
Caernarfon
Llandudno
Colwyn Bay
Rhyl
Prestatyn
Denbigh
Flint
Mold
Wrexham
Ruthin
Corwen
Llangollen
Bala
Oswestry
Whitchurch
Ellesmere
Shrewsbury
Telford
Wellington
Much Wenlock
Bridgnorth
Kidderminster
Stourport
Bewdley
Redditch
Droitwich
Worcester
Malvern
Tewkesbury
Cheltenham
Gloucester
Stroud
Cirencester
Tetbury
Malmesbury
Chippenham
Swindon
Wantage
Abingdon
Oxford
Woodstock
Bicester
Buckingham
Aylesbury
Tring
Hemel Hempstead
St. Albans
Watford
Amersham
Chesham
High Wycombe
Henley-on-Thames
Marlow
Maidenhead
Slough
Windsor
Staines
Reading
Wokingham
Bracknell
Newbury
Andover
Basingstoke
Alton
Farnham
Guildford
Dorking
Epsom
Kingston upon Thames
Richmond
Twickenham
Hounslow
Heathrow
Uxbridge
Harrow
Wembley
Hampstead
Finchley
Enfield
Barnet
Hertford
Ware
Stevenage
Hitchin
Letchworth
Luton
Dunstable
Leighton Buzzard
Aylesbury
Thame
Wallingford
Abingdon
Didcot
Faringdon
Cirencester
Tetbury
Malmesbury
Chippenham
Melksham
Trowbridge
Frome
Shepton Mallet
Wells
Glastonbury
Street
Yeovil
Sherborne
Dorchester
Weymouth
Bridport
Lyme Regis
Axminster
Honiton
Sidmouth
Exmouth
Exeter
Crediton
Okehampton
Tavistock
Launceston
Bodmin
Wadebridge
Newquay
Truro
Falmouth
Helston
Penzance
St. Ives
St. Austell
Fowey
Looe
Liskeard
Saltash
Plymouth
Dartmouth
Totnes
Brixham
Torquay
Newton Abbot
Teignmouth
Dawlish
Exmouth
Sidmouth
Honiton
Axminster
Lyme Regis
Bridport
Dorchester
Weymouth
Poole
Bournemouth
Christchurch
Wimborne Minster
Blandford Forum
Ringwood
Lymington
Yarmouth
Newport
Ventnor
Cowes
Ryde
Isle of Wight
Portsmouth
Gosport
Fareham
Havant
Eastleigh
Southampton
Romsey
Salisbury
Wilton
Amesbury
Andover
Winchester
Alton
Farnham
Haslemere

MAP INDEX

Political subdivisions

Avon .....51 26 N 2 30 W
Bedfordshire .....52 08 N 0 27 W
Berkshire .....51 30 N 1 15 W
Borders .....55 25 N 2 40 W
Buckinghamshire ...51 45 N 0 45 W
Cambridgeshire ....52 12 N 0 07 E
Central .....56 05 N 4 07 W
Cheshire .....53 15 N 2 30 W
Cleveland .....54 30 N 1 13 W
Clwyd .....53 05 N 3 30 W
Cornwall .....50 25 N 4 40 W
Cumbria .....54 37 N 3 00 W
Derbyshire .....53 05 N 1 35 W
Devon .....50 45 N 3 45 W
Dorset .....50 45 N 2 25 W
Dumfries and Galloway .....55 05 N 3 40 W
Durham .....54 40 N 1 45 W
Dyfed .....51 55 N 4 10 W
East Sussex .....50 55 N 0 15 E
Essex .....51 45 N 0 30 E
Fife .....56 10 N 3 05 W
Gloucestershire ....51 50 N 2 15 W
Grampian .....57 16 N 2 50 W
Gwent .....51 40 N 3 00 W
Gwynedd .....52 53 N 3 48 W
Hampshire .....51 00 N 1 15 W
Hereford and Worcester .....52 09 N 2 30 W
Hertfordshire .....51 50 N 0 15 W
Highland .....57 35 N 5 00 W
Humberside .....53 45 N 0 25 W
Isle of Wight .....50 40 N 1 15 W
Kent .....51 15 N 0 45 E
Lancashire .....53 55 N 2 35 W
Leicestershire .....52 40 N 1 10 W
Lincolnshire .....53 10 N 0 20 W
Lothian .....55 53 N 3 10 W
Mid Glamorgan ....51 40 N 3 40 W
Norfolk .....52 40 N 1 00 E
North Yorkshire ....54 13 N 1 25 W
Northamptonshire ..52 20 N 0 55 W
Northumberland ...55 15 N 2 00 W
Nottinghamshire ...53 05 N 1 00 W
Orkney Islands ....59 00 N 3 00 W
Oxfordshire .....51 45 N 1 10 W
Powys .....52 07 N 3 25 W
Shetland Islands ...60 30 N 1 00 W
Shropshire .....52 40 N 2 35 W
Somerset .....51 05 N 3 00 W
South Glamorgan ..51 25 N 3 20 W
Staffordshire .....52 50 N 2 00 W
Strathclyde .....56 00 N 5 10 W
Suffolk .....52 15 N 1 00 E
Surrey .....51 15 N 0 20 W
Tayside .....56 30 N 3 30 W
Warwickshire .....52 20 N 1 30 W
West Glamorgan ...51 35 N 3 45 W
West Sussex .....50 55 N 0 30 W
Western Isles .....57 45 N 7 00 W
Wiltshire .....51 20 N 1 55 W

Cities and towns

Aberdare (Aberdâr) .....51 43 N 3 27 W
Aberdeen .....57 09 N 2 08 W
Abergavenny .....51 49 N 3 01 W
Aberystwyth .....52 25 N 4 05 W
Airdrie .....55 52 N 3 59 W
Aith .....60 17 N 1 23 W
Alness .....57 42 N 4 15 W
Alnwick .....55 25 N 1 43 W
Alton .....51 09 N 0 59 W
Amersham .....51 41 N 0 36 W
Amesbury .....51 10 N 1 47 W
Ammanford .....51 48 N 4 00 W
Annan .....54 59 N 3 16 W
Antrim .....54 42 N 6 12 W
Arbroath .....56 34 N 2 35 W
Armagh .....54 21 N 6 40 W
Ashbourne .....53 01 N 1 44 W
Ashburton .....50 31 N 3 45 W
Aviemore .....57 11 N 3 50 W
Axminster .....50 47 N 3 00 W
Aylesbury .....51 49 N 0 49 W
Ayr .....55 28 N 4 37 W
Bae Colwyn, see Colwyn Bay
Bala .....52 55 N 3 36 W
Ballycastle .....55 13 N 6 14 W
Ballymena .....54 52 N 6 16 W
Ballymoney .....55 05 N 6 31 W
Banbridge .....54 21 N 6 17 W
Banbury .....52 04 N 1 20 W
Bangor .....53 13 N 4 08 W
Bangor .....54 39 N 5 42 W
Barnard Castle ....54 33 N 1 55 W
Barnsley .....53 33 N 1 29 W
Barnstaple .....51 05 N 4 03 W
Barrow-in-Furness ..54 07 N 3 14 W
Barry .....51 25 N 3 16 W
Basildon .....51 35 N 0 28 W
Basingstoke .....51 16 N 1 05 W
Bath .....51 23 N 2 22 W
Battle .....50 55 N 0 29 E
Beaumaris .....53 16 N 4 06 W
Bedford .....52 08 N 0 28 W
Bedlington .....55 08 N 1 35 W
Belfast .....54 35 N 5 56 W
Belper .....53 01 N 1 28 W
Berwick-upon-Tweed .....55 46 N 2 01 W
Beverley .....53 51 N 0 25 W
Bexhill .....50 51 N 0 28 E
Bicester .....51 54 N 1 09 W
Bideford .....51 01 N 4 12 W
Billingham .....54 36 N 1 17 W
Birkenhead .....53 23 N 3 01 W
Birmingham .....52 29 N 1 51 W
Bishop's Stortford .....51 52 N 0 10 E
Blackburn .....53 45 N 2 29 W
Blackpool .....53 49 N 3 03 W
Blandford Forum .....50 51 N 2 10 W
Bodmin .....50 29 N 4 44 W
Bognor Regis .....50 47 N 0 41 W
Bolton .....53 34 N 2 26 W
Boston .....52 59 N 0 01 W
Bournemouth .....50 43 N 1 53 W
Bracknell .....51 25 N 0 45 W
Bradford .....53 47 N 1 45 W
Bradford-on-Avon ..51 21 N 2 15 W
Brecon .....51 57 N 3 23 W
Brentwood .....51 37 N 0 19 E
Bridgend (Pen-y-bont Ar Ogwr) .....51 31 N 3 35 W
Bridgwater .....51 08 N 3 00 W
Bridlington .....54 05 N 0 11 W
Bridport .....50 44 N 2 46 W
Brighton .....50 51 N 0 08 W
Bristol .....51 26 N 2 35 W
Brixham .....50 23 N 3 31 W
Broadford .....57 15 N 5 54 W
Broadstairs .....51 21 N 1 26 E
Buckhaven .....56 11 N 3 02 W
Buckie .....57 41 N 2 59 W
Buckingham .....52 00 N 0 59 W
Burnham-on-Crouch .....51 38 N 0 50 E
Burnley .....53 47 N 2 15 W
Burton upon Trent .....52 48 N 1 38 W
Bury .....53 35 N 2 19 W
Bury St. Edmunds .....52 15 N 0 43 E
Buxton .....53 16 N 1 54 W
Caerleon .....51 37 N 2 58 W
Caernarfon .....53 08 N 4 16 W
Caerphilly .....51 34 N 3 13 W
Cambridge .....52 12 N 0 09 E
Campbeltown .....55 26 N 5 37 W
Canterbury .....51 17 N 1 05 E
Cardiff .....51 29 N 3 11 W
Cardigan .....52 05 N 4 40 W
Carlisle .....54 53 N 2 57 W
Carmarthen .....51 52 N 4 18 W
Carrickfergus .....54 43 N 5 44 W
Castell Nedd, see Neath
Castle Douglas ....54 57 N 3 56 W
Castlebay .....56 57 N 7 29 W
Castletown .....54 04 N 4 39 W
Catterick .....54 22 N 1 37 W
Chatham .....51 22 N 0 32 E
Cheddar .....51 17 N 2 47 W
Chelmsford .....51 44 N 0 28 E
Cheltenham .....51 54 N 2 05 W
Chester .....53 12 N 2 54 W
Chesterfield .....53 14 N 1 25 W
Chippenham .....51 28 N 2 07 W
Chorley .....53 39 N 2 38 W
Christchurch .....50 44 N 1 47 W
Cirencester .....51 43 N 1 58 W
Cleethorpes .....53 33 N 0 02 W
Cockermouth .....54 40 N 3 21 W
Colchester .....51 54 N 0 54 E
Coleraine .....55 08 N 6 40 W
Colwyn Bay (Bae Colwyn) .....53 17 N 3 42 W
Comber .....54 33 N 5 42 W
Congleton .....53 10 N 2 12 W
Cookstown .....54 39 N 6 45 W
Corby .....52 30 N 0 42 W
Coventry .....52 24 N 1 31 W
Cowbridge (Y Bont Faen) ....51 28 N 3 27 W
Cowes .....50 45 N 1 18 W
Craigavon .....54 28 N 6 25 W
Crawley .....51 07 N 0 11 W
Crediton .....50 48 N 3 39 W
Crewe .....53 06 N 2 27 W
Cricklade .....51 38 N 1 52 W
Cromer .....52 56 N 1 18 E
Crowborough .....51 03 N 0 10 E
Crowland .....52 41 N 0 10 W
Cumbernauld .....55 57 N 3 59 W
Cwmbran .....51 39 N 3 01 W
Darlington .....54 32 N 1 34 W
Dartford .....51 26 N 0 12 E
Dartmouth .....50 21 N 3 35 W
Deal .....51 13 N 1 24 E
Denbigh (Dinbych) .....53 11 N 3 25 W
Derby .....52 55 N 1 28 W
Devizes .....51 21 N 2 00 W
Dingwall .....57 36 N 4 26 W
Doncaster .....53 31 N 1 08 W
Dorchester .....50 43 N 2 27 W
Douglas .....54 09 N 4 28 W
Dover .....51 08 N 1 19 E
Downpatrick .....54 19 N 5 42 W
Droitwich .....52 16 N 2 09 W
Dromore .....54 25 N 6 09 W
Dudley .....52 31 N 2 06 W
Dumbarton .....55 57 N 4 35 W
Dumfries .....55 04 N 3 36 W
Dunbar .....56 00 N 2 31 W
Dundee .....56 29 N 3 02 W
Dunfermline .....56 04 N 3 26 W
Dungannon .....54 31 N 6 45 W
Dunoon .....55 57 N 4 55 W
Duns .....55 46 N 2 21 W
Dunstable .....51 53 N 0 32 W
Durham .....54 47 N 1 34 W
East Grinstead ....51 08 N 0 00 W
East Kilbride .....55 46 N 4 11 W
Eastbourne .....50 47 N 0 16 E
Eastleigh .....50 58 N 1 21 W
Ebbw Vale .....51 47 N 3 13 W
Edenbridge .....51 12 N 0 04 W
Edinburgh .....55 57 N 3 10 W
Elgin .....57 39 N 3 19 W
Ely .....52 24 N 0 16 E
Enniskillen .....54 22 N 7 38 W
Evesham .....52 05 N 1 57 W
Exeter .....50 43 N 3 31 W
Exmouth .....50 37 N 3 24 W
Exnaboe .....59 54 N 1 18 W
Falmouth .....50 09 N 5 04 W
Faversham .....51 19 N 0 54 E
Felixstowe .....51 58 N 1 21 E
Felling .....54 57 N 1 33 W
Fishguard .....52 00 N 4 59 W
Fleetwood .....53 55 N 3 02 W
Flint .....53 15 N 3 08 W
Folkestone .....51 05 N 1 10 E
Forfar .....56 39 N 2 53 W
Forres .....57 36 N 3 37 W
Fort William .....56 49 N 5 06 W
Fowey .....50 20 N 4 38 W
Fraserburgh .....57 42 N 2 01 W
Gainsborough .....53 24 N 0 47 W
Gairloch .....57 44 N 5 41 W
Galashiels .....55 37 N 2 48 W
Gateshead .....54 56 N 1 37 W
Gillingham .....51 23 N 0 33 E
Glasgow .....55 52 N 4 15 W
Glastonbury .....51 09 N 2 43 W
Glenrothes .....56 12 N 3 10 W
Gloucester .....51 52 N 2 14 W
Goole .....53 42 N 0 52 W
Gosport .....50 48 N 1 09 W
Grangemouth .....56 01 N 3 43 W
Grantham .....52 55 N 0 38 W
Grantown-on-Spey .....57 20 N 3 36 W
Gravesend .....51 27 N 0 23 E
Great Walsingham .....52 54 N 0 54 E
Great Yarmouth ...52 36 N 1 44 E
Greenock .....55 57 N 4 46 W
Grimsby .....53 34 N 0 05 W
Guildford .....51 14 N 0 35 W
Halesworth .....52 21 N 1 30 E
Halifax .....53 44 N 1 51 W
Halkirk .....58 31 N 3 29 W
Hamilton .....55 47 N 4 02 W
Harlech .....52 52 N 4 06 W
Harlow .....51 46 N 0 07 E
Harrogate .....54 00 N 1 32 W
Hartlepool .....54 41 N 1 13 W
Harwich .....51 56 N 1 16 E
Haslemere .....51 05 N 0 43 W
Hastings .....50 52 N 0 35 E
Havant .....50 51 N 0 59 W
Haverfordwest .....51 48 N 4 58 W
Hawick .....55 26 N 2 47 W
Haworth .....53 50 N 1 57 W
Helensburgh .....56 00 N 4 44 W
Helmsdale .....53 07 N 3 39 W
Helston .....50 06 N 5 16 W
Hemel Hempstead .....51 45 N 0 28 W
Henley-on-Thames ..51 32 N 0 54 W
Hereford .....52 04 N 2 43 W
Herne Bay .....51 22 N 1 07 E
Hexham .....54 58 N 2 06 W
High Wycombe ....51 38 N 0 45 W
Holyhead .....53 19 N 4 38 W
Holywell, see Treffynnon
Holywood .....54 38 N 5 49 W
Hove .....50 50 N 0 11 W
Huddersfield .....53 39 N 1 47 W
Huntly .....57 27 N 2 48 W
Hythe .....51 04 N 1 05 E
Ilfracombe .....51 21 N 4 08 W
Invergordon .....57 41 N 4 09 W
Inverness .....57 29 N 4 14 W
Inverurie .....57 17 N 2 22 W
Ipswich .....52 03 N 1 09 E
Irvine .....55 37 N 4 40 W
Jedburgh .....55 28 N 2 33 W
John o'Groats .....58 38 N 3 04 W
Keighley .....53 52 N 1 54 W
Keith .....57 32 N 2 58 W
Kelso .....55 36 N 2 26 W
Kendal .....54 19 N 2 45 W
Kenilworth .....52 20 N 1 34 W
Keswick .....54 36 N 3 08 W
Kettering .....52 24 N 0 43 W
Kidderminster .....52 23 N 2 16 W
Kilkeel .....54 04 N 6 01 W
Kilmarnock .....55 36 N 4 30 W
King's Lynn .....52 45 N 0 24 E
Kingston upon Hull .....53 45 N 0 20 W
Kinross .....56 12 N 3 25 W
Kirkcaldy .....56 07 N 3 10 W
Kirkwall .....58 59 N 2 57 W
Knaresborough ....54 01 N 1 28 W
Knutsford .....53 18 N 2 22 W
Lampeter .....52 07 N 4 05 W
Lanark .....55 40 N 3 46 W
Lancaster .....54 03 N 2 48 W
Langport .....51 02 N 2 49 W
Larne .....54 51 N 5 49 W
Launceston .....50 38 N 4 21 W
Leamington .....52 15 N 1 33 W
Leeds .....53 48 N 1 32 W
Leek .....53 06 N 2 01 W
Leicester .....52 38 N 1 08 W
Leominster .....52 14 N 2 44 W
Lerwick .....60 09 N 1 09 W
Letchworth .....51 58 N 0 13 W
Lewes .....50 52 N 0 01 W
Lichfield .....52 41 N 1 50 W
Limavady .....55 03 N 6 58 W
Lincoln .....53 14 N 0 32 W
Lisburn .....54 31 N 6 04 W
Liverpool .....53 25 N 2 57 W
Livingston .....55 54 N 3 30 W
Llandrindod (Llandrindod Wells) .....52 15 N 3 23 W
Llandudno .....53 19 N 3 50 W
Llanelli .....51 41 N 4 10 W
Llanelwy, see St. Asaph
Llangefni .....53 15 N 4 18 W
Llangollen .....52 58 N 3 10 W

Lochgilphead .....56 03 N 5 26 W
Lochmaben .....55 08 N 3 26 W
Lochmaddy .....57 36 N 7 10 W
Lockerbie .....55 07 N 3 22 W
London .....51 30 N 0 07 W
Londonderry .....55 00 N 7 20 W
Looe .....50 21 N 4 27 W
Lossiemouth .....57 43 N 3 17 W
Lostwithiel .....50 24 N 4 40 W
Loughborough .....54 46 N 1 13 W
Louth .....53 22 N 0 00 E
Lowestoft .....52 28 N 1 45 E
Ludlow .....52 22 N 2 43 W
Lurgan .....54 28 N 6 20 W
Luton .....51 53 N 0 25 W
Lutterworth .....52 27 N 1 12 W
Lydd .....50 57 N 0 55 E
Lyme Regis .....50 44 N 2 56 W
Macclesfield .....53 16 N 2 08 W
Magherafelt .....54 46 N 6 36 W
Maidenhead .....51 32 N 0 44 W
Maidstone .....51 16 N 0 32 E
Mallaig .....57 00 N 5 50 W
Malmesbury .....51 35 N 2 06 W
Malton .....54 08 N 0 48 W
Manchester .....53 29 N 2 15 W
Mansfield .....53 09 N 1 12 W
Margate .....51 23 N 1 23 E
Marlborough .....51 25 N 1 44 W
Marlow .....51 34 N 0 46 W
Matlock .....53 08 N 1 34 W
Mauchline .....55 31 N 4 23 W
Melrose .....55 36 N 2 43 W
Merthyr Tydfil .....51 45 N 3 22 W
Middlesbrough .....54 34 N 1 12 W
Milford Haven .....51 43 N 5 02 W
Mold .....53 10 N 3 09 W
Monmouth .....51 49 N 2 43 W
Montrose .....56 43 N 2 28 W
Morecambe .....54 04 N 2 52 W
Morpeth .....55 10 N 1 41 W
Mossbank .....60 28 N 1 12 W
Much Wenlock .....52 36 N 2 33 W
Musselburgh .....55 57 N 3 02 W
Nairn .....57 35 N 3 52 W
Neath (Castell Nedd) .....51 40 N 3 49 W
New Scone .....56 25 N 3 24 W
Newark-on-Trent .....53 04 N 0 49 W
Newbury .....51 24 N 1 19 W
Newcastle .....54 12 N 5 53 W
Newcastle-under-Lyme .....53 01 N 2 13 W
Newcastle upon Tyne .....54 58 N 1 36 W
Newmarket .....52 15 N 0 25 E
Newport .....50 42 N 1 18 W
Newport .....51 35 N 2 59 W
Newquay .....50 25 N 5 05 W
Newry .....54 11 N 6 20 W
Newton Abbot .....50 32 N 3 36 W
Newton Aycliffe .....54 37 N 1 35 W
Newtown .....52 31 N 3 18 W
Newtown St. Boswells .....55 34 N 2 41 W
Newtownabbey .....54 40 N 5 57 W
Newtownards .....54 35 N 5 40 W
North Walsham .....52 49 N 1 24 E
Northallerton .....54 21 N 1 25 W
Northampton .....52 15 N 0 54 W
Northwich .....53 15 N 2 31 W
Norwich .....52 38 N 1 18 E
Nottingham .....52 58 N 1 10 W
Oakham .....52 41 N 0 44 W
Okehampton .....50 45 N 4 00 W
Oldham .....53 32 N 2 07 W
Omagh .....54 36 N 7 18 W
Oswestry .....52 52 N 3 03 W
Oxford .....51 45 N 1 15 W
Paisley .....55 50 N 4 25 W
Peebles .....55 39 N 3 11 W
Peel .....54 14 N 4 42 W
Pembroke .....51 40 N 4 55 W
Pen-y-bont Ar Ogwr, see Bridgend
Penarth .....51 26 N 3 10 W
Penrith .....54 40 N 2 45 W
Penryn .....50 10 N 5 06 W
Perth .....56 24 N 3 25 W
Peterborough .....52 35 N 0 14 W
Peterhead .....57 31 N 1 47 W
Peterlee .....54 45 N 1 20 W
Plymouth .....50 23 N 4 09 W
Pontefract .....53 41 N 1 19 W
Pontypool .....51 42 N 3 02 W
Poole .....50 43 N 1 59 W
Port of Ness .....58 29 N 6 14 W
Port Talbot .....51 35 N 3 48 W
Portadown .....54 26 N 6 27 W
Porthcawl .....51 29 N 3 42 W
Portree .....57 25 N 6 11 W
Portrush .....55 13 N 6 39 W
Portsmouth .....50 49 N 1 04 W
Preston .....53 46 N 2 42 W
Prestwick .....55 30 N 4 37 W
Ramsey .....52 27 N 0 07 W
Ramsey .....54 19 N 4 23 W
Ramsgate .....51 20 N 1 25 E
Reading .....51 27 N 0 58 W
Redditch .....52 18 N 1 56 W
Renfrew .....55 52 N 4 24 W
Rhondda .....51 39 N 3 29 W
Rhyl .....53 19 N 3 29 W
Richmond .....54 25 N 1 44 W
Ripon .....54 08 N 1 31 W
Rochdale .....53 37 N 2 10 W
Rochester .....51 23 N 0 29 E
Romsey .....51 00 N 1 30 W
Ross-on-Wye .....51 55 N 2 35 W
Rotherham .....53 26 N 1 21 W
Rothesay .....55 50 N 5 03 W
Royal Tunbridge Wells .....51 08 N 0 16 E
Rugby .....52 23 N 1 16 W
Runcorn .....53 19 N 2 44 W
Ryde .....50 44 N 1 10 W
Rye .....50 57 N 0 44 E
Saffron Walden .....52 01 N 0 15 E
St. Albans .....51 45 N 0 20 W
St. Andrews .....56 20 N 2 47 W
St. Asaph (Llanelwy) .....53 16 N 3 27 W
St. Austell .....50 20 N 4 47 W
St. David's .....51 53 N 5 16 W
St. Helens .....53 27 N 2 45 W
St. Helier .....49 11 N 2 06 W
St. Ives .....50 13 N 5 29 W
St. Ives .....52 20 N 0 05 W
St. Peter Port .....49 27 N 2 32 W
Salford .....53 29 N 2 18 W
Salisbury .....51 04 N 1 48 W
Sandhurst .....51 21 N 0 48 W
Sandwich .....51 17 N 1 21 E
Scalloway .....60 08 N 1 16 W
Scarborough .....54 17 N 0 25 W
Scunthorpe .....53 35 N 0 39 W
Selkirk .....55 33 N 2 50 W
Sevenoaks .....51 17 N 0 11 E
Shawbost .....58 19 N 6 40 W
Sheffield .....53 23 N 1 28 W
Shrewsbury .....52 42 N 2 45 W
Skegness .....53 09 N 0 20 E
Slough .....51 30 N 0 34 W
Solihull .....52 25 N 1 46 W
South Shields .....54 59 N 1 26 W
Southampton .....50 55 N 1 24 W
Southend-on-Sea .....51 33 N 0 42 E
Southport .....53 39 N 3 00 W
Stafford .....52 49 N 2 07 W
Staines .....51 26 N 0 30 W
Stamford .....52 39 N 0 29 W
Stevenage .....51 54 N 0 12 W
Stirling .....56 07 N 3 56 W
Stockport .....53 25 N 2 10 W
Stockton-on-Tees .....54 35 N 1 20 W
Stoke-on-Trent .....53 01 N 2 11 W
Stonehaven .....56 58 N 2 12 W
Stornoway .....58 12 N 6 23 W
Strabane .....54 48 N 7 28 W
Stranraer .....54 54 N 5 02 W
Stratford-upon-Avon .....52 12 N 1 43 W
Stromness .....58 58 N 3 18 W
Sudbury .....52 02 N 0 44 E
Sunderland .....54 55 N 1 23 W
Sutton Coldfield .....52 33 N 1 48 W
Swansea .....51 38 N 3 58 W
Swindon .....51 34 N 1 46 W
Tain .....57 48 N 4 03 W
Tamworth .....52 38 N 1 40 W
Tarbert .....57 54 N 6 48 W
Tavistock .....50 33 N 4 08 W
Teignmouth .....50 33 N 3 29 W
Telford .....52 41 N 2 28 W
Tewkesbury .....51 59 N 2 09 W
Thetford .....52 25 N 0 45 E
Thurso .....58 36 N 3 31 W
Tilbury .....51 28 N 0 22 E
Tobermory .....56 38 N 6 04 W
Todmorden .....53 43 N 2 05 W
Torquay .....50 29 N 3 32 W
Totnes .....50 26 N 3 41 W
Treffynnon (Holywell) .....53 16 N 3 13 W
Troon .....55 32 N 4 39 W
Trowbridge .....51 19 N 2 13 W
Truro .....50 16 N 5 03 W
Ullapool .....57 54 N 5 10 W
Ventnor .....50 36 N 1 12 W
Wakefield .....53 41 N 1 30 W
Wallsend .....55 00 N 1 32 W
Walsall .....52 35 N 1 59 W
Wantage .....51 35 N 1 25 W
Warrenpoint .....54 09 N 6 16 W
Warrington .....53 24 N 2 36 W
Warwick .....52 17 N 1 35 W
Watford .....51 40 N 0 24 W
Wellingborough .....52 18 N 0 41 W
Wellington .....50 59 N 3 13 W
Wells .....51 12 N 2 39 W
Welshpool (Y Trallwng) .....52 40 N 3 09 W
Welwyn Garden City .....51 48 N 0 12 W
West Bromwich .....52 32 N 2 01 W
Westbury .....51 15 N 2 12 W
Weston-super-Mare .....51 21 N 2 58 W
Weymouth .....50 37 N 2 28 W
Whitby .....54 29 N 0 37 W
Whitehaven .....54 33 N 3 35 W
Whitstable .....51 22 N 1 02 E
Wick .....58 26 N 3 06 W
Widnes .....53 22 N 2 44 W
Wigan .....53 33 N 2 38 W
Wilton .....51 05 N 1 52 W
Wimborne Minster .....50 48 N 1 59 W
Winchester .....51 04 N 1 19 W
Wisbech .....52 40 N 0 10 E
Wolverhampton .....52 35 N 2 08 W
Woodbridge .....52 06 N 1 18 E
Woodstock .....51 51 N 1 21 W
Worcester .....52 12 N 2 13 W
Workington .....54 38 N 3 34 W
Worksop .....53 19 N 1 07 W
Worthing .....50 49 N 0 23 W
Wrexham .....53 03 N 3 00 W
Y Bont Faen, see Cowbridge
Y Trallwng, see Welshpool
York .....53 57 N 1 06 W

Physical features and points of interest

Albyn, see Mor, Glen
Alderney, *island* .....49 43 N 2 12 W
Anglesey (Ynys Môn), Isle of .....53 17 N 4 21 W
Antrim, Mountains of .....55 00 N 6 10 W
Aran Fawddwy, *peak* .....52 47 N 3 41 W
Arran, Island of .....55 37 N 5 15 W
Atholl Mountains .....56 49 N 3 53 W
Atlantic Ocean .....56 00 N 8 00 W
Avon, *river* .....52 00 N 2 09 W
Awe, Loch .....56 17 N 5 14 W
Ayre, Point of .....54 25 N 4 22 W
Bann, *river* .....55 10 N 6 45 W
Bardsey Island (Ynys Enlli) .....52 46 N 4 47 W
Barnstaple (Bideford) Bay .....51 05 N 4 25 W
Barra, *island* .....56 59 N 7 28 W
Beachy Head, *cape* .....50 44 N 0 15 E
Beacon Hill .....52 23 N 3 12 W
Bell Rock (Inchcape), *island* .....56 26 N 2 24 W
Benbecula, *island* .....57 27 N 7 19 W
Berwyn Mountains .....52 52 N 3 27 W
Bideford, see Barnstaple Bay
Black Mountains .....51 56 N 3 06 W
Blackmoor Vale .....50 55 N 2 24 W
Bodmin Moor .....50 34 N 4 33 W
Brecon Beacons, *mountains* .....51 53 N 3 26 W
Brecon Beacons National Park .....51 53 N 3 24 W
Bristol Channel .....51 32 N 4 12 W
Broad Bay (Loch a Tuath) .....58 15 N 6 16 W
Broad Law, *mountain* .....55 30 N 3 21 W
Broads, *region* .....52 40 N 1 30 E
Buchan Ness, *point* .....57 28 N 1 45 W
Butt of Lewis, *cape* .....58 31 N 6 16 W
Cader Idris, *mountains* .....52 42 N 3 54 W
Caernarfon Bay .....53 04 N 4 22 W
Cairngorm Mountains .....57 06 N 3 40 W
Calais, Pas de, see Dover, Strait of
Caledonian Canal .....56 53 N 5 02 W
Calf of Man, *island* .....54 03 N 4 49 W
Cambrian Mountains .....52 25 N 3 30 W
Campsie Fells, *hills* .....56 01 N 4 18 W
Cardigan Bay .....52 30 N 4 25 W
Carlingford Lough, *bay* .....54 03 N 6 12 W
Carmarthen Bay .....51 38 N 4 33 W
Carron, *river* .....57 22 N 5 31 W
Celtic Sea .....51 00 N 8 00 W
Channel Islands .....49 30 N 2 30 W
Cheviot, *peak* .....55 29 N 2 09 W
Cheviot Hills .....55 23 N 2 20 W
Cleveland Hills .....54 26 N 1 04 W
Clwyd, *river* .....53 12 N 3 21 W
Clyde, *river* .....55 51 N 4 15 W
Clyde, Firth of .....55 38 N 5 01 W
Coll, *island* .....56 37 N 6 34 W
Colonsay, *island* .....56 04 N 6 13 W
Cotswold Hills .....51 40 N 2 10 W
Croob, Slieve, *mountain* .....54 20 N 5 58 W
Cuillin Hills .....57 13 N 6 14 W
Cumbrian Mountains .....54 31 N 5 08 W
Dartmoor Forest, *upland* .....50 37 N 3 57 W
Dartmoor National Park .....50 35 N 3 55 W
Dearg', Beinn, *mountain* .....57 47 N 4 56 W
Dee, *river* .....53 19 N 3 11 W
Dee, *river* .....54 54 N 4 00 W
Dee, *river* .....57 06 N 2 13 W
Derwent, *river* .....53 45 N 0 57 W
Derwent Water, *lake* .....54 35 N 3 09 W
Deveron, *river* .....57 31 N 2 36 W
Don, *river* .....53 28 N 1 18 W
Donard, Slieve, *peak* .....54 11 N 5 55 W
Doon, Loch .....55 15 N 4 22 W
Dornoch Firth .....57 52 N 3 55 W
Dover (Pas de Calais), Strait of .....51 00 N 1 30 E
Dungeness, *cape* .....50 55 N 0 59 E
Dunkery Beacon, *peak* .....51 10 N 3 35 W
Earn, *river* .....56 20 N 3 36 W
Eden, *river* .....54 55 N 2 59 W
Eden Valley .....54 42 N 2 40 W
Eige, Carn, *peak* .....57 17 N 5 07 W
Eigg, *island* .....56 54 N 6 10 W
English Channel (La Manche) .....50 00 N 1 00 W
Ericht, Loch .....56 51 N 4 21 W
Erne, *river* .....54 30 N 8 16 W
Exe, *river* .....50 48 N 3 31 W
Exmoor Forest, *upland* .....51 08 N 3 43 W
Exmoor National Park .....51 10 N 3 45 W
Fair Isle .....59 32 N 1 36 W
Fens, *region* .....52 46 N 0 03 W
Fetlar, *island* .....60 36 N 0 52 W
Flamborough Head, *cape* .....54 07 N 0 05 W
Forth, *river* .....56 05 N 3 46 W
Forth, Firth of .....56 06 N 2 41 W

Foula, *island* .....60 08 N 2 05 W
Foyle, *river* .....54 59 N 7 18 W
Foyle, Lough, *bay* .....55 06 N 7 08 W
Galloway, Mull of, *cape* .....54 38 N 4 53 W
Giant's Causeway, *cape* ..55 15 N 6 30 W
Gigha, *island* .....55 41 N 5 45 W
Gower Peninsula ...51 36 N 4 08 W
Grampian Mountains .....56 48 N 4 21 W
Great Ouse, *river* ..52 21 N 0 03 E
Great Stour, *river* ..51 19 N 1 12 E
Greenwich Meridian 0
Guernsey, *island* ...49 27 N 2 36 W
Harris, *island* .....57 57 N 6 50 N
Hebrides, Sea of the .....57 07 N 6 55 W
Helvellyn, Mount ...54 32 N 3 01 W
Highlands, *region* ..57 30 N 4 30 W
Hirta, see St. Kilda
Holy Island (Ynys Gybi) .....53 17 N 4 37 W
Hoy, *island* .....58 53 N 3 19 W
Humber, *river* .....53 37 N 0 07 E
Humber Bridge ....53 42 N 0 27 W
Inchcape, see Bell Rock
Inner Hebrides, *islands* .....56 35 N 6 40 W
Inner Sound .....57 24 N 5 55 W
Iona, *island* .....56 20 N 6 25 W
Irish Sea .....53 40 N 5 00 W
Islay, *island* .....55 46 N 6 09 W
Jersey, *island* .....49 13 N 2 07 W
Jura, *island* .....55 59 N 5 54 W
Jura, Sound of ....55 59 N 5 45 W
Kebock Head, *cape* .....58 03 N 6 22 W
Kilbrannan Sound .....55 37 N 5 26 W
Kintyre, Mull of, *cape* .....55 18 N 5 47 W
Kintyre Peninsula ..55 36 N 5 37 W
La Manche, see English Channel
Lagan, *river* .....54 37 N 5 54 W
Lake District National Park .....54 30 N 3 10 W
Lammermuir Hills ..55 50 N 2 44 W
Land's End, *cape* ..50 04 N 5 43 W
Lawers, Ben, *mountain* .....56 32 N 4 13 W
Lea (Lee), *river* ....51 34 N 0 04 W
Leeds–Liverpool Canal .....53 59 N 2 10 W
Leven, Loch .....56 12 N 3 23 W
Lewis, Isle of .....58 20 N 6 35 W
Linn of Dee Falls ...56 59 N 3 32 W
Little Minch, *strait* .....57 37 N 6 44 W
Liverpool Bay .....53 27 N 3 09 W
Lizard Point .....49 58 N 5 13 W
Lleyn Peninsula ....52 55 N 4 28 W
Loch a Tuath, see Broad Bay
Lochnagar, Mount .....56 57 N 3 14 W
Lomond, Loch .....56 05 N 4 36 W
Lorn, Firth of .....56 18 N 5 46 W
Lower Lough Erne .....54 30 N 7 50 W
Loyal, Loch .....58 24 N 4 21 W
Luce Bay .....54 45 N 4 50 W
Lundy, *island* .....51 11 N 4 40 W
Lune, *river* .....54 03 N 2 50 W
Macdui, Ben, *peak* .....57 05 N 3 40 W
Mainland, *island* ...60 16 N 1 16 W
Mainland (Pomona), *island* ..59 00 N 3 15 W
Man, Isle of .....54 14 N 4 32 W
Menai Strait .....53 13 N 4 12 W
Merrick, Mount ....55 08 N 4 29 W
Minch, *strait* .....57 45 N 6 30 W
Monadhliath Mountains .....57 10 N 4 13 W
Moor Foot Hills ....55 45 N 3 05 W
Mor (Albyn), Glen ..57 14 N 4 30 W
Moray Firth .....57 37 N 4 01 W
Morecambe Bay ...54 07 N 3 00 W
Mourne Mountains .....54 05 N 6 05 W
Mull, Island of .....56 27 N 5 59 W
Neagh, Lough .....54 37 N 6 25 W
Ness, Loch .....57 17 N 4 29 W
Nevis, Ben, *mount* .....56 48 N 5 00 W
Nith, *river* .....54 57 N 3 33 W
Nithsdale, *valley* ...55 11 N 3 44 W
North Channel, *strait* .....55 15 N 6 00 W
North Downs, *hills* .....51 13 N 0 54 W
North Sea .....55 20 N 2 00 E
North Sound .....59 18 N 2 46 W
North Uist, *island* ..57 36 N 7 18 W
North York Moors, *hills* .....54 25 N 0 45 W
North York Moors National Park .....54 20 N 0 45 W
Northumberland National Park .....55 17 N 2 15 W
Ochil Hills .....56 16 N 3 33 W
Orkney Islands ....59 01 N 3 06 W
Outer Hebrides, *islands* .....57 44 N 7 13 W
Oykel, *river* .....57 58 N 4 43 W
Papa Westray, *island* .....59 21 N 2 53 W
Peak District National Park .....53 21 N 1 50 W
Pembrokeshire Coast National Park .....51 50 N 5 05 W
Pennines, *uplands* .....54 10 N 2 10 W
Pentland Firth .....58 43 N 3 07 W
Pentland Hills .....55 48 N 3 26 W
Plynlimon (Pumlumon Fawr) Uplands .....52 28 N 3 47 W
Pomona, see Mainland
Radnor Forest, *region* .....52 16 N 3 11 W
Rannoch, Loch, *reservoir* .....56 41 N 4 18 W
Reidh, Rubha, *point* .....57 51 N 5 49 W
Rhum, *island* .....57 00 N 6 20 W
Ribble, *river* .....53 45 N 2 43 W
Rona, *island* .....59 07 N 5 49 W
Ross of Mull Peninsula .....56 18 N 6 14 W
St. Abb's Head, *cape* .....55 55 N 2 08 W
St. Austell Bay ....50 19 N 4 43 W
St. Bees Head, *cape* .....54 30 N 3 38 W
St. George's Channel .....52 30 N 5 30 W
St. Kilda (Hirta), *island* .....57 49 N 8 35 W
St. Magnus Bay ...60 24 N 1 32 W
Salisbury Plain ....51 13 N 1 55 W
Sanday, *island* ....59 15 N 2 35 W
Sark, *island* .....49 27 N 2 22 W
Sawel Mountain ....54 49 N 7 02 W
Scafell Pike, *mountain* .....54 27 N 3 13 W
Scilly, Isles of .....49 55 N 6 18 W
Severn, *river* .....51 48 N 2 23 W
Severn Road Bridge .....51 37 N 2 39 W
Sheppey, Isle of ...51 23 N 0 49 E
Shetland Islands ...60 23 N 1 15 W
Shiel, Loch .....56 49 N 5 32 W
Shin, Loch .....58 06 N 4 32 W
Sidlaw Hills .....56 31 N 3 11 W
Sinclair's Bay .....58 30 N 3 05 W
Skye, Island of ....57 18 N 6 15 W
Snowdonia, *peak* .....53 04 N 4 05 W
Snowdonia National Park .....52 50 N 3 50 W
Solent, *channel* ....50 47 N 1 17 W
Solway Firth .....54 44 N 3 35 W
South Downs, *hills* .....50 55 N 0 44 W
South Ronaldsay, *island* .....58 47 N 2 58 W
South Uist, *island* .....57 17 N 7 19 W
Southern Uplands, *region* ...55 30 N 3 30 W
Sperrin Mountains .....54 48 N 7 05 W
Spey, *river* .....57 40 N 3 06 W
Stour, *river* .....51 57 N 1 12 E
Strangford Lough ..54 30 N 5 36 W
Strathmore Valley .....56 39 N 3 00 W
Stronsay, *island* ...59 06 N 2 35 W
Sumburgh Head, *cape* .....59 51 N 1 17 W
Sunart, Loch .....56 41 N 5 43 W
Tamar, *river* .....50 29 N 4 13 W
Taw, *river* .....51 00 N 3 59 W
Tay, *river* .....56 23 N 3 22 W
Tay, Firth of .....56 27 N 2 59 W
Tees, *river* .....54 30 N 1 27 W
Teifi, *river* .....52 05 N 4 40 W
Teviot, *river* .....55 31 N 2 30 W
Thames, *river* .....51 39 N 0 14 E
Tiree, *island* .....56 30 N 6 53 W
Torridge, *river* .....51 01 N 4 12 W
Trent, *river* .....53 41 N 0 42 W
Tummel, Loch .....56 43 N 3 56 W
Tweed, *river* .....55 45 N 2 03 W
Tyne, *river* .....54 59 N 1 46 W
Ullswater, Lake ....54 35 N 2 52 W
Unst, *island* .....60 45 N 0 53 W
Upper Lough Erne .....54 15 N 7 35 W
Wash, *bay* .....52 55 N 0 15 E
Waveney, *river* ....52 28 N 1 37 E
Wear, *river* .....54 51 N 1 33 W
Westray, *island* ....59 18 N 2 59 W
Wharfe, *river* .....53 53 N 1 16 W
Wharfe Dale .....53 54 N 1 27 W
Wight, Isle of .....50 41 N 1 18 W
Windermere, Lake .....54 21 N 2 57 W
Wolds, *upland* .....53 34 N 0 21 W
Wrath, Cape .....58 37 N 5 00 W
Wye, *river* .....51 37 N 2 39 W
Yare, *river* .....52 35 N 1 29 E
Yell, *island* .....60 36 N 1 07 W
Ynys Enlli, see Bardsey Island
Ynys Gybi, see Holy Island
Ynys Môn, see Anglesey
Yorkshire Dales National Park .....54 16 N 2 05 W

Austell (Hensbarrow); Carn Brea; and the spectacular extremity of Land's End. Granite reappears above the sea in the Isles of Scilly, 28 miles further southwest. Despite this variation in general elevation, the landscape, like that of so many other parts of the United Kingdom, has a quite marked uniformity of summit heights, with a high series occurring between 1,000 and 1,400 feet, a middle group between 700 and 1,000 feet, and coastal plateaus ranging between 200 and 400 feet. A network of deep, narrow valleys alternates with flat-topped, steplike areas rising inland. The South West derives much of its renowned physical attraction from its peninsular nature, for, in addition to magnificent drowned estuaries created by sea-level changes, the coastline is unsurpassed for its diversity.

The Lowland zone. Gauged by the 700-foot contour, the Lowland zone starts around the Solway Firth in the northwest, with a strip of low-lying ground extending up the fault-directed Eden Valley. Southward, the narrow coastal plain bordering the Lake District broadens into the flat, glacial-drift-covered Lancashire Plain, with its slow-flowing rivers. East of the Pennine ridge the lowlands are continuous, except for the limestone plateau north of the River Tees and, to the south, the North York Moors, with large, exposed tracts over 1,400 feet. West of the North York Moors lies the wide Vale of York, which the broad lower Trent valley links to the younger rocks of the Midland Plain, terminating against the Welsh Massif on the west. The lowland continues southward along the flat landscapes bordering the lower River Severn, becomes constricted by the complex Bristol–Mendip upland, and opens out once more into the extensive and flat plain of Somerset. The eastern horizon of much of the Midland Plain is the scarp face of the Cotswolds, part of the discontinuous outcrop of limestones and sandstones that arcs from the Dorset coast in southern England, continuing in the Cleveland Hills as far as the coast of North Yorkshire. The more massive limestones and sandstones give rise to noble 1,000-foot-high escarpments, yet the dip slope is frequently of such a low angle that the countryside resembles a dissected plateau, passing gradually on to the clay vales of Oxford, White Horse, Lincoln, and Pickering. The flat, often reclaimed landscapes of the Fens are also underlain by these clays, and the next scarp, the western-facing chalk outcrop (cuesta), undergoes several marked directional changes in the vicinity of the Wash, a shallow arm of the North Sea.

Eastern chalklands and vales

The chalk outcrop is a more conspicuous and continuous feature than its sandstone and limestone predecessor: it begins in the north with the open rolling country known as the Yorkshire Wolds, where heights of 750 feet occur, is breached by the River Humber, and then continues in the Lincolnshire Wolds. East of the Fens the outcrop is very low, barely attaining 150 feet, but it then rises gradually to the 807-foot Ivinghoe Beacon in the attractive Chiltern

Hills. The outcrop is interrupted by several wind gaps, or former river courses, and the Thames actually cuts through it, in the Goring Gap. Where the dip slope of the chalk is almost horizontal, as is the case in the open Salisbury Plain, the landscape is that of a large dissected plateau of 350 to 500 feet. Only the main valleys contain rivers, while the other valleys remain dry.

The chalk outcrop continues into Dorset, but in the south the chalk has been folded along west-to-east lines. Downfolds, subsequently filled in by geologically recent sands and clays, now floor the London and Hampshire basins. The former, an asymmetrical synclinal (or structurally downwarped) lowland rimmed by chalk, is occupied mainly by gravel terraces and valley-side benches and has relatively little floodplain; the latter is similarly cradled by a girdle of chalk, but the southern rim, or monocline, has been cut by the sea in two places to form the scenic Isle of Wight.

Between these two synclinal areas rises the anticlinal, or structurally upwarped, dome of the Weald region of Kent and East Sussex. The arch of this vast geologic upfold has long since been eroded away, and the bounding chalk escarpments of the North and South Downs (uplands so named because of their open, rolling, treeless grassland) are therefore inward-facing and enclose a concentric series of exposed clay vales and sandstone ridges. On the coast the upfold has been eaten into by the waters of the English Channel to produce a dazzling succession of chalk cliffs facing the European mainland, 21 miles distant at the narrowest point.

DRAINAGE

The main watershed in Great Britain runs from north to south, keeping well to the west until the basin of the River Severn. Westward-flowing streams empty into the Atlantic in relatively short distances; and the Clyde in Scotland, the Eden and Mersey in northwestern England, and the Dee, Teifi, and Tywi in Wales are the only significant rivers. The drainage complex that debouches into the Severn estuary covers a large part of Wales and southwestern England. Thereafter, the Avon, flowing through Bristol, and the Parret watershed lie somewhat to the east, but subsequently, with the exception of the Taw–Torridge valleys, they run very close to the western coast in Devon and Cornwall.

Severn estuary

The rivers draining east from the main watershed are longer, several coalescing into wide estuaries. The fast-flowing Spey, Don, Tay, Forth, and Tweed of eastern Scotland run generally across impermeable rocks, and their discharges increase rapidly after rain. From the northern Pennines the Tyne, Wear, and Tees flow independently to the North Sea, but thereafter significant estuary groupings occur. A number of rivers, including the Trent, drain into the Humber after they leave the Pennines. To the south another group of rivers enters the Wash after sluggishly draining a large, flat countryside. The large drainage complex of the Thames dominates southeastern England; its source is in the Cotswolds, and, after being joined by many tributaries as it flows over the Oxford Clay, the mainstream breaches the chalk escarpment in the Goring Gap. A number of tributaries add their discharges farther downstream, and the total drainage converging on the Thames estuary is nearly 4,000 square miles. The important rivers flowing into the English Channel are the Tamar, Exe, Fowey, Avon, Test, Arun, and Ouse. The major rivers in Northern Ireland are the Erne, Foyle, and Bann.

SOILS

The regional pattern of soil formation can be correlated usefully with local variations of relief and climate. Although changes are gradual and can be complicated by local factors, a division of Britain into four climatic regimes (see below *Climate*) goes far to explain the distribution of soils.

At the higher altitudes of the Highland zone, particularly in Scotland, the weather is characterized by a cold, wet regime of more than 40 inches (1,000 millimetres) rainfall and less than 47° F (8° C) mean temperature annually; there are found blanket peat and peaty podzol soils, with their organic surface layer resting on a gray, leached base. A similarly wet regime, but with a mean annual temperature exceeding 47° F, obtains over most of the remainder of the Highland zone, particularly on the lower parts of the Southern Uplands, the Solway Firth–Lake District area, the peripheral plateaus of Wales, and most of southwestern England. These areas are covered by acid brown soils and weakly podzolized associates. On the lower-lying areas within the Highland zone, more particularly in eastern Scotland and the eastern flanks of the Pennines, a relatively cold, dry regime gives rise to soils intermediate between the richer brown earths and the podzols.

Over the whole of the Lowland zone, which also has a mean annual temperature above 47° F but less than 40 inches of rainfall, leached brown soils are characteristic. Calcareous, and thus alkaline, parent materials are widespread, particularly in the southeast, and so acid soils and podzols are confined to the most quartz-laden parent materials. In Northern Ireland at elevations of about 460 feet, brown earths are replaced by semipodzols, and these grade upslope into more intensively leached podzols. This is particularly the case in the Sperrins and the Mournes, but, between them in the Lough Neagh lowland, rich brown earth soils are extensively developed.

CLIMATE

The climate of the United Kingdom is broadly determined by its setting within the pattern of the atmosphere's general circulation and its position in relation to the form and distribution of land and sea. Regional diversity does exist, but the boundaries of major world climatic systems do not pass through the country. Britain's marginal position between the European landmass to the east and the ever-present, relatively warm Atlantic waters to the west ensures the modification of both the thermal and moisture characteristics of the principal types of air reaching the country's shores. These, according to their source regions, are Arctic, polar, and tropical; by their route of travel, the polar and tropical may in each case be maritime or continental. For much of the year the weather is dominated by the sequence of disturbances within the midlatitude westerlies that bring in mostly polar maritime and, occasionally, tropical maritime air. In winter, occasional high-pressure areas to the east allow biting Arctic and polar continental air to sweep over Britain. All of these atmospheric systems tend to fluctuate rapidly in their paths and to vary both in frequency and intensity throughout the seasons of the year and also from year to year for any given season. Variability contributes much to the character of British weather, and extreme conditions, though rare, can be very important for the life of the country.

Air-mass characteristics

The polar maritime winds that reach the country in winter create a temperature distribution that does not reflect latitudinal differences. Thus the north-to-south run of the 40° F (4° C) January isotherm, or line of equal temperature, from the coast in northwestern Scotland south to the Isle of Wight betrays the moderating influence of the winds blowing off the Atlantic Ocean. In summer, polar maritime air is less common, and a 9° difference of latitude and distance from the sea assume more importance, with temperatures increasing from north to south and from the coast inland. Above-average temperatures are usually associated with tropical continental air, particularly in anticyclonic, or high-pressure, conditions. On rare occasions these southerly or southeasterly airstreams can bring to southern England heat waves with temperatures of 90° F (32° C). The mean annual temperature ranges from 46° F (8° C) in the Hebrides to 52° F (11° C) in southwestern England. In spring and autumn a variety of airstreams and temperature conditions may be experienced.

Rainfall patterns

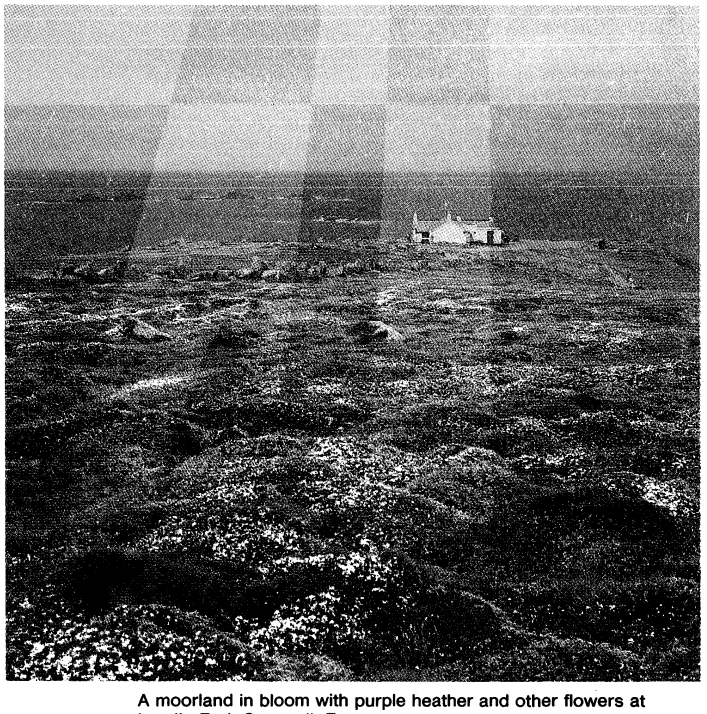
Rain-producing atmospheric systems arrive from a westerly direction, and some of the bleak summits of the highest peaks of the Highland zone can receive as much as 200 inches (5,100 millimetres) of rainfall a year. Norfolk, Suffolk, and the Thames estuary, in contrast, can expect as little as 20 inches (500 millimetres). Rain is fairly well distributed throughout the year: June, on the whole, is the driest month throughout Britain; May is the next driest in the eastern and central parts of England, but April is

drier in parts of the west and north. The wettest months are usually October, December, and August, but in any particular year almost any month can prove to be the wettest, and the association of Britain with seemingly perpetual rainfall (a concept popularly held among foreigners) is based on a germ of truth. Some precipitation falls as snow, which increases with altitude and from southwest to northeast. The average number of days with snow falling can vary from as many as 30 in blizzard-prone northeastern Scotland to as few as five in southwestern England. Average daily hours of sunshine vary from less than three in the extreme northeast to about 4½ along the southeastern coast.

PLANT AND ANIMAL LIFE

Except for northern Scotland, the highest hills of the north and west, the saturated fens and marshes, and the seacoast fringes, the natural vegetation of the British Isles is deciduous summer forest dominated by oak. With human encroachment, only scattered woodlands and areas of wild or seminatural vegetation lie outside the enclosed cultivated fields. Few of these fine moorlands and heathlands, wild though they may appear, can lay claim to any truly natural plant communities: nearly all show varying degrees of adjustment to grazing, swaling (controlled burning), or other activities. Woodland now covers less than 10 percent of the country, and, although the Forestry Commission has been active since its creation in 1919, nearly two-thirds of this woodland remains in private hands. The largest areas of woodland are now to be found in northeastern Scotland, Kielder and other forests in Northumberland, Ashdown Forest in East Sussex, Gwynedd in Wales, and Breckland in Norfolk.

© Hans Hoefer from TSW—CLICK/Chicago



A moorland in bloom with purple heather and other flowers at Land's End, Cornwall, Eng.

Flora of the moorlands and heathlands

The moorlands and heathlands occupy about a fourth of the total area of the United Kingdom. They consist of the possibly true Arctic-Alpine vegetation on some mountain summits in Scotland and the much more extensive peat moss, heather, bilberry, and thin *Molinia* and *Nardus* grass moors of the Highland zone. A similar vegetation exists on high ground in eastern Northern Ireland and on the Mournes, with considerable areas of peat moss vegetation on the Mountains of Antrim. In the Lowland zone, where light sandy soils are found, the vegetation is dominated typically by the common heather—the deep purple of which adds a splash of colour to the autumn countryside—but sometimes by bilberry or bell heather. A strip of land immediately bordering the coastline has also largely escaped the attention of humans and domesticated animals, so that patches of maritime vegetation can frequently be found in approximately their natural state.

The survival of the wild mammals, amphibians, and reptiles of the United Kingdom depends on their ability to come to terms with the changing environment and to protect themselves from attacks by their enemies—the most dangerous being human. British mammals survive in a greater range of habitats than do amphibians and reptiles. Most of the formerly abundant larger mammals—such as boars, reindeer, and wolves—have become extinct, but red deer survive in the Scottish Highlands and in Exmoor Forest, and roe deer in the wooded areas of Scotland and southern England. The carnivores (badgers, otters, foxes, stoats, and weasels) thrive in most rural areas; also widely distributed are the rodents (rats, squirrels, mice) and insectivores (hedgehogs, moles, shrews). Rabbits are widespread, and their numbers are increasing. The other nocturnal vegetarian, the brown hare, is found in open lowland country, while the mountain hare is native to Scotland. Amphibians are represented by three species of newt and five species of frogs and toads, while reptiles consist of three species of snakes, of which only the adder is venomous, and three species of lizard. There are no snakes in Northern Ireland.

In many respects the British Isles are an ornithologist's paradise. The islands lie at the focal point of a migratory network, and the coastal, farmland, and urban habitats for birds are diverse. Some 200 species of birds are found in the United Kingdom, of which more than one-half are migratory. Many species are sufficiently versatile to adapt to changing conditions, and it is estimated that suburban gardens have a higher bird density than any kind of woodland. The most common game birds are the wild pigeon, pheasant, and grouse. Most numerous are the sparrow, blackbird, chaffinch, and starling.

Marshland reclamation has resulted in movement of waterfowl to various bird sanctuaries. A continuous effort on the part of ornithological organizations promotes and encourages research and conservation. As a result, many bird refuges, sanctuaries, and reserves have been created. These developments, along with a more sympathetic and enlightened attitude, may help to redress some of the worst effects of environmental changes on birdlife.

Many of the British rivers, once renowned for their salmon, trout, roach, perch, pike, and grayling, have become polluted, which has led to a decline in inland fisheries. Freshwater fishing is now largely for recreation and sport. The Dogger Bank in the North Sea, one of the richest fishing grounds in the world, has provided excellent fishing for centuries. Other good waters for fishing are those of the Irish Sea and those off the western coast of Scotland. Chief offshore species are cod, haddock, whiting, mackerel, coalfish, turbot, herring, and plaice.

SETTLEMENT PATTERNS

The conditions of regional awareness

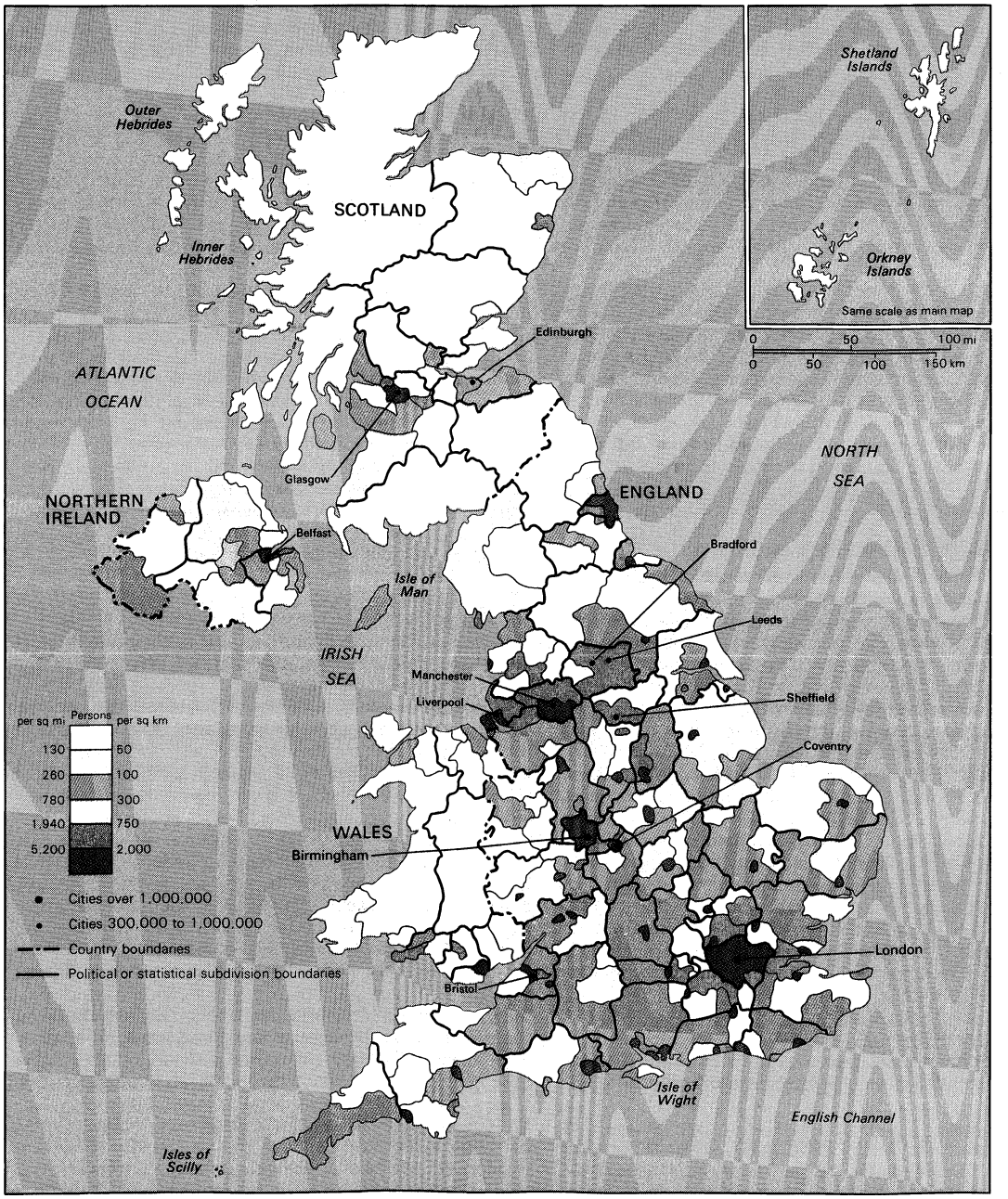
When a long view is taken of the slow possession of Britain by its peoples, it seems an inescapable conclusion that the man–land relationship probably reached its greatest intimacy late in the 18th century: communications at that time were good enough to bind the community into a unity but were not yet so good as to destroy the sense of belonging to a specific locality, as well as to a particular nation—Scotland, England, Wales, or Ireland. Generations of toiling hands had won the necessities of life from the landscape, and generations of tongues had shaped the national languages and dialects for the expression of local things. The regional character of British life is still recognizable, but its heyday has passed. Yet the consciousness of being a Northern Irelander, a Scot, a Welshman, or a Cornishman—to say nothing of the rivalry between a North and South Walian, or a Highland and Lowland Scot—is as marked as is the obvious geographic identity of these parts of the Highland zone.

Rural settlement. The forms and patterns of settlement are remarkably varied in the United Kingdom and reflect not only the physical variety of the landscape but also the successive movements of peoples arriving as settlers,

refugees, conquerors, or traders from continental Europe. The social and economic advantages that led people to cluster and, on the other hand, the equally strong desire for separateness on the part of some individuals are apparent in settlement forms from very early times, and so regional contrasts in the degree of dispersion and nucleation are frequent.

The single farmstead, together with many survivals of the old clachan (cluster or hamlet), interspersed with the occasional village and small town, is still characteristic of much of the Highland zone. Radical alteration has nevertheless occurred in some nucleated settlement patterns: in Wales the breakup of hamlets began in the late Middle Ages as a result of the related processes of consolidation and enclosure that accompanied the decline in the numerical strength of the bond (feudally tied) population. This trend was reinforced by the Black Death of 1349, which spread quickly among the poorer inhabitants. Many surviving bondsmen took advantage of the turmoil caused by the nationalistic uprising led by Owen Glendower (Welsh: Owain Glyndŵr) to escape their servile obligations by flight, and thus many Welsh hamlets fell into decay by 1410, when the rebellion was crushed. In Scotland great changes accompanied the Highland clearances in the mid-18th century; in Northern Ireland, as late as the 1880s, many clachans disappeared as part of a deliberate policy of reallocating land to new dispersed farmsteads. Great changes have also occurred in the Lowland zone, where the swing to individual ownership or tenancy from the medieval custom of landholding in common brought about not only dispersion and deserted villages but also the enclosure of fields by hedges and walls. Nucleations remain remarkably stable features of the rural landscape of Britain, and linear, round, oval, and ring-shaped villages survive, many with their ancient greens still held in common by the community.

Urban settlement. By any standards Britain must be regarded as the most urbanized of countries, for towns are not only particularly expressive of the national way of life but are also unusually significant elements in the geography of the country. The greatest overall change in settlement was, in fact, the massive urbanization that accompanied Britain's early industrial development. The increasing percentage of employees in offices and service industries ensures continued contemporary urban growth. Of every 10 people in the United Kingdom, nine live in



Population density of the United Kingdom.

towns and more than three of them in one of the country's seven conurbations. The Greater London conurbation—the greatest port, the largest centre of industry, the most important centre of office employment, and the capital city—is by far the largest of these. The need for accommodating business premises has involved the displacement of population from inner London, and this development, in part, has led to the designation of New Towns outside the 10-mile-wide Green Belt, which surrounds London's built-up area.

The role of Greater London

Large conurbations have also formed on or near the exposed coalfields. The extensive built-up area of the West Midlands conurbation is dominated by Birmingham, but the industrial Black Country—named for its formerly polluted skies and grimy buildings—also has several large and flourishing towns. In the South East Lancashire conurbation, with a similar number of inhabitants, urbanization accompanied the mechanization of the cotton textile industry. Across the Pennines similar mechanization of wool textiles created the West Yorkshire conurbation, with Leeds and Bradford as its twin centres. The Tyneside and the Central Clydeside conurbations are also located on coalfields, the latter housing about one-third of Scotland's people. Merseyside is not on the Lancashire coalfield, but it has close economic links with it. About one-fifth of Northern Ireland's population live in Belfast. In addition to these large metropolitan areas, there are also a number of other minor urban agglomerations and large towns, a number of which are strung out along the coast.

With so much urban and suburban concentration, the problems of air, water, and noise pollution have become subjects of much concern in the United Kingdom. Considerable progress has been made in controlling air pollution as a result of changes in fuel usage and through clean-air legislation, which has led to the establishment of smoke-control areas in most cities and towns. Pollution of the rivers remains a large problem, particularly in the highly industrialized parts of the United Kingdom, but vigilance, research, and control on the part of the National River Authorities and general public concern for the environment are encouraging features of contemporary Britain. Several statutory and voluntary organizations support measures to protect the environment; they have as their aim the conservation of the natural amenity and beauty not only of the countryside but also of the towns and cities.

The people

ETHNIC COMPOSITION

For centuries people have been going to the British Isles from many parts of the world, some to avoid political or religious persecution, others to find a better way of life or to escape from poverty. The Irish have long made homes in Britain, as have Jews, many of the latter arriving toward the end of the 19th century and in the 1930s. After 1945 large numbers of other European refugees settled in the country. The large communities from the West Indies and South Asian subcontinent date from the 1950s and '60s. There are also sizable groups of Americans, Australians, and Chinese, as well as various other Europeans, such as Greeks, Russians, Poles, Serbs, Estonians, Latvians, Armenians, Turkish Cypriots, Italians, and Spaniards. Since the early 1970s, Ugandan Asians and immigrants from Latin America, Southeast Asia, and Sri Lanka have sought refuge in Britain. Persons of Indian, Pakistani, and Bangladeshi origin account for more than one-half of the total ethnic minority. The foreign-born element of the population is disproportionately concentrated in inner-city areas, particularly in the South East.

20th-century immigration

LINGUISTIC COMPOSITION

All the traditional languages spoken in the United Kingdom are descended from a common Indo-European original, a tongue so ancient that, over the centuries, it has split into a variety of languages, each with its own peculiarities in sounds, grammar, and vocabulary. A separate idiom in what became the United Kingdom was initiated when peoples from the Continent were cut off, in their new homes, from regular intercourse with their continental kindred.

Of the surviving languages the earliest to arrive were the two forms of Celtic: the Goidelic (from which Irish Gaelic, Manx, and Scottish Gaelic are derived) and Brythonic (from which are descended the old Cornish language and modern Welsh). Among the contemporary Celtic languages Welsh is the strongest: about one-fifth of the total population of Wales are able to speak it, and there are extensive interior upland areas and regions facing the Irish Sea where the percentage rises to 50 percent and more. Scottish Gaelic is at its strongest among the inhabitants of the islands of Outer Hebrides and Skye. It is still heard in the nearby North West Highlands. Because less than 2 percent of the Scottish people are able to speak Gaelic, it has long since ceased to be a national language, and even in the northwestern part, where it remains the language of religion, business, and social activity, Gaelic is losing ground. In Northern Ireland very little Gaelic is spoken. Similarly, Manx is now used by very few individuals indeed, although as late as 1870 it was spoken by about half the people of the Isle of Man. Cornish became extinct in the early 18th century.

Germanic language links

The second link with Indo-European is through the ancient Germanic language group, two branches of which, the North Germanic and the West Germanic, were destined to make contributions to the English language. Modern English is derived mainly from the four Germanic dialects spoken by the Angles, Saxons, and Jutes (who all arrived in Britain in the 5th century AD) and by the Danes, whose long series of raids began about 790. The Humber became an important linguistic as well as geographic boundary, and the English-speaking portion of what became England was divided into a Northumbrian and a Southumbrian province (in which the most important kingdoms were Mercia, Wessex, and Kent). In the 8th century Northumbria was foremost in literature and culture, followed for a short time by Mercia; finally, Wessex remained the linguistic centre until the time of King Edward the Confessor.

The Normans, although also of Viking stock, were at first regarded as much more of an alien race than the Danes. Under the Norman and Angevin kings, England formed part of a continental empire, and the prolonged connection with France retained by its new rulers and landlords made a deep impression on the English language. An Anglo-French hybrid speech developed and remained the official language, sometimes even displacing Latin in public documents, until the mid-14th century. Many additions to the English language have been made since that date, but the Normans were the last important linguistic group to enter Britain.

RELIGIONS

The various Christian denominations in the United Kingdom have emerged from the schisms that divided the church. The greatest of these occurred in England in the 16th century, when Henry VIII rejected the absolute supremacy of the pope. This break with Rome facilitated the adoption of some Protestant tenets and became the foundation of the Church of England, still the established church. In Scotland the Reformation gave rise to a church governed by presbyteries—local bodies composed of ministers and elders—rather than by bishops, as was the case in England. Roman Catholicism in Ireland as a whole was almost undisturbed by these events, but what became Northern Ireland came strongly under the influence of the Anglican and Presbyterian churches. In the 17th century further schisms divided the Church of England; these were associated with the rise of the Puritan movement, which, with its desire for simpler forms of worship and government, led to a proliferation of nonconformist churches, such as those of the Baptists and the Congregationalists. The Society of Friends (Quakers) also originated at that time. Religious revivals of the mid-18th century gave to Wales a form of Protestantism closely linked with the Welsh language; Calvinistic Methodism is still the most powerful religious influence in the principality. The great Evangelical revivals of the 18th century, associated with John Wesley and others, led to the foundation of Methodist churches, particularly in the industrial areas; Northum-

The established church

berland, Durham, North Yorkshire, and Humberside in northeastern England and Cornwall in the southwestern peninsula still have the largest percentages of adherents to this denomination. In the 19th century the Salvation Army and various fundamentalist sects grew from minor schisms. The 19th century also saw the introduction of sects from the United States as well as a marked increase in the number of Jews in Britain. The first Jewish community in Britain after their expulsion in 1290 was that established in London during the 17th century, and in the 19th century Jews also settled in many of the large provincial cities. More than half of all British Jews live in London, and the rest are essentially members of urban communities. Jewish congregations in Britain now form the second largest Jewish community in Europe.

Britain has a long tradition of religious tolerance, a feature that has been particularly advantageous since the 1950s, when a large variety of religious beliefs began to be brought in by immigrants. There are large and growing communities of Muslims, Hindus, and Sikhs. The largest number of Muslims came from Pakistan and Bangladesh, with sizable groups from India, Cyprus, the Arab world, Malaysia, and parts of Africa. The large Sikh and Hindu communities originated in India, the Sikhs being the larger group of the two, and there are also many Buddhist groups.

DEMOGRAPHIC TRENDS

Population growth. Since the making in 1086 of the Domesday Survey, a detailed compilation that enables the earliest reasonable estimate of England's population to be made (the survey did not cover other areas), the number of people has been increasing. This growth has continued despite some setbacks, by far the most serious of which was the Black Death in the mid-14th century, in which it is estimated that about one-third of the population died. There is little concrete information, however, concerning variation in rates of births to deaths and of immigration to emigration until 1801, when the first official census was taken. The assumption is that a population of about two million lived in what became the United Kingdom at the end of the 11th century and that this figure had increased to about 12 million by 1801. This slow growth rate, in contrast with that of more modern times, resulted mainly from the fact that a high birth rate was accompanied by an almost equally high death rate. Family monuments in old churches show many examples of men whose "quivers were full" but whose hearths were not crowded. It is estimated that in the first half of the 18th century three-quarters of the children born in London died before they reached puberty. Despite the appalling living conditions it produced, the Industrial Revolution resulted in an acceleration of the birth rate; gradually the greater medical knowledge, improved nutrition, and concern for public health that characterized the 19th and 20th centuries bore fruit in a lower mortality rate and an overall increase in population.

First official census

Since the 1930s the population has experienced a complete cycle in its pattern of growth. A low rate of increase during the 1930s was followed by a postwar marriage boom that gave rise to an acceleration in the rate of growth, culminating in a peak during the mid-1960s. Since 1964 a considerable fall in the birth rate has brought about a dramatic decline in growth, even giving rise to an absolute decline between 1974 and 1978. Although the main cause of these abrupt reverses was the erratic nature of the birth rate, it must nevertheless be associated with a general decline in fertility, plus a rising longevity and a reduction in mortality. Such processes have also affected the age composition of the population, which has assumed a decidedly older cohort. There has been a decline in the proportion of those under 16 and an increase in the proportion of older people, especially those aged 85 and over.

Migration patterns. After 1957 the immigration of non-white ("New Commonwealth") people from such developing nations as India, Pakistan, and the countries of the West Indies became significant, and from that year until 1962 there was a net migration gain. Since then restriction on the entry of New Commonwealth citizens has lessened the primary inflow, but dependents of immigrants already in the United Kingdom are still admitted. The reasons for restricting entry were in part economic but were also associated with the problems of assimilating so many peoples of different cultures. Nevertheless, the United Kingdom continues to be a net importer of people from the New Commonwealth.

"New Commonwealth" immigrants

Although historical records make reference to emigration to North America in the 17th and 18th centuries, there is little quantitative information about such movements before the middle of the following century. The greatest numbers appear to have left Great Britain in the 1880s and between 1900 and the outbreak of World War I. Emigration, particularly to Canada and Australia ("Old Commonwealth"), continued at a high rate after the war until 1930, when unfavourable economic conditions in the British Empire and in the United States reversed the movement. During the same years there also was an influx of refugees from Europe. After World War II both inward and outward movements reached considerable dimensions. Emigration to the countries of the Old Commonwealth and, to a lesser degree, to the United States continued, but until 1951 the net migration balance of the United Kingdom with the rest of the world was not large. Since the mid-1960s there has been a slackening of emigration as Canada and Australia no longer maintained an open-door policy to citizens of the United Kingdom, accepting only those whose skills were in demand. Nevertheless, the United Kingdom continues to be an exporter of population, albeit on a declining scale, to the Old Commonwealth, but this is offset by emigration to the European Communities nations and other foreign countries.

Migration within the United Kingdom has at times reached sizable dimensions. Until 1700 the small population was sparsely distributed and largely rural and agricultural, much as it had been in medieval times. From the mid-18th century, scientific and technological innovations created the first modern industrial state, while, at the same time, agriculture was undergoing technical and tenurial changes and revolutionary improvements in transport made easier the movement of materials and people. As a result, by the first decade of the 19th century, a previously mainly rural population had been largely replaced by a nation made up of industrial workers and town dwellers.

The rural exodus was a long process. The breakdown of communal farming started before the 14th century; and subsequently enclosures advanced steadily, especially after 1740, until a century later open fields had virtually disappeared from the landscape. Many of the landless agricultural labourers so displaced were attracted to the better opportunities for employment and the higher wage levels existing in the growing industries; their movements, together with those of the surplus population produced by the contemporary rapid rise in the birth rate, resulted in a high volume of internal migration that took the form of a movement toward the towns.

Industry, as well as the urban centres that inevitably grew up around it, was increasingly located near the coalfields, while the railway network, which grew rapidly after 1830, enhanced the commercial importance of many towns. The migration of people, especially young people, from the country to industrialized towns took place at an unprecedented rate in the early railway age, and such movements were relatively confined geographically. Migration from agricultural Ireland provided an exception, for, when the disastrous potato disease of 1845–46 led to widespread famine, large numbers moved to Britain to become the urban workers of Lancashire, Clydeside, and London. The rural exodus continued, but on a greatly reduced scale, after 1901.

Soon after World War I, new interregional migration flows commenced when the formerly booming 19th-century industrial and mining districts lost much of their economic momentum. Declining or stagnating heavy industry in Clydeside, northeastern England, South Wales, and parts of Lancashire and Yorkshire swelled the ranks of the unemployed, and the consequent outward migration became the drift to the relatively more prosperous Midlands and southern England. This movement of peo-

The drift south

ple continued until it was arrested by the relatively full employment conditions that obtained soon after the outbreak of World War II.

In the 1950s, opportunities for employment in the United Kingdom improved with government-sponsored diversification of industry, and this did much to reduce the magnitude of the prewar drift to the south. The decline of certain northern industries—coal mining, shipbuilding, and cotton textiles in particular—had nevertheless reached a critical level by the late 1960s, and the emergence of new growth points in the West Midlands and southeastern England made the drift to the south a continuing feature of British economic life. Subsequently, the area of most rapid growth shifted to East Anglia, the South West, and the East Midlands. This particular spatial emphasis resulted from the deliberately planned movement of people to the New Towns in order to relieve the congestion around London. (W.Ra.)

The economy

The United Kingdom has a fiercely independent, developed, and international trading economy that was at the forefront of the 19th-century Industrial Revolution. The country emerged from World War II as a military victor but with a debilitated manufacturing industry. Postwar recovery was relatively slow, and it took nearly 40 years, with additional stimulation from joining the European Economic Community (now the European Communities, or EC) in 1973, for the U.K. economy to improve its competitiveness significantly. The United Kingdom now ranks among the top industrial countries in growth rates, productivity, and competitiveness. The share of manufacturing's contribution to gross domestic product has declined to about one-fourth of the total, with services providing the source of greatest growth. The United States is a major influence and trading partner, and Japan has become a significant investor in local production, Japanese companies often choosing the United Kingdom as their European base. In addition, other fast-developing East Asian countries with export-oriented economies include the United Kingdom's open market among their important outlets. The United Kingdom's trading pattern has thus switched from being dominantly Commonwealth-linked to one in which just over half its visible trade is with the EC.

The 1980s, during which the Conservative Party governed, saw an acceleration of privatization, or denationalization, of publicly owned corporations that had been nationalized by previous governments. There also was an improvement in the standard of living, though there was, in general, greater prosperity in the South East, which includes London, than in the more northerly and westerly regions. Unemployment and inflation rates were gradually reduced but remained high. The country's role as a major world financial centre and its discovery of natural gas in 1965 and oil in 1969 in the North Sea and their commercial exploitation (beginning in 1967 and 1975, respectively) reduced the dependence on more traditional sources of energy and were major influences on the health of the internal economy and on national economic policies.

National economic policies

Prevailing government policies include the close monitoring and frequent adjustment of interest rates; a gradual reduction in the level of direct personal taxation; in favour of local bargaining, a reduction in the levels of power and influence of national trade unions in national labour negotiations; the encouragement of wider home ownership and of individuals' share holdings in companies; and the recognition of the need for increased provision of training, both for those entering employment for the first time and for those seeking to change the course of their work. Considerable emphasis is placed on improvement in productivity and competitiveness and on increased exposure of the economy to market forces.

Just under half of the total population is in the labour force. Within this group are small numbers of self-employed workers, as well as members of the armed forces and of work-related government training programs. The highest proportion of employees (more than two-thirds) are in the service sectors, with financial services and distribution the largest. Manufacturing, although it has declined, employs more than one-fifth of all workers. Smaller numbers are in construction, energy, agriculture, forestry, and fishing. The number of part-time workers has increased considerably.

RESOURCES

The range of mineral resources of economic value in the United Kingdom is relatively limited. The once-important extraction of iron ore has dwindled almost to nothing. Other metals of importance that are mined include tin, which supplies about half of the domestic demand, and zinc. There are adequate supplies of nonmetallic minerals, including sand and gravel, limestone, dolomite, chalk, slate, barite, talc, clay and clay shale, kaolin (china clay), ball clay, fuller's earth, celestite, and gypsum. Sand, gravel, limestone, and other crushed rocks are quarried for use in construction.

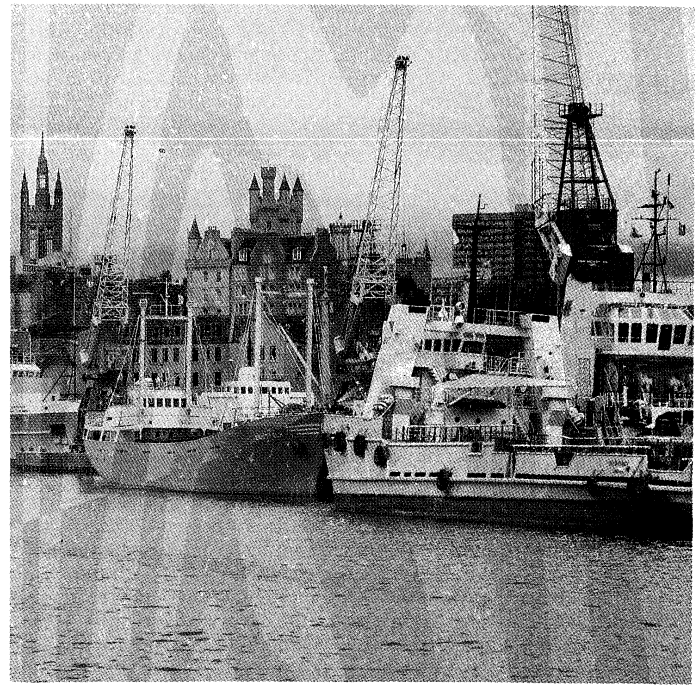
By contrast, the United Kingdom has larger energy resources—including oil, natural gas, and coal—than has any other EC member. Coal, the fuel once vital to the British economy, has continued to decrease in importance. Compared with its peak year of 1913, when production was more than 300 million tons produced by over one million workers, current output has fallen by two-thirds, with an even greater reduction in the labour force. Power stations are the major customers for coal, but, with growth in the use of competitive fuels and the increasing number of pits that have become uneconomic to operate, the industry remains under considerable pressure.

North Sea oil

Discovery of oil in the North Sea, and the apportionment of blocks of its area to several adjoining countries, led to the rapid development of oil exploitation. From the start of production in 1975, the quantities brought ashore have grown each year, making the United Kingdom virtually self-sufficient in oil and even an exporter. With an average output of 2.6 million barrels per day, the country is the world's sixth largest producer. The balance of payments has benefited considerably from oil revenues, and a substantial proportion has been invested abroad, with the income generated to offset diminishing oil income in the future. Proven reserves are estimated at 770 million tons.

Since offshore natural gas supplies from the North Sea began to be available in quantity from 1967, they have displaced the previously coal-based supplies of town gas. A national network of distribution pipelines has been cre-

Milt and Joan Mann—CAMERAMANN INTERNATIONAL



Ships serving North Sea oil platforms at dock in the port of Aberdeen, Scot.

ated. Proven reserves of natural gas are estimated at 22.7 trillion cubic feet.

AGRICULTURE, FORESTRY, AND FISHING

Agriculture. The United Kingdom is unusual among western European and other industrial countries in having only a small proportion of its employed population (less than 2 percent) in agriculture. With commercial intensification of yields and a high level of mechanization, supported initially by national policy and subsequently by the EC's Common Agricultural Policy (CAP), the output of some agricultural products has exceeded demand. Employment in agriculture has declined gradually, and, with a reversal of policies to achieve reduction of surpluses, the trend is likely to continue. To create alternative employment opportunities in rural areas, some of which are remote from towns, efforts have been made to diversify into other activities. The land area used for agriculture (about three-quarters of the total) has also declined, with the arable share falling in favour of grazing.

Official policy in agriculture is undertaken under the aegis of the CAP. The policy's aims have been to improve agricultural productivity in order to ensure stable agricultural markets while simultaneously affording producers a fair standard of living and guaranteeing consumers regular food supplies at reasonable prices. To achieve these aims, a system of minimum prices for domestic goods and levies on imports, to support lower domestic prices, has been in operation. Exports are encouraged by refunds that close the price gap between the EC and the rest of the world. For a few products, particularly beef and sheep, there are additional payments made directly to producers. More recent policies have included milk quotas, land set-asides (to compensate farmers for taking land out of agricultural use), and reliance on the price mechanism as a regulator.

Principal crops

The most important farm crops are wheat, barley, oats, sugar beets, potatoes, and, more recently, rapeseed. While significant proportions of wheat, barley, and rapeseed are turned into animal feed, much of the remainder is used for human consumption through flour milling (wheat), malting and distilling (barley), and vegetable oil (rapeseed). The main livestock products are derived from cattle and calves, sheep and lambs, pigs, and poultry. A high level of self-sufficiency has been achieved in most of the main agricultural products, although less so in sugar and cheese.

Forestry. More than 8 percent of the United Kingdom's land area is devoted to productive forestry. The government-supported Forestry Commission manages almost half of these woodlands, the rest being in private hands. Domestic timber production supplies only about 12 percent of the United Kingdom's requirements. The majority of new plantings are of conifers in upland areas, but encouragement is also given to planting broad-leaved trees where appropriate.

Fishing. Although the United Kingdom is one of Europe's leading fishing countries and is two-thirds self-sufficient in this sector, the industry has been in long-term decline. Fishing limits were extended to 200 miles offshore in the mid-1970s, and, because a significant part of the area fished by other EC members lies within British waters, it has been necessary to regulate catches on a Community-wide basis. Conversely, the United Kingdom has lost opportunities to fish in some more-distant waters (*e.g.,* those off Iceland), which has reduced its total catch more than that of other EC nations. The most important fish landed are cod, haddock, mackerel, whiting, and plaice, as well as shellfish (including nephrops [Norway lobsters], lobsters, crabs, and oysters). There has been significant development of inshore fish farming, mainly of trout and salmon.

INDUSTRY

As noted above, coal mining has declined and the energy sector has been transformed by virtue of self-sufficiency in oil and natural gas. Nuclear fuel has slightly expanded its contribution to electricity generation, while hydroelectric power contributes a small proportion, particularly of importance in Scotland. The bulk of electricity is produced in conventional steam power stations.

Manufacturing. The manufacturing industry as a whole has continued to shrink both in employment and in its contribution (now less than one-fourth) to the gross domestic product (GDP). This part of the economy was the major contributor to the rapid rise in unemployment in the early 1980s. Once economic growth returned, however, there was great improvement in productivity and profits and in expansion in the manufacturing labour force. In terms of their relative importance to the GDP, the most important manufacturing industries are engineering; food, beverages (including alcoholic beverages), and tobacco; chemicals; paper, printing, and publishing; metals and minerals; and textiles, clothing, footwear, and leather. The fastest growing sectors have been metals, chemicals, man-made fibres, timber, furniture, rubber, and plastics. Within the chemical industry, pharmaceuticals and specialty products have shown the largest increases. Food, beverages, tobacco, textiles, clothing, footwear, leather, and engineering as a whole have had below-average growth. Within the engineering industry, electrical and instrument engineering and transport engineering—including motor vehicles and aerospace equipment—have grown faster than mechanical engineering and metal goods, with electronic products showing the fastest growth.

Major manufactures

The 1980s witnessed considerable increases in imports of manufactured products, and, though exports of finished manufactured products increased by value, the surplus in the balance of trade disappeared and was transformed into a large deficit. Nevertheless, the maintenance of a competitive manufacturing base is regarded as of prime importance for international trade, where it has greater potential than the service sectors.

Construction. Construction has grown at a faster rate than manufacturing. About one-third of the labour force in construction is self-employed. More than half of all construction work is on new projects, the remainder on repair and maintenance. Housing represents more than two-fifths of the value of construction, with repair and maintenance absorbing a greater proportion of resources than new building. There has been a marked switch away from housing funded and owned by public authorities toward private development. Considerable efforts have also been made to encourage tenants of publicly owned rented houses to become owner-occupiers, with the result that the proportion of these grew from one-half in the early 1970s to two-thirds by the end of the 1980s. The supply of privately rented accommodations became scarcer because of statutory rent controls that discouraged new construction, but changes both in the economic climate and in official policy began to stimulate the supply. The average price of a new house, particularly in London and the South East, has continued to increase more rapidly than the prevailing rate of inflation. In turn, this phenomenon has created considerable pressures on the availability of land for housing, which has been relatively tightly controlled. Here, too, public policy has been changing in favour of greater permissiveness.

The remainder of construction is taken up by new private industrial and commercial work and other public building and by the repair and maintenance of existing public and private works.

Services. The most remarkable phenomenon characterizing the economy of the United Kingdom has been the growth of service industries. This reflects the rise in real personal incomes and the way in which patterns of consumer expenditure have changed. The providers of personal, financial, and leisure services have been the beneficiaries. Although some services—for example, public transport, laundries, and cinemas—have declined in favour of privately owned goods—such as automobiles, washing machines, and television sets—this has stimulated increased demand for the related services that distribute, maintain, and repair such products. Other types of services for which demand has accelerated include hotels and catering, air travel and other leisure-related activities, distribution (particularly retailing), and finance. A multiplicity of other business-support services has become more prominent, *e.g.,* computing systems and software, management consultancy, advertising, and market research, as

Growth of services

well as the provision of exhibition and conference facilities. Britain is also the base for some of the world's leading art auction houses.

FINANCE

While the United Kingdom has traditionally been a focus of financial activity worldwide, the 1980s saw considerable changes in the structure, composition, and regulation of financial institutions. These embrace banking, insurance, building societies, the Stock Exchange, shipping, and commodity markets. Some long-standing distinctions in specialization have become less clear-cut; for example, housing loans used to be primarily the responsibility of building societies, but increasingly banks and insurance companies have been attracted into this area of lending. Two related developments have occurred: the transformation of building-society branch offices into virtual banks with personal cashing facilities, and the diversification of all three of these types of institutions into real estate services. Building societies also participate to a limited extent in investment services, insurance, trusteeship, executorship, and land services.

London has continued to grow in size and influence as a centre of international financial operations. Capital flows have increased, as have foreign exchange and securities trading; in consequence, a large number of foreign banks are represented there. Increased competition and developments in technology have accelerated change—*e.g.,* the International Stock Exchange was reorganized, and the historical two-tier structure of brokers, who were formerly limited to executing investors' instructions to buy and sell stocks and shares, and jobbers, who "made" markets in these securities, was abolished. As a result, new companies have been created to link British and foreign banks with former brokers and jobbers. The Financial Services Act of 1986, the Building Societies Act of 1987, and the Banking Act of 1987 regulate these new financial organizations. New supervisory bodies have been set up to police the new circumstances and to ensure that the rules of conduct determined by the acts are observed.

Bank of England

All commercial banks are supervised by the Bank of England, which also has the sole right to issue bank notes in England and Wales (banks in Scotland and Northern Ireland have limited rights to do this in their own areas). The Bank of England licenses retail banks (those with a significant branch network), merchant banks, discount houses, and other British or foreign banks. The dividing lines in this sector have also become blurred, the major retail banks diversifying into merchant banking, mortgage banking, insurance, securities dealing, unit trusts (mutual funds), leasing, and factoring. The Bank of England also controls the Minimum Lending Rate, which influences the general structure of interest rates. It intervenes actively in foreign exchange markets and acts as banker to the Sterling Area and as the government's banker. Pound sterling is a major internationally traded currency, and London one of the world's important trading centres.

Saving by individuals is channeled into investments through a variety of institutions. Examples are insurance companies, pension funds, and investment and unit trusts. Other institutions specialize in particular types of financing; thus, finance houses provide home mortgages and corporate lending and leasing. There are also companies that finance the leasing of business equipment; factoring companies that provide immediate cash to creditors and subsequently collect the corporate debts owed; and finance corporations that supplement the medium- and long-term capital markets, otherwise supplied by the banks or the Stock Market, including the so-called venture capital funding of innovations or of high-risk companies.

The United Kingdom has a number of organized financial markets. The securities markets consist of the International Stock Exchange, which deals in officially listed stocks and shares (including government issues, traded options, stock index options, and currency options); the Unlisted Securities Market, for smaller companies; and a Third Market, for small unlisted companies. Money market activities include the trading of bills, certificates of deposit, short-term deposits, and, increasingly, sterling commercial paper. Other markets are those dealing in Eurocurrency, Eurobonds, foreign exchange, financial futures, gold, ship brokerage, freight futures, and agricultural and other commodity futures.

The share of invisible trade (receipts and payments from financial services; interest, profits, and dividends; and transfers between the United Kingdom and other countries) has been rising steadily, from about one-third to one-half of the country's total foreign earnings. Within this area, transactions in services have grown rapidly, with financial services growing the fastest.

TRADE

Because of the limitations of its natural resources and the increasing demands of rising living standards, the United Kingdom has been spurred to greater efforts to export more. At the same time, imports supply a growing share of domestic needs. Products that feature prominently in British exports are machinery, automobiles and other transport equipment, computers, aerospace equipment, electrical and electronics goods, and oil. The country imports one-tenth of its foodstuffs and about one-third of its machinery and transport equipment.

Imports and exports

An increasing share of the United Kingdom's trade is with other developed countries. Joining the EC has caused a major reorientation of trade flows; half of all trade is with European partners, though the United States remains the United Kingdom's single largest export market and its second largest supplier. West Germany is the leading supplier and the second most important export market.

The United Kingdom's balance of payments, which was favourable during the early 1980s, has since been transformed into deficit because of the acceleration in visible imports compared with visible exports. There have been no controls on the outflow of funds since 1979, and consequently there has been considerable overseas investment, with growth in related earnings. It has been official policy to support trade liberalization, and every opportunity is taken to participate in international organizations that endorse such measures.

TRANSPORTATION

The United Kingdom, which is relatively small in area and has a fairly high population density, has undergone considerable change in its patterns of transport. The growth of automobile ownership (nearly two-thirds of all households have one automobile, and some have two or more), the decline in the use of local buses, and the transfer of much internal freight from rail to road has emphasized the importance attached to maintaining and developing road networks, particularly motorways and trunk roads. Intercity rail services have been improved, as have commuter services around major conurbations. Similarly, air traffic has grown, particularly international flights. Seven airports (Heathrow, Gatwick, Stansted, Glasgow, Edinburgh, Prestwick, and Aberdeen) are owned and managed by the British Airports Authority, and together they handle about three-fourths of the passenger air traffic in the country. Although there has been a downward trend in shipping and sea travel, the majority of overseas trade is still carried by sea. Substantial traffic of both passengers and cargo is handled between the ports of the United Kingdom, Ireland, and Europe. Oil and natural gas, which each share a national bulk distribution pipeline system, are independent of the road and rail networks.

It is arguable whether investment in transportation is adequate in relation to demand—for example, the M25 motorway around London showed signs of overload soon after it was opened in 1986; there is overcrowding on commuter services, both rail and underground; congested traffic moves at a snail's pace in cities; and there is continuous pressure to build more motorways and airports to serve London. (U.M.S.)

Administration and social conditions

GOVERNMENT

Central government. The United Kingdom is a constitutional monarchy and a parliamentary democracy. The

country's head of state is the reigning king or queen. The head of government is the prime minister, who is the leader of the political party that can command a majority of votes in the House of Commons.

Sources of the constitution

The British constitution is partly unwritten and is flexible. Its basic sources are parliamentary legislation and decisions by courts of law. Matters for which there is no formal law, such as the resignation of office by a government, are determined by important conventions but are always open to development or modification.

The main elements of the government are the legislature, the executive, and the judiciary, with some overlap since there is no formal separation of powers. Sovereignty resides in Parliament, which consists of the monarch, the hereditary and appointive House of Lords, and the elected House of Commons. The sovereignty of Parliament is expressed in its legislative enactments, which are binding on all, though individuals may contest in the courts the legality of any action under a specific statute. In addition, individuals may in certain circumstances seek protection under European law.

All political power is concentrated in the prime minister and the Cabinet, and the monarch must act on their advice. Members of the Cabinet are chosen by the prime minister from his or her political party in Parliament. Most Cabinet ministers are heads of government departments. The prime minister's authority has increased over the 20th century, and some decisions previously made by the Cabinet as a whole have increasingly been made by the prime minister, alone or with one or two colleagues. A prime minister has nevertheless been overruled by the Cabinet on many occasions and must generally have its support to exercise his or her powers.

Because it is supported by the party with a majority in the House of Commons, the Cabinet exercises the sovereignty of Parliament. The royal right of veto has not been exercised since the 18th century, and the legislative power of the House of Lords was reduced in 1911 to the right to delay legislation. The Cabinet plans and lays before Parliament all important bills. While the Cabinet thus controls the lawmaking machinery, it is also subject to Parliament; it must expound and defend its policy in debate, and its continuation in office depends on its retaining the confidence of the House of Commons.

Administrative decisions of the Cabinet are carried out by the executive apparatus. The Cabinet secretariat, which developed after World War I, converts the Cabinet's decisions into executive action. It prepares the Cabinet's agenda and records its conclusions in a form that can speedily be put into effect.

Within the United Kingdom, Scotland has a distinct legal system based on Roman law. Both Scotland and Wales regard themselves as nations on a par with England, and each is the responsibility of a Cabinet minister. Until the early 1970s Northern Ireland had a Parliament and a Cabinet with autonomy in domestic affairs. This arrangement was abolished by the Northern Ireland Constitution Act of 1973. Since 1974 Northern Ireland has been subject to direct rule by the British secretary of state for Northern Ireland.

Local government. Local government was first established on a uniform basis between 1835 and 1888 and remained broadly unchanged until the present system was adopted between 1965 and 1975. London's local government was reorganized in 1965. In 1974 there was a radical change in the numbers, structure, and functions of local authorities in England and Wales (and in 1975 in Scotland), leading to a two-tier system of 47 counties (and nine regions in Scotland) and 390 lower-tier authorities. Within that structure, six of the counties covering large conurbations were designated metropolitan counties, which were subdivided into metropolitan districts. In 1986 the metropolitan counties and the Greater London Council were abolished and their responsibilities handed over to the lower tier, to specially constituted boards, or to the central government.

Organization and powers

Historically, local government had considerable powers, though they have declined since the 1930s. Central governments over the years have increasingly taken local powers. Until 1990 (1989 in Scotland), spending by local authorities in the United Kingdom was financed by a combination of a local property tax (the rates), government grants, and charges for services. In 1990 (1989 in Scotland), domestic rates were replaced by a flat-rate poll tax known as the community charge.

POLITICS

Elections. The franchise has been progressively widened since the early 19th century; the vote is now given to all citizens at age 18. Citizens vote in parliamentary and local elections and also in elections to the European Parliament. All other public posts are filled by appointment. Each member of the House of Commons represents one parliamentary constituency. Constituency boundaries are drawn so that each contains roughly the same number of voters.

Registration of electors is compulsory and carried out annually. Candidates for election to Parliament or a council are normally chosen by the local parties. There are no primary elections along U.S. lines, for example, nor would such a system be easy, because the timing of general elections is unpredictable.

The House of Commons is elected for a maximum term of five years, reduced in 1911 from seven. At any time during those five years the prime minister has the right to request the monarch to dissolve Parliament and call a general election. A government with a working majority is expected to govern for the greater part of its term, though it rarely runs to the end. An early election may take place if there is no working majority, and only three weeks' notice need be given of a general election. The expenses of parliamentary candidates are strictly limited. These provisions and uncertainty about the timing of an election produce inexpensive campaigns of unusual brevity.

Political parties. A two-party system has existed in the United Kingdom since the late 17th century. From the mid-1920s the dominant groupings have been the Conservative Party and the Labour Party, despite the existence of smaller parties such as the Social and Liberal Democratic Party (which includes the rump of the former Liberal Party) and the Social Democratic Party.

The two-party system is one of the outstanding features of British politics and has generally produced firm and decisive government. The practice of simple majority voting and the establishment of single-member constituencies of uniform population size have tended to exaggerate the majority of the winning party and thus to eliminate third parties.

The Opposition

The two-party system, together with uncertainty about the timing of a general election, has produced the British phenomenon of the Opposition. Its decisive characteristic is that it attempts to form an alternative government, ready at any time to take office, in recognition of which the leader of the Opposition is paid an official salary.

JUSTICE

Judges in the United Kingdom are appointed and virtually irremovable; there is no election to judicial office. Judges are recruited from successful practicing lawyers. The courts alone declare the law, but any act of Parliament is accepted by the courts as part of the law. No court can declare a statute invalid.

An accused person is presumed innocent until proved guilty. The courts strictly enforce the law of contempt to prevent newspapers or television from prejudicing the trial of the accused before a jury. Since 1967, jurors have been able to give a majority verdict in criminal cases (in Scotland a simple majority; in England and Wales with no more than two dissenting votes). Hanging for murder was abolished in 1965. Some 98 percent of defendants in criminal cases in the Crown Courts (in Scotland the High Court of Justiciary), which deal with all serious cases, are granted publicly funded legal aid.

More than 90 percent of criminal cases in England and Wales are tried and determined by some 28,000 justices of the peace, who are unpaid laymen, or by the more than 60 stipendiary (paid) magistrates, who are trained lawyers. More serious crimes also come initially before a magistrates' court. The system is similar in Northern Ire-

land, but, in Scotland, District and Sheriff courts try most criminal cases. The police must bring an arrested person before a magistrate within 36 hours, but the magistrate can authorize further detention without charge for up to 96 hours. Only 1 percent of suspects are held without charge for more than 24 hours, however. The magistrate decides whether the accused should be held on bail or in custody.

The vast majority of civil actions in England, Wales, and Northern Ireland are tried in local county courts, whose jurisdiction is limited by the nature of the action and the amount of money at stake. In Scotland, Sheriff Courts and the Court of Session try all civil actions.

Appeals in civil and criminal matters lie from the High and Crown courts to the Court of Appeal, which, in cases of legal importance, can allow a final appeal to the judges in the House of Lords. In Scotland only civil matters may be appealed to the House of Lords.

POLICE SERVICES

British police

Britain has no national police force nor any minister exclusively responsible for the police. Each provincial force is maintained by a police authority, a committee elected by a number of local authorities. One of its important tasks is to appoint and dismiss the chief constable. Once appointed, the chief constable has the sole right of appointment, promotion, discipline, and deployment of his force.

The Home Secretary is the police authority for the Metropolitan Police in London. The commissioner of the Metropolitan Police has status similar to that of a chief constable. Scotland Yard (the criminal investigation department of the Metropolitan Police) assists other police forces and handles the British business of the International Criminal Police Organization (Interpol).

The British police wear a uniform that is nonmilitary in appearance. Their only regular weapon is a short wooden truncheon, which is kept out of sight and may not be employed except in self-defense or to restore order. Police engaged on dangerous missions may carry firearms for that specific occasion.

THE ARMED FORCES

Responsibility for national defense rests with the prime minister and the Cabinet. The secretary of state for defense formulates defense policy. His ministry, created in 1970 by a merger of the three original service ministries (Admiralty, War, and Air), has responsibility for the armed forces. The secretary of state is advised by the chief of the defense staff aided by the three service chiefs.

EDUCATION

Primary and secondary education. The modern structure of preuniversity education was set out in the Education Act of 1944. The responsibility for education in England rests with the secretary of state for education, who is accountable to Parliament and is responsible for the Department of Education and Science. The secretaries for Wales, Scotland, and Northern Ireland are responsible for preuniversity education in their respective countries. Under the 1944 act, primary and secondary education became a local responsibility. Local Education Authorities (LEAs) employ the teachers and are the major providers of education. In addition, a few schools are run by voluntary bodies, mostly religious. There is also a small private sector.

Primary education is free and compulsory from age five to 11. Secondary education, which is organized in a variety of ways, is provided by LEAs for children aged 11 to 19; it is free and compulsory to age 16 (the minimum leaving age being raised from 15 in 1972). Teachers employed by the LEAs are paid on an agreed national scale. Primary and secondary education is financed almost entirely by the state out of central and local tax revenues. Most expenditures take place at the local level, though about half of local revenues derive from the central government.

Under the 1944 act children were examined at age 11 (the so-called eleven plus), on the basis of which they were assigned either to grammar schools, which prepared them for higher education, or to secondary modern schools, which they usually left at the minimum leaving age, often without having passed public examinations or having been awarded a certificate. In 1964 the Labour government reorganized secondary education on comprehensive lines; that is, each school accepted pupils of all abilities. The eleven plus was abolished in many areas, and (except in Northern Ireland, where secondary education remains selective) the great majority of children now go to comprehensive schools.

The Education Reform Act of 1988 brought about substantial changes. Of its provisions for schools, five were particularly important. It laid on the secretary of state a duty to establish a national curriculum, applicable to all state schools; introduced a system by which individual schools controlled their own management and finance; and permitted schools to apply to opt out of control by local authorities. In addition, schools were required not unduly to restrict numbers (known as "open enrollment"), and the Inner London Education Authority, which provided education in the inner London boroughs, was abolished.

Private schools. Alongside the state sector a small number of private schools (often called "public" schools) provide education for about 6 percent of children. Their existence is controversial. It is argued that private schools skim off scarce resources of gifted children and teachers and of finance and that they perpetuate economic and social divisions. The counterarguments focus on their high quality, the beneficial effects of competition, and the freedom of parents to choose their favoured education.

Higher education. Universities historically have been independent and self-governing. But they have close links with the central government, in that a large proportion of their income derives from public funds. Higher education takes place also in polytechnics and some other colleges. The latter were run by local authorities until 1989, when they became autonomous institutions similar to the universities.

Students do not have a right to a place at a university; they are carefully selected by examination performance, and the dropout rate is low by international standards. The tuition fees are paid out of public funds; in addition, most students receive a tax-financed maintenance grant, inversely related to parental income, to cover living expenses. Foreign students and British students taking a degree at an overseas university are not generally eligible for public funding.

Public funds flow to universities through recurrent grants and in the form of tuition fees; universities derive income also from foreign students and from various private-sector sources. Prior to the Education Reform Act of 1988, the recurrent grant was channeled to institutions through the University Grants Committee (UGC), whose main purpose was to allocate public money independently from the government.

After a major expansion in the 1960s, the system came under pressure in the 1980s. Public funding became more stringent and, partly in consequence, the UGC became more interventionist. In part because of the high taxpayer cost involved, the proportion of the age group entering higher education in the United Kingdom was small by international standards. The grant system was also criticized, not least because it no longer supported students adequately. The Education Reform Act of 1988 abolished the UGC and replaced it with a new body, the Universities Funding Council, with a similar body for polytechnics. The measure was criticized by universities for its centralizing tendency but received support from polytechnics for freeing them from local authority control.

The Open University

In 1971 the Open University—something unique of its kind in the world—was begun. It is a degree-granting institution that provides courses of study for adults who missed the chance of higher education, and it uses television, radio, and local study programs. Applicants must apply for a number of places limited at any time by the availability of teachers.

HEALTH AND WELFARE

The National Health Service. In 1948 the National Health Service (NHS) replaced earlier, more piecemeal arrangements. In England the NHS is the responsibility

of the secretary of state for health, who is answerable to Parliament. The secretaries of state for Wales, Scotland, and Northern Ireland are responsible for the NHS in their respective countries.

Medical care under the NHS has a tripartite structure of primary care, hospitals, and community health care. The main element in primary care is the system of general practitioners (family doctors). Everyone can be registered with a general practitioner, who provides preventive and curative care and who refers patients to hospital and specialist services. All consultations with a general practitioner under the NHS are free.

The other major types of primary medical care are dentistry and pharmaceutical and opthalmic services. These are the only services of the NHS for which charges are levied, though people under 16, past retirement, or with low incomes are usually exempt. Everyone else must pay charges that are below the full cost of the services involved.

Under the Department of Health in England are 14 regions, under which are 193 districts, whose major responsibility is to run the hospital service. (Health authorities in Scotland, Wales, and Northern Ireland are the responsibilities of the respective secretaries of state.) Hospitals absorb more than two-thirds of the NHS budget. All hospital treatment under the NHS is free, including consultations with doctors, nursing, drugs, and intensive care, whatever the type of medical problem and however long the hospital stay. Hospital doctors are paid a salary rather than a fee for service but can combine salaried work for the NHS with private practice.

The Community Health Service has three functions: preventive health services; liaison with the local authority, especially over matters of public health; and cooperation with local authority personal social service departments, to enable health and personal care to be dealt with together wherever possible.

Patient choice

Individuals can register with any NHS general practitioner in their area who is prepared to add them to his or her list of patients. Anyone who wishes to change to another doctor may do so. Except in emergencies, patients are referred to a hospital by their general practitioner, allowing an element of patient choice.

Apart from the charges mentioned above, treatment under the NHS is free to the patient. The service is almost entirely funded from government revenues, less than 5 percent of NHS revenue coming from charges. This arrangement is unique among industrialized countries. There is no substantial reliance on private medical insurance (as, for example, in the United States).

The NHS budget, like that for any other government service, is determined by negotiation between the Treasury and the spending departments, as modified by subsequent discussion in the Cabinet. The resulting figure is a budget for the NHS as a whole. The division of money throughout the United Kingdom is partly constrained by a formula designed to improve the geographic distribution of medical resources. Each regional authority divides its total funds among the district health authorities.

Alongside the NHS is a system of private medical care both for primary care and hospital treatment. Though it grew somewhat over the 1980s, the sector remains small, involving only about 5 percent of the total expenditure on doctors and hospital in-patient care. Most private care is financed by voluntary private medical insurance, which is inexpensive, partly because the most serious cases are generally treated by the NHS.

Although the NHS, which was subject to reforms in the 1970s and '80s, is a popular institution, it is not without problems. Resources are scarce; many hospital buildings are old; there are waiting lists for nonurgent conditions; the distribution of health care by social class and by region is less equal than many would wish; and management needs to be improved.

The advantages, however, are enormous. The NHS is very inexpensive by international standards: in the late 1980s, for example, the United Kingdom spent less than 6 percent of its gross domestic product on health care; many other countries spent half as much again, and the United States spent more than 11 percent. Despite such low spending, health in the United Kingdom, measured in terms of infant mortality and life expectancy, matches that in comparable countries. The variation in the quality and quantity of treatment by income level is smaller than in most other countries. The system is able to direct resources toward specific regions and specific types of care. Treatment is free, whatever the extent and duration of illness; no one is denied treatment because of low income, and no one goes in fear of financial ruin.

Cash benefits. The current system of cash benefits, though substantially modified since its introduction in 1946, is based on the 1942 "Beveridge Report." Every employed person pays a national insurance contribution, which since 1975 has taken the form of a percentage of earnings. The contribution is collected by employers, and there is also an employer contribution. Separate arrangements exist for the self-employed. The revenue from contributions from individuals and employers goes into the National Insurance Fund.

Insurance benefits

Insured individuals are entitled to unemployment compensation, cash benefits during sickness or disability, and a retirement pension (at age 65 for men, 60 for women). There are also benefits for individuals injured in work-related accidents and for widows. Anyone, whether or not receiving an insurance benefit, may also be eligible for noncontributory benefit.

In 1965 a statutory system of redundancy, or severance, pay was introduced. Employees who lose their jobs through no fault of their own receive lump-sum payments whose cost is met in part by their employers and in part from a general levy on employers.

The major noncontributory benefits, paid out of general tax revenues, offer poverty relief to individuals and families whose income and savings fall below some prescribed level. The benefit of last resort is income support (formerly called supplementary benefit); it is payable to individuals whose entitlement to insurance benefits has been exhausted or has left them with a very low income and to those who never had any entitlement to an insurance benefit. Other means-tested benefits assist low-paid working families with children and help people on low incomes with their housing costs. An important class of noncontributory benefit is not means-tested, the major example being child benefit, a weekly, tax-free payment for each child, usually payable to the mother.

The 1946 system changed substantially over the years, with a burst of reform in the mid-1970s. The original Beveridge scheme was modeled on private insurance and offered flat-rate benefits in return for flat-rate contributions. The contribution system was criticized, not least because it bore heavily on the lower-paid. Under the Social Security Act of 1973 the flat-rate contribution was replaced from April 1975 for all employed persons by a percentage contribution of earnings up to a ceiling.

Pension arrangements were criticized in the 1950s and '60s because the flat-rate pension failed to keep many elderly people out of poverty. The Social Security Pensions Act of 1975 overhauled the system, most notably by introducing a substantial earnings-related pension, to be phased in over 20 years, so as to raise the income of pensioners.

The Beveridge arrangements offered financial support to families through a dual system of income tax allowances and a taxable family allowance for all children after the first. The result was complex and failed to give the greatest assistance to the poorest families. Under the Child Benefit Act of 1975, the tax allowance and family allowance were both abolished and replaced by the unified child benefit described above.

There was a further review of the benefit system in 1985, which changed the detailed workings of several benefits in 1988 but which left the basic structure intact.

Welfare aspects of housing. Under the Housing (Homeless Persons) Act of 1977 (which amended older legislation), local authorities have a statutory obligation in certain circumstances to find housing for homeless families. Partly for that reason, they own a substantial stock of housing for rent and maintain waiting lists and allocate housing according to need. (N.A.Ba.)

Cultural life

GENERAL CONSIDERATIONS

The cultural life of the United Kingdom is perforce that of the dominant English culture, as leavened and enriched by the contributions of Scotland, Wales, and Northern Ireland and, to a lesser degree, by those of the dozens of cultures that the period of English colonialism brought into contact with the homeland. Scotland, Wales, and Northern Ireland share fully in this common culture while preserving for themselves lively traditions that predate political union with England. (For discussions of the various arts and of major cultural institutions, see the sections *Cultural life* below under *England, Scotland, Wales,* and *Northern Ireland.*) That said, some general comment on British cultural life may be offered.

Widespread changes in the United Kingdom's cultural life occurred after 1945. The most remarkable was perhaps the emergence first of Liverpool and then of London in the 1960s as a world centre of popular culture. The Beatles were only the first and best-known of the many British rock music groups to win a world following. British clothing designers for a time led the world as innovators of new styles of dress for both men and women, and the brightly coloured outfits sold in Carnaby Street and King's Road shops briefly became more symbolic of Britain than the traditionally staid tailoring of Savile Row.

The social settings of the arts

Underlying this development, and a similar if less remarked renewal of vigour in more traditional fields in the decades after World War II, were several important social developments. Most evident was the rising standard of education. The number of pupils going on to higher education increased dramatically after World War II and was matched by a major expansion in the number of universities and other institutions of higher education. In society in general there was a marked increase in leisure.

During this period, successive governments shifted their policies toward the arts. The independent Arts Council, formed in 1946, supports many kinds of contemporary creative and performing arts. This support has supplemented the great expansion of the cultural market, mainly commercial, and of audiences and viewers for the arts generally. As in many other highly developed countries, the clash of tastes and values between generations and, to some extent, between social classes, has occasionally been sharp, especially in the 1960s and '70s.

MASS MEDIA

The communications media—press, publishing, broadcasting, and entertainment—provide mass coverage in Great Britain, but their scale ranges from the multimillion audiences for television, radio, and national newspapers to the small minority audiences for local papers, specialist periodicals, or experimental theatre and film.

The press. In both sales and reputation the national papers dominate. Within the national-newspaper business in the United Kingdom, a distinction has grown up between popular papers with multimillion circulation and quality papers with relatively small sales. Four populars account for about 85 percent of the total morning paper circulation. A tradition of editorial independence is recognized in newspaper groups, which give substantial freedom to the running of each of their newspapers and periodicals.

The Welsh press includes several daily papers as well as a number of weekly English-language, bilingual, or Welsh-language newspapers. Scotland also has national daily newspapers and a number of regional weeklies as well. Northern Ireland's daily papers are all published in Belfast. There is a large periodical press in the United Kingdom that ranges from such traditional publications as *Punch* to more specialized and, often, more mercurial journals.

Broadcasting. A monopoly of both radio and television broadcasting was in the hands of the BBC, which had been established as an independent public corporation in 1927, until 1954, when the Independent Television Authority (ITA) was established. The ITA was intended to provide the facilities for a number of commercial television program companies. In 1972 the ITA became the Independent Broadcasting Authority (IBA), with responsibility also for commercial radio.

The BBC and ITA

The BBC draws its revenue from license fees (on a scale fixed by the government) from persons owning receiving sets. It is governed by 12 individuals appointed by the monarch, with separate governors for Scotland, Wales, and Northern Ireland. The IBA, which also has a governing board of 12 members—these appointed by the Home Secretary—owns transmitters that it leases to the commercial program companies, which obtain their revenue from selling advertising time. The BBC operates two television channels, and the IBA operates one (with a second operated via a subsidiary that began operations in 1982). In its second television channel the BBC has tended to put programs of above-average intellectual and cultural interest—competition that the IBA commercial channel meets with its own cultural programs. The BBC also operates a comprehensive external service, broadcasting around the world in 40 languages, as well as a world service in English 24 hours a day.

Both the BBC and the IBA supply educational programs for schools and for adult studies. The Open University, offering degree courses to people who are without formal academic qualifications, uses educational programs that are broadcast by the BBC; these programs are backed by correspondence courses.

Both the BBC and IBA are public bodies that in the last resort can be controlled by the government, and Parliament can alter the terms of their authority. The government has the statutory power to veto a broadcast, but only rarely does it interfere with the day-to-day management of the BBC and IBA. More than 30 BBC local radio stations serve England and the Channel Islands, and regional and community radio stations also serve Scotland, Wales, and Northern Ireland.

For statistical data on the land and people of the United Kingdom, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (Ed.)

HISTORY OF ENGLAND AND GREAT BRITAIN

Ancient Britain

Apart from a few short references in classical literature, knowledge of Britain before the Roman conquest (begun AD 43) is derived entirely from archaeological research. It is thus lacking in detail, for archaeology can rarely identify personalities, motives, or exact dates. All that is available is a picture of successive cultures and some knowledge of economic development. But even in Roman times Britain lay on the periphery of the civilized world, and Roman historians, for the most part, provide for that period only a framework into which the results of archaeological research can be fitted. Britain truly emerged into the light of history only after the Saxon settlements in the 5th century AD.

Until late in the Mesolithic period, Britain formed part of the continental landmass and was easily accessible to migrating hunters. The cutting of the land bridge, *c.* 6000–5000 BC, had important effects: migration became more difficult and remained for long impossible to large numbers. Thus Britain developed insular characteristics, absorbing and adapting rather than fully participating in successive continental cultures. And within the island geography worked to a similar end; the fertile southeast was more receptive of influence from the adjacent continent than were the less-accessible hill areas of the west and north. Yet in certain periods the use of sea routes brought these too within the ambit of the continent.

From the end of the Ice Age (*c.* 11,000 BC), there was a gradual amelioration of climate leading to the replacement of tundra by forest and of reindeer hunting by that of red deer and elk. Valuable insight on contemporary conditions was gained by the excavation of a lakeside settlement at Star Carr, Humberside, which was occupied for about 20 successive winters by hunting people in the 8th millennium BC.

PRE-ROMAN BRITAIN

Neolithic period. A major change occurred *c.* 4000 BC with the introduction of agriculture by Neolithic immigrants from the coasts of western and possibly northwestern Europe. They were pastoralists as well as tillers of the soil. Tools were commonly of flint won by mining, but axes of volcanic rock were also traded by prospectors exploiting distant outcrops. The dead were buried in communal graves of two main kinds: in the west, tombs were built out of stone and concealed under mounds of rubble; in the stoneless eastern areas the dead were buried under long barrows (mounds of earth), which normally contained timber structures. Other evidence of religion comes from enclosures (*e.g.*, Windmill Hill, Wiltshire), which are now believed to have been centres of ritual and of seasonal tribal feasting. From them developed, late in the 3rd millennium, more clearly ceremonial ditch-enclosed earthworks known as henge monuments. Some, like Durrington Walls, Wiltshire, are of great size and enclose subsidiary timber circles. British Neolithic culture thus developed its own individuality.

Ritual centres

Bronze Age. Early in the 2nd millennium or perhaps even earlier, from *c.* 2300 BC, changes were introduced by the Beaker folk from the Low Countries and the middle Rhine. These people buried their dead in individual graves, often with the drinking vessel that gives their culture its name. The earliest of them still used flint; later groups, however, brought a knowledge of metallurgy and were responsible for the exploitation of gold and copper deposits in Britain and Ireland. They may also have introduced an Indo-European language. Trade was dominated by the chieftains of Wessex, whose rich graves testify to their success. Commerce was far-flung, in one direction to Ireland and Cornwall and in the other to central Europe and the Baltic, whence amber was imported. Amber bead spacers from Wessex have been found in the shaft graves at Mycenae in Greece. It was, perhaps, this prosperity that enabled the Wessex chieftains to construct the remarkable monument of shaped sarsens (large sandstones) known as Stonehenge III. Originally a late Neolithic henge, Stonehenge was uniquely transformed in Beaker times with a circle of large bluestone monoliths transported from southwest Wales.

Stonehenge

Little is known in detail of the early and middle Bronze Age. Because of present ignorance of domestic sites, these periods are mainly defined by technological advances and changes in tools or weapons. In general, the southeast of Britain continued in close contact with the continent and the north and west with Ireland.

From about 1200 BC there is clearer evidence for agriculture in the south; the farms consisted of circular huts in groups with small oblong fields and stock enclosures. This type of farm became standard in Britain down to and into the Roman period. From the 8th century onward, expansion of continental Urnfield and Hallstatt groups brought new people (mainly the Celts) to Britain; they came at first, perhaps, in small prospecting groups, but soon their influence spread, and new settlements developed. Some of the earliest hill forts in Britain were constructed in this period (*e.g.*, Beacon Hill, near Ivinghoe, Buckinghamshire; or Finavon, Angus); though formally belonging to the late Bronze Age, they usher in the succeeding period.

Iron Age. Knowledge of iron, introduced in the 7th century, was a merely incidental fact: it does not signify a change of population. The centuries 700–400 BC saw a succession of small migrations, and the newcomers mingled easily with existing inhabitants. Yet the greater availability of iron facilitated land clearance and thus the growth of population. The earliest ironsmiths made daggers of the Hallstatt type but of a distinctively British form. The settlements were also of a distinctively British type, with the traditional round house, the "Celtic" system of farming with its small fields, and storage pits for grain. Thus Britain absorbed the newcomers.

Celtic field system

The century following 600 BC saw the building of many large hill forts; these suggest the existence of powerful chieftains and the growth of strife as increasing population created pressures on the land. By 300 BC swords were making their appearance once more in place of daggers. Finally, beginning in the 3rd century, a British form of La Tène Celtic art was developed to decorate warlike equipment such as scabbards, shields, and helmets, and eventually also bronze mirrors and even domestic pottery. During the 2nd century the export of Cornish tin, noted before 300 by Pytheas of Massalia, a Greek explorer, continued; evidence of its destination is provided by the Paul (Cornwall) hoard of north Italian silver coins. In the 1st century BC this trade was in the hands of the Veneti of Brittany; their conquest (56 BC) by Julius Caesar, who destroyed their fleet, seems to have put an end to it.

By 200 Britain had fully developed its insular Celtic character. The emergence, however, of the British tribes known to Roman historians was due to a further phase of settlement by tribesmen from Belgic Gaul. Coin finds suggest that the earliest movements of this migration began before the end of the 2nd century; the decisive settlements were made in the 1st century probably as a result of pressures in Gaul created by Germanic and Roman expansion. The result was a distinctive culture in southeast Britain (especially in Kent and north of the Thames) which represented a later phase of the continental Celtic La Tène culture. Its people used coins and the potter's wheel and cremated their dead, and their better equipment enabled them to begin the exploitation of heavier soils for agriculture.

ROMAN BRITAIN

The conquest. Julius Caesar conquered Gaul between 58 and 50 BC and invaded Britain in 55 or 54 BC, thereby bringing the island into close contact with the Roman world. Caesar's description of Britain at the time of his invasions is the first coherent account extant. From about 20 BC it is possible to distinguish two principal powers: the Catuvellauni north of the Thames led by Tasciovanus,

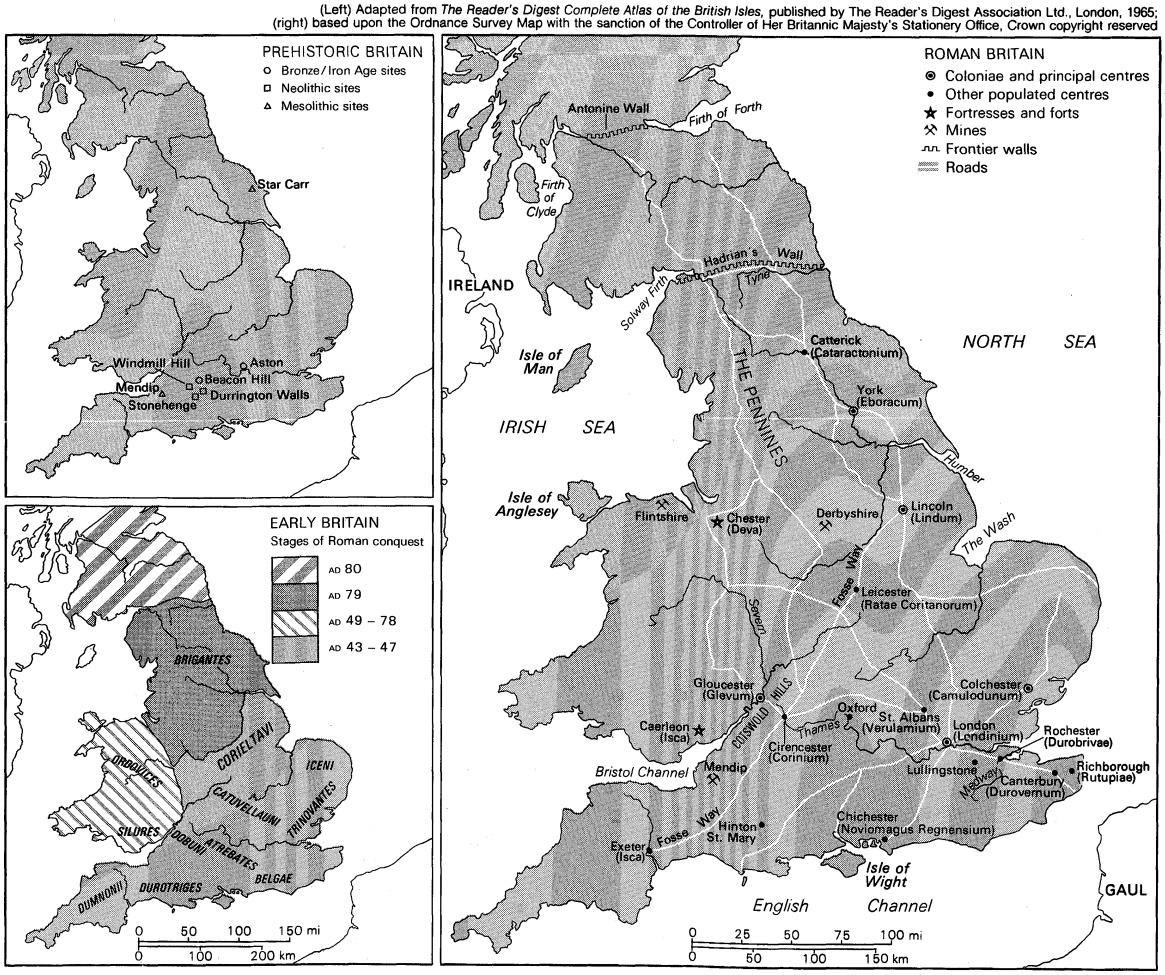
successor of Caesar's adversary Cassivellaunus, and, south of the river, the kingdom of the Atrebates ruled by Commius and his sons Tincommius, Eppillus, and Verica. Tasciovanus was succeeded in about AD 5 by his son Cunobelinus, who, during a long reign, established power all over the southeast, which he ruled from Camulodunum (Colchester). Beyond these kingdoms lay the Iceni in what is now Norfolk, the Corieltavi in the Midlands, the Dobuni (Dobunni) in the area of Gloucestershire, and the Durotriges in that of Dorset, all of whom issued coins and probably had Belgic rulers. Behind these again lay further independent tribes—the Dumnonii of Devon, the Brigantes in the north, and the Silures and Ordovices in Wales. The Belgic and semi-Belgic tribes later formed the civilized nucleus of the Roman province and thus contributed greatly to Roman Britain.

Claudius' conquest of Britain

The client relationships that Caesar had established with certain British tribes were extended by Augustus. In particular, the Atrebatic kings welcomed Roman aid in their resistance to Catuvellaunian expansion. The decision of the emperor Claudius to conquer the island was the result partly of his personal ambition, partly of British aggression. Verica had been driven from his kingdom and appealed for help, and it may have been calculated that a hostile Catuvellaunian supremacy would endanger stability across the Channel. Under Aulus Plautius an army of four legions was assembled, together with a number of auxiliary regiments consisting of cavalry and infantry raised among warlike tribes subject to the empire. After delay caused by the troops' unwillingness to cross the ocean, which they then regarded as the boundary of the human world, a landing was made at Richborough, Kent, in AD 43. The British under Togodumnus and Caratacus, sons and successors of Cunobelinus, were taken by surprise and defeated. They retired to defend the Medway crossing near Rochester but were again defeated in a hard battle. The way to Camulodunum lay open, but Plautius halted at the Thames to await the arrival of the emperor, who took personal command of the closing stages of the campaign. In one short season the main military opposition had been crushed: Togodumnus was dead and Caratacus had fled to Wales. The rest of Britain was by no means united, for Belgic expansion had created tensions. Some tribes submitted, and subduing the rest remained the task for the year 44. For this purpose smaller expeditionary forces were formed consisting of single legions or parts of legions with their *auxilia* (subsidiary allied troops). The best-documented campaign is that of Legion II under its legate Vespasian starting from Chichester, where the Atrebatic kingdom was restored; the Isle of Wight was taken and the hill forts of Dorset reduced. Legion IX advanced into Lincolnshire, and Legion XIV probably across the Midlands toward Leicester. Colchester was the chief base, but the fortresses of individual legions at this stage have not yet been identified.

By the year 47, when Plautius was succeeded as commanding officer by Ostorius Scapula, a frontier had been established from Exeter to the Humber, based on the road known as the Fosse Way; from this fact it appears that Claudius did not plan the annexation of the whole island but only of the arable southeast. The intransigence of the tribes of Wales, spurred on by Caratacus, however, caused Scapula to occupy the lowlands beyond the Fosse Way up to the River Severn and to move forward his forces into this area for the struggle with the Silures and Ordovices. The Roman forces were strengthened by the addition of Legion XX, released for this purpose by the foundation of a veteran settlement (*colonia*) at Camulodunum in the year 49. The *colonia* would form a strategic reserve as well as setting the Britons an example of Roman urban organization and life. A provincial centre for the worship of the emperor was also established. Scapula's right flank

(Left) Adapted from *The Reader's Digest Complete Atlas of the British Isles*, published by The Reader's Digest Association Ltd., London, 1965; (right) based upon the Ordnance Survey Map with the sanction of the Controller of Her Britannic Majesty's Stationery Office, Crown copyright reserved



Ancient Britain.

was secured by the treaty relationship that had been established with Cartimandua, queen of the Brigantes. Hers was the largest kingdom in Britain, occupying the whole area between Derbyshire and the Tyne; unfortunately it lacked stability, nor was it united behind its queen, who lost popularity when she surrendered the British resistance leader, Caratacus, to the Romans. Nevertheless, with occasional Roman military support, Cartimandua was maintained in power until 69 against the opposition led by her husband, Venutius, and this enabled Roman governors to concentrate on Wales.

By AD 60 much had been achieved; Suetonius Paulinus, governor from 59 to 61, was invading the island of Anglesey, the last stronghold of independence, when a serious setback occurred: this was the rebellion of Boudicca, queen of the Iceni. Under its king Prasutagus the tribe of the Iceni had enjoyed a position of alliance and independence; but on his death (60) the territory was forcibly annexed and outrages occurred. Boudicca was able to rally other tribes to her assistance; chief of these were the Trinovantes of Essex, who had many grievances against the settlers of Camulodunum for their arrogant seizure of lands. Roman forces were distant and scattered; and, before peace could be restored, the rebels had sacked Camulodunum, Verulamium (St. Albans), and London, the three chief centres of Romanized life in Britain. Paulinus acted harshly after his victory, but the procurator of the province, Julius Classicianus, with the revenues in mind and perhaps also because, as a Gaul by birth, he possessed a truer vision of provincial partnership with Rome, brought about his recall.

Boudicca's rebellion

In the first 20 years of occupation some progress had been made in spreading Roman civilization. Towns had been founded, the imperial cult had been established, and merchants were busily introducing the Britons to material benefits. It was not, however, until the Flavian period, AD 69–96, that real advances were made in this field. With the occupation of Wales by Julius Frontinus (governor from 74 to 78) and the advance into northern Scotland by Gnaeus Julius Agricola (78–84), troops were removed from southern Britain, and self-governing civitates, administrative areas based for the most part on the indigenous tribes, took over local administration. This involved a large program of urbanization and also of education, which continued into the 2nd century; Tacitus, in his biography of Agricola, emphasizes the encouragement given to it. Roman conquest of Wales was complete by 78, but Agricola's invasion of Scotland failed because shortage of manpower prevented him from completing the occupation of the whole island. Moreover, when the British garrison was reduced (*c.* AD 90) by a legion because of continental needs, it became evident that a frontier would have to be maintained in the north. After several experiments, the Solway–Tyne isthmus was chosen, and there the emperor Hadrian built his stone wall (*c.* 122–130).

Condition of the province. There was a marked contrast in attitude toward the Roman occupation between the lowland Britons and the inhabitants of Wales and the hill country of the north. The economy of the former was that of settled agriculture, and they were largely of Belgic stock; they soon accepted and appreciated the Roman way of life. The economy of the hill dwellers was pastoral, and the urban civilization of Rome threatened their freedom of life. Although resistance in Wales was stamped out by the end of the 1st century AD, Roman influences were nonetheless weak except in the Vale of Glamorgan. In the Pennines until the beginning of the 3rd century there were repeated rebellions, the more dangerous because of the threat of assistance from free Scotland.

Army and frontier. After the emperor Domitian had reduced the garrison in about the year 90, three legions remained; their permanent bases were established at York, Chester, and Caerleon. The legions formed the foundation of Roman military power, but they were supplemented in garrison duty by numerous smaller auxiliary regiments both of cavalry and infantry, either 1,000 or 500 strong. These latter garrisoned the wall and were stationed in a network of other forts established for police work in Wales and northern England. With 15,000 legionaries and about 40,000 auxiliaries, the army of Britain was very powerful; its presence had economic as well as political results. Hadrian's Wall was the most impressive frontier work in the Roman Empire. Despite a period in the following two reigns when another frontier was laid out on the Glasgow–Edinburgh line—the Antonine Wall, built of turf—the wall of Hadrian came to be the permanent frontier of Roman Britain. The northern tribes only twice succeeded in passing it, and then at moments when the garrison was fighting elsewhere. In the late Roman period, when sea raiding became prevalent, the wall lost its preeminence as a defense for the province, but it was continuously held until the end of the 4th century. But although they withdrew to Hadrian's line not later than the year 180, the Romans never abandoned interest in southern Scotland. In the 2nd century their solution was military occupation. In the 3rd, after active campaigning (208–211) by the emperor Septimius Severus and his sons during which permanent bases were built on the east coast of Scotland, the solution adopted by the emperor Caracalla was regulation of relationship by treaties. These, perhaps supported by subsidies, were enforced by supervision of the whole Lowlands by patrols based on forts beyond the wall. During the 4th century more and more reliance was placed on friendly native states, and patrols were withdrawn.

Hadrian's Wall

Administration. Britain was an imperial province. The governor represented the emperor, exercising supreme military as well as civil jurisdiction. As commander of three legions he was a senior general of consular rank. From the late 1st century he was assisted on the legal side by a *legatus juridicus.* The finances were in the hands of the provincial procurator, an independent official of equestrian status whose staff supervised imperial domains and the revenues of mines in addition to normal taxation. In the early 3rd century Britain was divided into two provinces in order to reduce the power of its governor to rebel, as Albinus had done in 196: Britannia Superior had its capital at London and a consular governor in control of two legions and a few auxiliaries; Britannia Inferior, with its capital at York, was under a praetorian governor with one legion but many more auxiliaries.

Local administration was of varied character. First came the chartered towns. By the year 98 Lincoln and Gloucester had joined Camulodunum as *coloniae,* and by 237 York had become a fourth. *Coloniae* of Roman citizens enjoyed autonomy with a constitution based on that of republican Rome, and Roman citizens had various privileges before the law. It is likely that Verulamium was chartered as a Latin municipium (free town); in such a town the annual magistrates were rewarded with Roman citizenship. The remainder of the provincials ranked as *peregrini* (subjects). In military districts control was in the hands of fort prefects responsible to legionary commanders; but by the late 1st century local self-government, as already stated, was granted to *civitates peregrinae,* whose number tended to increase with time. These also had republican constitutions, being controlled by elected councils and annual magistrates and having responsibility for raising taxes and administering local justice. In the 1st century there were also client kingdoms whose rulers were allied to Rome; Cogidubnus, Verica's successor, who had his capital at Chichester, is the best known. But Rome regarded these as temporary expedients, and none outlasted the Flavian Period (69–96).

Society. Pre-Roman Celtic tribes had been ruled by kings and aristocracies; the Roman civitates remained in the hands of the rich because of the heavy expense of office. But since trade and industry now yielded increasing profits and the old aristocracies no longer derived wealth from war but only from large estates, it is likely that new men rose to power. Roman citizenship was now an avenue of social advancement, and it could be obtained by 25 years' service in the auxiliary forces as well as (more rarely) by direct grants. Soldiers and traders from other parts of the empire significantly enhanced the cosmopolitan character of the population, as did the large number of legionaries, who were already citizens and many of whom must have settled locally. The population of Roman Britain at its peak amounted perhaps to about two million.

Roman citizenship

Economy. Even before the conquest, according to the Greek geographer Strabo, Britain exported gold, silver, iron, hides, slaves, and hounds in addition to grain. A Roman gold mine is known in Wales, but its yield was not outstanding. Iron was worked in many places but only for local needs; silver, obtained from lead, was of more significance. But the basis of the economy was agriculture, and the conquest greatly stimulated production because of the requirements of the army. According to Tacitus, grain to feed the troops was levied as a tax; correspondingly more had to be grown before a profit could be made. The pastoralists in Wales and the north probably had to supply leather, which the Roman army needed in quantity for tents, boots, uniforms, and shields. A military tannery is known at Catterick. A profit could, nonetheless, be won from the land because of the increasing demand from the towns. At the same time the development of a system of large estates (villas) relieved the ancient Celtic farming system of the necessity of shouldering the whole burden. Small peasant farmers tended to till the lighter, less-productive, more easily worked soils. Villa estates were established on heavier, richer soils, sometimes on land recently won by forest clearance, itself a result of the enormous new demand for building timber from the army and the new towns and for fuel for domestic heating and for public baths. The villa owners had access to the precepts of classical farming manuals and also to the improved equipment made available by Roman technology. Their growing prosperity is vouched for by excavation: there are few villas that did not increase in size and luxury as corridors and wings were added or mosaics and bath blocks provided. At least by the 3rd century some landowners were finding great profit in wool; Diocletian's price edict (AD 301) shows that at least two British cloth products had won an empire-wide reputation. Archaeological evidence indicates that the Cotswold district was one of the centres of this industry.

Villa estates

Trade in imported luxury goods ranging from wine to tableware and bronze trinkets vastly increased as traders swarmed in behind the army to exploit new markets. The profits of developing industries went similarly at first to foreign capitalists. This is clearly seen in the exploitation of silver-lead ore and even in the pottery industry. The Mendip lead field was being worked under military control as early as the year 49, but under Nero (54–68) both there and in Flintshire, and not much later also in the Derbyshire lead field, freedmen—the representatives of Roman capital—were at work. By Vespasian's reign (69–79) organized companies (societates) of prospectors are attested. Roman citizens, who must in the context be freedmen, are also found organizing the pottery industry in the late 1st century. Large profits were made by continental businessmen in the first two centuries not only from such sources but also by the import on a vast scale of high-class pottery from Gaul and the Rhineland and on a lesser scale of glass vessels, luxury metalware, and Spanish oil and wine. A large market existed among the military, and the Britons themselves provided a second. Eventually this adverse trade balance was rectified by the gradual capture of the market by British products. Much of the exceptional prosperity of 4th-century Britain must have been due to its success in retaining available profits at home.

A final important point is the role of the Roman army in the economic development of the frontier regions. The presence as consumers of large forces in northern Britain created a revolution in previous patterns of trade and civilized settlement. Cereal production was encouraged in regions where it had been rare, and large settlements grew up in which many of the inhabitants must have been retired soldiers with an interest in the land as well as in trade and industry.

Towns. Belgic Britain had large centres of population but not towns in the Roman sense of having not merely streets and public buildings but also the amenities and local autonomy of a city. In Britain these had therefore to be provided if Roman civilization and normal methods of provincial administration were to be introduced. Thus a policy of urbanization existed in which the legions, as the nearest convenient source of architects and craftsmen, played an organizing role. The earlier towns consisted of half-timbered buildings; before AD 100 only public buildings seem to have been of stone. The administrative capitals had regular street grids, a forum with basilica (public hall), public baths, and temples; a few had theatres and amphitheatres, too. With few exceptions they were undefended. In the 3rd century, town walls were provided, not so much as a precaution in unsettled times but as a means of keeping operational the earthwork defenses already provided during a crisis at the end of the 2nd century. These towns grew in size to about 100–130 acres with populations of about 5,000; a few were twice this size. The majority of towns in Roman Britain seem to have developed out of traders' settlements in the vicinity of early garrison-forts: those that were not selected as administrative centres remained dependent for their existence on economic factors, serving either as centres of trade or manufacture or else as markets for the agricultural peasantry. They varied considerably in size. In the north, where garrisons were permanently established, quite large trading settlements grew up in their vicinity, and at least some of these would rank as towns.

Urbanization

Villas. Apart from the exceptional establishment at Fishbourne, in West Sussex, whose Italian style and luxurious fittings show that it was the palace of King Cogidubnus, the houses of Romano-British villas had simple beginnings and were of a provincial type. A few owners were prosperous enough in the 2nd century to afford mosaics; but the great period of villa prosperity lay in the 4th century, when many villas grew to impressive size. Their importance was economic and has already been described. Much remains to be learned from full excavation of their subsidiary work buildings. Larger questions of tenure and organization are probably insoluble in the absence of documentary evidence, for it is dangerous to draw analogies from classical sources since conditions in Celtic Britain were very different from those of the Mediterranean world.

Religion and culture. A great variety of religious cults were to be found. In addition to numerous Celtic deities of local or wider significance, the gods of the classical pantheon were introduced and were often identified with their Celtic counterparts. In official circles the worship of the state gods of Rome and of the imperial cult was duly observed. In addition merchants and soldiers introduced oriental cults, among them Christianity. The latter, however, made little headway until the late 4th century, though the frescoes at Lullingstone in Kent and the mosaics at Hinton St. Mary in Dorset attest its presence among villa owners. Although classical temples are sometimes found in towns, the normal temple was of the Romano-Celtic type consisting of a small square shrine and surrounding portico; temples of this type are found in town and country alike.

Romanization was strongest in the towns and among the upper classes, as would be expected; there is evidence that in the countryside Celtic continued to be spoken, though it was not written. Many people were bilingual: graffiti prove that even artisans wrote Latin. Evidence of the classical education of the villa owners is provided by their mosaics, which prove an acquaintance with classical mythology and even with the *Aeneid* of Virgil. Sculpture and wall painting were both novelties in Roman Britain. Statues or busts in bronze or marble were imported from Gaulish or Mediterranean workshops, but British sculptors soon learned their trade and at their best produced attractive works in a provincial idiom, often for votive purposes. Many cruder works were also executed whose interest lies in the proof they afford that the conventions of the classical world had penetrated even to the lower classes. Mosaic floors, found in towns and villas, were at first, as at Fishbourne, laid by imported craftsmen. But there is evidence that by the middle of the 2nd century a local firm was at work at Colchester and Verulamium, and in the 4th century a number of local mosaic workshops can be recognized by their styles. One of the most skilled of these was based in Cirencester.

Mosaics

Roman civilization thus took root in Britain; its growth was more obvious in urban circles than among the peas-

ants and weakest in the resistant highland zone. It was a provincial version of Roman culture, but one with recognizably British traits.

The decline of Roman rule. The reforms of Diocletian ended the chaos of the 3rd century and ushered in the late imperial period. Britain, however, for a short time became a separate empire through the rebellion (286/287) of Carausius. This man had been in command against the Saxon pirates in the Channel and by his naval power was able to maintain his independence. His main achievement was to complete the new system of Saxon Shore forts around the southeastern coasts. At first he sought recognition as coemperor, but this was refused. In 293 the fall of Boulogne to Roman forces led to his murder and the accession of Allectus, who, however, fell in his turn when Constantius I invaded Britain in 296. Allectus had withdrawn troops from the north to oppose the landing, and Hadrian's Wall seems to have been attacked, for Constantius had to restore the frontier as well as reform the administration. He divided Britain into four provinces, and in the same period the civil power was separated from the military. Late Roman sources show three separate commands respectively under the *dux Britanniarum* (commander of the Britains), the *comes litoris Saxonici* (count of the Saxon Shore), and the *comes Britanniarum,* though the dates of their establishment are unknown and may not have been identical.

The 4th century was a period of great prosperity in towns and countryside alike. Britain had escaped the "barbarian" invasions of the 3rd century and may have seemed a safe refuge for wealthy continentals. Its weakness lay in the fact that its defense was ultimately controlled by distant rather than local rulers. The garrison was perhaps weakened by withdrawals for the civil war of Magnentius (350–351); at any rate in 367 a military disaster occurred due to concerted seaborne attacks from the Picts of Scotland and the Scots of Ireland. But, though the frontier and forts behind it suffered severely, there is little trace of damage to towns or villas. Count Theodosius in 369 restored order and strengthened the defenses of the towns with external towers designed to mount artillery. Prosperity continued, but the withdrawals of troops by Magnus Maximus in 383 and again at the end of the century by Stilicho weakened security. Thus, when Constantine III, who was declared emperor by the army in Britain in 407, took further troops to Gaul, the forces remaining in the island were insufficient to provide protection against increasing Pictish and Saxon raids. The Britons appealed to the legitimate emperor, Honorius, who was unable to send assistance but authorized the cities to provide for their own defense (410). This marks the end of Roman Britain, for the central government never reestablished control, but for a generation there was little other outward change.

The rise of tyrants

Power fell gradually into the hands of tyrants. Chief of these was Vortigern (*c.* 425), who, unlike earlier usurpers, made no attempt to become Roman emperor but was content with power in Britain. Independence was producing separate interests. By this date Christianity had made considerable headway in the island, but the leaders followed the heretical teaching of Pelagius, himself a Briton, who had emphasized the importance of the human will over divine grace in the achievement of salvation. It has been held that the self-reliance shown in the maintenance of national independence was inspired by this philosophy. Yet there was also a powerful Roman Catholic party anxious to reforge the links with Rome, in support of whom St. Germanus of Auxerre visited Britain in 429. It may have been partly to thwart the plans of this party that Vortigern made the mistake (*c.* 430; the date given by the Anglo-Saxon Benedictine scholar Bede [d. 735] is between 446 and 454) of inviting Saxons to settle and garrison strategic areas of the east coast, though he certainly also had in mind the need to ward off seaborne raids by Picts, which at this time were troublesome. Planned settlement of this sort is the best explanation for the earliest Saxon settlements found around the mouths of the east-coast estuaries and also in the central southeast region around Oxford. For a time the system worked successfully, but, when in 442 these Saxon foederati (allies) rebelled and called in others of their race to help them, it was found that they had been given a stranglehold on Britain. A long period of warfare and chaos was inaugurated, which was economically disastrous. It was probably this period that saw the disintegration of the majority of the villa estates; with the breakdown of markets and the escape of slaves, villas ceased to be viable and must have gradually fallen into ruin, though the land itself did not cease to be cultivated. A few villas met a violent end. The towns, under the protection of their strong defenses, at first provided refuge at any rate for the rich who could leave their lands; but by degrees decay set in as trade declined and finally even the supply of food was threatened. In about 446 the British made a vain appeal for help to the Roman general Aetius (the "Groans of the Britons" mentioned in the *De excidio et conquestu Britanniae* of the British writer Gildas). For several decades they suffered reverses; many emigrated to Brittany. In the second half of the 5th century Ambrosius Aurelianus and the shadowy figure of Arthur began to turn the tide by the use of cavalry against the ill-armed Saxon infantry. A great victory was won at Mons Badonicus (a site not identifiable) toward 500: now it was Saxons who emigrated, and the British lived in peace all through the first half of the 6th century, as Gildas records. But in the second half the situation slowly worsened. (S.S.F.)

Anglo-Saxon England

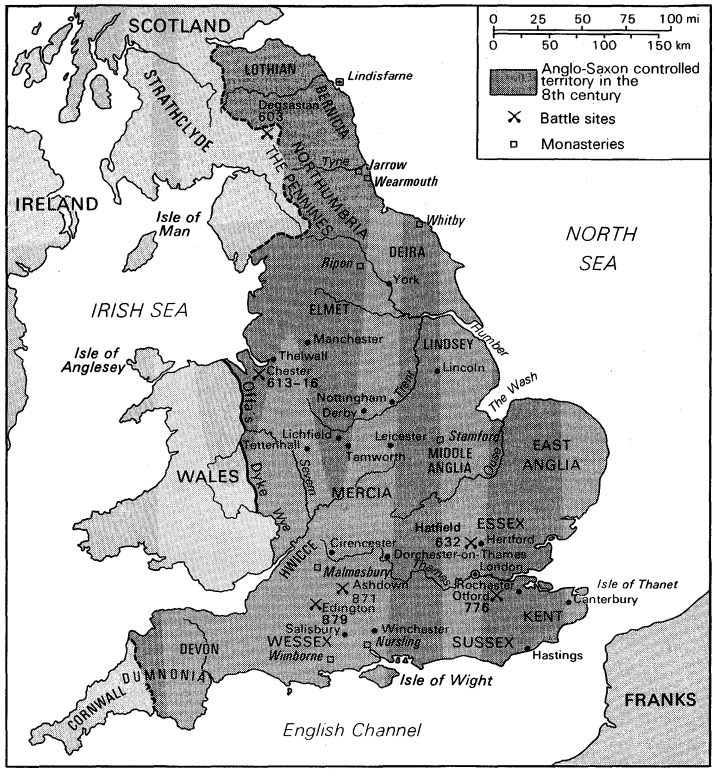
THE INVADERS AND THEIR EARLY SETTLEMENTS

Although Germanic foederati, allies of Roman and post-Roman authorities, had settled in England in the 4th century AD, tribal migrations into Britain began about the middle of the 5th century. The first arrivals, according to the 6th-century British writer Gildas, were invited by a British king to defend his kingdom against the Picts and Scots. A tradition reached Bede that the first mercenaries were from three tribes—the Angles, Saxons, and Jutes—which he locates on the Cimbric Peninsula, and by implication the coastlands of northwestern Germany. Archaeology, however, suggests a more complex picture showing many tribal elements, Frankish leadership in the first waves, and Frisian contacts. Revolt by these mercenaries against their British employers in the southeast of England led to large-scale Germanic settlements near the coasts and along the river valleys. Their advance was halted for a generation by native resistance, which tradition associates with the names of Ambrosius Aurelianus and Arthur, culminating in victory about 500 by the Britons at the Battle of Mons Badonicus at an unidentified location. But a new Germanic drive began about 550, and before the century had ended, the Britons had been driven west to the borders of Dumnonia (Cornwall and Devon) and to the Welsh Marches, while invaders were advancing west of the Pennines and northward into Lothian.

Battle of Mons Badonicus

The fate of the native British population is difficult to determine. The case against its large-scale survival rests largely on linguistic evidence, such as the scarcity of Romano-British words continuing into English and the use of English even by Northumbrian peasants. The case against wholesale extermination also rests on linguistic evidence, such as place-names and personal names, as well as on evidence provided by urban and rural archaeology. Certainly few Britons in England were above servile condition. By the end of the 7th century people regarded themselves as belonging to "the nation of the English," though divided into several kingdoms. This sense of unity was strengthened during long periods when all kingdoms south of the Humber acknowledged the overlordship (called by Bede an imperium) of a single ruler, known as a bretwalda, a word first recorded in the 9th century.

The first such overlord was Aelle of Sussex, in the late 5th century; the second was Ceawlin of Wessex, who died in 593. The third overlord, Aethelberht of Kent, held this power in 597 when the monk Augustine led a mission from Rome to Kent; Kent was the first English kingdom to be converted to Christianity. The Christian church provided another unifying influence, overriding political divisions, although it was not until 669 that the church in England acknowledged a single head.



Anglo-Saxon England.

Adapted from R. Treharne and H. Fullard (eds.), *Muir's Historical Atlas: Ancient, Medieval and Modern*, 9th ed. (1965); George Philip & Son Ltd., London

The social system. Aethelberht set down in writing a code of laws; although it reflects Christian influence, the system underlying the laws was already old, brought over from the Continent in its main lines. The strongest social bond of this system was that of kinship; every freeman depended on his kindred for protection, and the social classes were distinguished by the amount of their wergild (the sum that the kindred could accept in place of vengeance if a man were killed). The normal freeman was the ceorl, an independent peasant landowner; below him in Kent were persons with lower wergilds, who were either freedmen or, as were similar persons in Wessex, members of a subject population; above the ceorls were the nobles—some perhaps noble by birth but more often men who had risen by service as companions of the king—with a wergild three times that of a ceorl in Kent, six times that of a ceorl elsewhere. The tie that bound a man to his lord was as strong as that of the kindred. Both nobles and ceorls might possess slaves, who had no wergild and were regarded as chattels.

Early traditions, embodied in king lists, imply that all Anglo-Saxon kingdoms except Sussex were established by rulers deemed to have descended from the gods. No invading chieftain is described by the *Anglo-Saxon Chronicle* as "king"—although the title was soon used—and chieftainship, as before the conquest, remained central to Germanic tribal society. The sacral character of kingship later increased and changed in meaning as the Christian ruler was set apart by coronation and anointment. In the established English kingdoms the king had special rights—compensations for offenses committed in his presence or his home or against anyone under his protection; rights to hospitality, which later became a food rent charged on all land; and rights to various services. He rewarded his followers with grants of land, probably at first for their lifetime only, but the need to provide permanent endowment for the church brought into being a type of land that was free from most royal dues and that did not revert to the king. From the latter part of the 7th century such land was sometimes conferred by charter. It became common to make similar grants by charter to laymen, with power to bequeath; but three services—the building of forts and bridges and service in the army—were almost invariably excepted from the immunity. The king received fines for various crimes; but a man's guilt was established in an assembly of freemen, where the accused tried to establish his innocence by his oath—supported by oath helpers—and, if this failed, by ordeal. On matters of importance the king normally consulted his witan (wise men).

The rights of the king

There were local variations in the law, and over a period of time the law developed to meet changed circumstances. As kingdoms grew larger, for example, an official called an ealderman was needed to administer part of the area, and later a sheriff was needed to look after the royal rights in each shire. The acceptance of Christianity made it necessary to fit the clergy into the scale of compensations and assign a value to their oaths and to fix penalties for offenses such as sacrilege, heathen practices, and breaches of the marriage law. But the basic principles were little changed.

The Anglo-Saxons left England a land of villages, but the continuity of village development is uncertain. In the 7th–8th centuries, in what is called the "Middle Saxon shuffle," many early villages were abandoned, and others, from which later medieval villages descended, were founded. The oldest villages are not, as previously thought, those with names ending in *-ingas* but rather those ending in *-ham* and *-ingham*. English trading towns, whose names often end in *-wich*, from the Latin *vicus* ("village"), developed in the Middle Saxon period, and other urban settlements grew out of and date from the Alfredian and later defenses against Viking attack.

The conversion to Christianity. Place-names containing the names of gods or other heathen elements are plentiful enough to prove the vitality of heathenism and to account for the slow progress of conversion in some areas. In Kent, the first kingdom to accept Christianity, King Wihtred's laws in 695 contained clauses against heathen worship. The conversion renewed relations with Rome and the Continent; but the full benefit of this was delayed because much of England was converted by the Celtic church, which had lost contact with Rome.

King Wihtred's laws

Augustine's mission in 597 converted Kent; but it had only temporary success in Essex, which reverted to heathenism in 616. A mission sent under Bishop Paulinus from Kent to Northumbria in 627 converted King Edwin and many of his subjects in Northumbria and Lindsey. It received a setback in 632 when Edwin was killed and Paulinus withdrew to Kent. About 630 Archbishop Honorius of Canterbury sent a Burgundian, Felix, to convert East Anglia, and the East Anglian church thenceforth remained faithful to Canterbury. Soon after, the West Saxons were converted by Birinus, who came from Rome. Meanwhile, King Oswald began to restore Christianity in Northumbria, bringing Celtic missionaries from Iona. And it was the Celtic church that began in 653 to spread the faith among the Middle Angles, the Mercians, and the peoples of the Severn valley; it also won back Essex.

At first there was little friction between the Roman and Celtic missions. Oswald of Northumbria joined with Cynegils of Wessex in giving Dorchester-on-Thames as seat for Birinus' bishopric; the Irishmen Maildubh in Wessex and Fursey in East Anglia worked in areas converted by the Roman church; and James the Deacon continued Paulinus' work in Northumbria. Later, however, differences in usage—especially in the calculation of the date of Easter—caused controversy, which was settled in favour of the Roman party at the Synod of Whitby in 664. The adherents of Celtic usage either conformed or withdrew, and advocates of Roman practice became active in the north, the Midlands, and Essex. Theodore of Tarsus (arrived 669), the first Roman archbishop to be acknowledged all over England, was active in establishing a proper diocesan system, whereas in the Celtic church bishops tended to move freely without fixed sees and settled boundaries; he held the first synod of the English church at Hertford in 672, and this forbade a bishop to interfere in another's diocese or any priest to move into another diocese without his bishop's permission. Sussex and the Isle of Wight—the last outposts of heathenism—were converted by Bishop Wilfrid and his followers from 681 to 687 and thenceforth followed Roman usages.

The Anglo-Saxons attributed their conversion to Pope

"The Apostle of the English"

Gregory I, "the Apostle of the English," who had sent Augustine. This may seem less than fair to the Celtic mission. The Celtic church made a great impression by its asceticism, fervour, and simplicity, and it had a lasting influence on scholarship. Yet the period of Celtic dominance was only 30 years. The decision at Whitby made possible a form of organization better fitted for permanent needs than the looser system of the Celtic church.

The golden age of Bede. Within a century of Augustine's landing, England was in the forefront of scholarship. This high standard arose from a combination of influences: that from Ireland, which had escaped the decay caused elsewhere by the barbarian invasions, and that from the Mediterranean, which reached England mainly through Archbishop Theodore and his companion, the abbot Adrian. Under Theodore and Adrian, Canterbury became a famous school, and men trained there took their learning to other parts of England. One of these men was Aldhelm, who had been a pupil of Maildubh (the Irish founder of Malmesbury); under Aldhelm, Malmesbury became an influential centre of learning. Aldhelm's own works, in Latin verse and prose, reveal a familiarity with many Latin authors; his writings became popular among admirers of the ornate and artificial style he had learned from his Celtic teachers. Before long a liberal education could be had at such other West Saxon monasteries as Nursling and Wimborne.

The finest centre of scholarship was Northumbria. There Celtic and classical influences met: missionaries brought books from Ireland, and many Englishmen went to Ireland to study. Other Northumbrians went abroad, especially to Rome; among them was Benedict Biscop. Benedict returned from Rome with Theodore (668–669), spent some time in Canterbury, and then brought the learning acquired there to Northumbria. He founded the monasteries at Wearmouth (674) and Jarrow (682), where Bede spent his life. Benedict and Ceolfrith, abbot of Jarrow, brought books from the Continent and assembled the fine library that was available to Bede.

Bede (*c.* 672–735) is remembered as a great historian whose work never lost its value; but he was also a theologian regarded throughout the Middle Ages as second only to the Church Fathers. Nonetheless, even though he was outstanding, he did not work in isolation. Other Northumbrian houses—Lindisfarne, Whitby, and Ripon—produced saints' lives, and Bede was in touch with many learned men, not only in Northumbria; there are also signs of scholarly activity in London and in East Anglia.

Moreover, in this period religious poetry was composed in the diction and technique of the older secular poetry in the vernacular. *Beowulf,* considered the greatest Old English poem, is sometimes assigned to this age, but the dating is uncertain. Art flourished, with a combination of native elements and influences from Ireland and the Mediterranean. The Hiberno-Saxon (or Anglo-Irish) style of manuscript illumination was evolved, its greatest example—the Lindisfarne Gospels—also showing classical influence. Masons from Gaul and Rome built stone churches. In Northumbria stone monuments with figure sculpture and vine-scroll patterns were set up. Churches were equipped with precious objects—some from abroad, some of native manufacture (even in heathen times the English had been skilled metalworkers). Manuscripts and works of art were taken abroad to churches founded by the English missions, and these churches, in turn, became centres of production. The great Sutton Hoo ship burial, discovered in 1939 at the burial site of the East Anglian royal house and perhaps the cenotaph of the bretwalda Raedwald (d. *c.* 625), is further evidence of influences from abroad, revealing important Anglo-Saxon contacts with Scandinavia, Byzantium, France, and the Mediterranean.

Visual arts and architecture

THE HEPTARCHY

The supremacy of Northumbria and the rise of Mercia. When Northumbria became eminent in scholarship, its age of political importance was over. This political dominance had begun when Aethelfrith, ruling over the united Northumbrian kingdoms of Bernicia and Deira, defeated the Dalriadic Scots at Degsastan in 603 and the Welsh at Chester in 613–616. Aethelfrith was himself defeated and killed in 616 by Edwin, the exiled heir to Deira, with the help of Raedwald of East Anglia, then overlord of the southern peoples.

Edwin continued to defeat the Welsh and became the acknowledged overlord of all England except Kent: he annexed the British kingdom of Elmet, invaded North Wales, and captured Anglesey and the Isle of Man. But he fell at Hatfield in 632 before the forces of Cadwallon, king of Gwynedd, and of Penda, a Mercian chieftain. A year later Aethelfrith's son Oswald destroyed Cadwallon and restored the kingdom of Northumbria, and he became overlord of all the lands south of the River Humber. But Mercia was becoming a serious rival; originally a small kingdom in the northwest Midlands, it had absorbed the peoples of the Severn valley, including the Hwicce, a West Saxon people annexed in 628 after a victory by Penda at Cirencester.

Penda threw off Northumbrian control when he defeated and killed Oswald in 641. He drove out Cenwalh of Wessex, who took refuge in East Anglia from 645 to 648. Penda's control of Middle Anglia, where he made his son subking in 653, brought him to the East Anglian frontier; he invaded this kingdom three times, killing three of its kings. He was able to draw an army from a wide area, including East Anglia, when he invaded Northumbria in 654; nevertheless, he was defeated and killed by Oswiu, Oswald's successor.

Rule of Wulfhere

For a short time Oswiu was overlord of southern England, but a Mercian revolt put Penda's son Wulfhere on the throne in 657, and he greatly extended Mercian power to the southeast and south. Wulfhere became overlord of Essex, with London, and of Surrey. He also held the West Saxon lands along the middle Thames and blocked any eastward advance of the West Saxons by capturing the Isle of Wight and the mainland opposite and giving them to his godson, Aethelwalh of Sussex. Yet Wulfhere's reign ended in disaster; the Kentish monk Aedde, in his *Life of St. Wilfrid,* said Wulfhere roused all the southern peoples in an attack on Ecgfrith of Northumbria in 674 but was defeated and died soon after.

Ecgfrith took possession of Lindsey, a section of modern Lincolnshire, but he lost it to Aethelred of Mercia after the Battle of the Trent in 678. Thenceforward Northumbria was no threat to Mercian dominance because it was occupied in fighting the Picts in the north. After Ecgfrith was slain by them in 685, his successors took little part in external affairs.

Yet Mercian power was threatened from the south. Caedwalla had added Surrey, Sussex, and the Isle of Wight to the West Saxon kingdom and thus came near to uniting all lands south of the Thames into a single kingdom that might have held its own against Mercia. But this kingdom was short-lived. Kent became free from foreign interference in 694, two years after the accession of Wihtred, who reestablished the Kentish royal line. Sussex appears again as an independent kingdom; and Caedwalla's successor, Ine, was mainly occupied in extending his territory to the west. After Wihtred's death in 725 and Ine's abdication in 726, both Kent and Wessex had internal troubles and could not resist the Mercian kings Aethelbald and Offa.

The great age of Mercia. Aethelbald succeeded in 716 to the rule of all the Midlands and to the control of Essex and London. By 731 all provinces south of the Humber were subject to him. Some of his charters use a regnal style suited to this dignity, such as "king not only of the Mercians but also of all provinces . . . of the South English" and *rex Britanniae* (a Latinization of bretwalda). Aethelbald held this position, with only occasional warfare, until his death, in 757—far longer than any previous holder of the imperium. St. Boniface praised the good order he maintained in his kingdom, though complaining of his immoral life and his encroachment on church privileges. Aethelbald was murdered by his own household.

Offa did not at once attain the powerful position that later caused Charles the Great (Charlemagne) to treat with him on equal terms; Cynewulf of Wessex recovered West Saxon lands by the middle Thames and did not submit until 779. Offa was overlord in Kent by 764, in Sussex

and the district of Hastings by 771; he apparently lost his authority in Kent after the Battle of Otford in 776 but recovered it in 785. His use of an East Anglian mint shows him supreme there. He claimed greater powers than earlier overlords—subkings among the Hwicce and in Sussex dropped their royal titles and appeared as ealdermen, and he referred to a Kentish king as his thegn. The English scholar Alcuin spoke of the blood shed by Offa to secure the succession of his son, and fugitives from his kingdom sought asylum with Charles the Great. Charles treated Offa as if he were sole king of England, at least of the region south of the Humber; the only other king he acknowledged was the Northumbrian ruler. Offa seemed not to have claimed authority beyond the Humber but instead allied himself with King Aethelred of Northumbria by giving him his daughter in 794.

Charlemagne and Offa

Offa appears on the continental scene more than had any previous English king. Charles wrote to him as "his dearest brother" and wished for a marriage between his own son Charles and Offa's daughter. Offa's refusal, unless Charles let one of his daughters marry Offa's son Ecgfrith, led to a three-year quarrel in which Charles closed his ports to traders from England. This and a letter about regulating trade, written when the quarrel was over, provide evidence for the importance of cross-Channel trade, which was one reason for Offa's reform of the coinage.

Imitating the action of Pepin in 755, Offa took responsibility for the coinage, and thenceforward the king's name normally appeared on coins. But the excellent quality in design and workmanship of his coins, especially those with his portrait, served an additional purpose: they had a propaganda value in bringing home the preeminence of the Mercian king not only to his English subjects but also to people on the Continent. Pope Adrian I regarded Offa with awe and respect.

Because Offa's laws are lost, little is known of his internal government, though Alcuin praises it. Offa was able to draw on immense resources to build a dike to demarcate his frontier against Wales. In the greatness of its conception and the skill of its construction, the dike forms a fitting memorial to him. It probably belongs to his later years, and it secured Mercia from sudden incursions.

The church and scholarship in Offa's time. Northumbria was still preeminent in scholarship, and the fame of the school of York, founded by Bede's pupil Archbishop Egbert, attracted students from the Continent and from Ireland. Eventually it supplied Alcuin to take charge of the revival of learning inaugurated by Charles the Great; Alcuin's writings exercised great influence on theological, biblical, and liturgical studies, and his pupils carried on his work well into the 9th century.

Learning was not confined to Northumbria; one Latin work was produced in East Anglia, and recent attribution of manuscripts to Lichfield suggests that Mercian scholarship has been underestimated. Offa himself took an interest in education, and men from all areas corresponded with the missionaries. The Mercian schools that supplied Alfred with scholars in the 9th century may go back to this period. Vernacular poetry was composed, perhaps including *Beowulf* and the poems of Cynewulf.

A steady advance was made in the creation of parishes, and monasticism flourished and received support from Offa. A great event in ecclesiastical history was the arrival of a papal legation in 787, the first since the conversion. It drew up reforming statutes, which were accepted by the two ecclesiastical provinces, meeting separately under the presidency of Offa and Aelfwald of Northumbria. Offa used the visit to secure the consecration of his son—the first recorded coronation ceremony in England—and also to have Mercia made into a metropolitan province with its see at Lichfield. The latter seemed desirable partly because he disliked the Kentish archbishop of Canterbury, Jaenberht, but also because it would seem fitting to him that the leading kingdom should be free from external interference in ecclesiastical affairs. This move was unpopular with the church, and in 802, when relations with Canterbury had improved, the archbishopric of Lichfield was abolished.

The decline of Mercia and the rise of Wessex. Offa died in 796, and his son died a few weeks later. Cenwulf, their successor, suppressed revolts in Kent and East Anglia, but he never attained Offa's position. Cenwulf allowed Charles to intervene in Northumbria in 808 and restore Eardwulf (who had been driven from his kingdom) to the throne—a unique incident in Anglo-Saxon history. Mercian influence in Wessex was ended when Egbert became king there in 802, though there is no recorded warfare between the kingdoms for many years, during which Egbert conquered Cornwall and Cenwulf fought in Wales. But in 825 Egbert defeated Beornwulf of Mercia and then sent an army into Kent, with the result that he was accepted as king of Kent, Surrey, Sussex, and Essex. In that same year the East Angles threw off the Mercian yoke, killing Beornwulf. In 829 Egbert became ruler of Mercia and all regions south of the Humber, which caused the chronicler to add his name to Bede's list of kings who held the imperium, calling him bretwalda. The Northumbrians accepted Egbert without fighting. Yet he held this proud position only one year; then Wiglaf recovered the Mercian throne and ruled without subjection to Egbert.

Mercian-West Saxon alliance

By this time Danish Viking raids were a grave menace, and Aethelwulf, who succeeded his father Egbert in 839, had the wisdom to see that Mercia and Wessex must combine against the Vikings. Friendly relations between them were established by marriage alliances and by a peaceful settlement of boundaries; this paved the way for the acceptance in 886 of Alfred, king of Wessex, as lord of all the English who had not fallen under Danish rule.

THE PERIOD OF THE SCANDINAVIAN INVASIONS

Viking invasions and settlements. Small scattered Viking raids began in the last years of the 8th century; in the 9th century large-scale plundering incursions were made in Britain and in the Frankish empire as well. Though Egbert defeated a large Viking force in 838 that had combined with the Britons of Cornwall and Aethelwulf won a great victory in 851 over a Viking army that had stormed Canterbury and London and put the Mercian king to flight, it was difficult to deal with an enemy that could attack anywhere on a long and undefended coastline. Destructive raids are recorded for Northumbria, East Anglia, Kent, and Wessex.

A large Danish army came to East Anglia in the autumn of 865, apparently intent on conquest. By 871, when it first attacked Wessex, it had already captured York, been bought off by Mercia, and had taken possession of East Anglia. Many battles were fought in Wessex, including one that led to a Danish defeat at Ashdown in 871. Alfred the Great, a son of Aethelwulf, succeeded to the throne in the course of the year and made peace; this gave him a respite until 876. Meanwhile the Danes drove out Burgred of Mercia, putting a puppet king in his place, and one of their divisions made a permanent settlement in Northumbria.

Alfred's defense of Wessex

Alfred was able to force the Danes to leave Wessex in 877, and they settled northeastern Mercia; but a Viking attack in the winter of 878 came near to conquering Wessex. That it did not succeed is to be attributed to Alfred's tenacity. He retired to the Somerset marshes, and in the spring he secretly assembled an army that routed the Danes at Edington. Their king, Guthrum, accepted Christianity and took his forces to East Anglia, where they settled.

The importance of Alfred's victory cannot be exaggerated. It prevented the Danes from becoming masters of the whole of England. Wessex was never again in danger of falling under Danish control, and in the next century the Danish areas were reconquered from Wessex. Alfred's capture of London in 886 and the resultant acceptance of him by all the English outside the Danish areas was a preliminary to this reconquest. That Wessex stood when the other kingdoms had fallen must be put down to Alfred's courage and wisdom, to his defensive measures in reorganizing his army, to his building fortresses and ships, and to his diplomacy, which made the Welsh kings his allies. Renewed attacks by Viking hosts in 892–896, supported by the Danes resident in England, caused widespread damage but had no lasting success.

Alfred's government and his revival of learning. Good internal government contributed to Alfred's successful resistance to the Danes. He reorganized his finances and the services due from thegns, issued an important code of laws, and scrutinized carefully the exercise of justice. Alfred saw the Viking invasions as a punishment from God, especially because of a neglect of learning, without which men could not know and follow the will of God. He deplored the decay of Latin and enjoined its study by those destined for the church, but he also wished all young freemen of adequate means to learn to read English, and he aimed at supplying men with "the books most necessary for all men to know," in their own language.

Alfred had acquired an education despite great difficulties, and he translated some books himself with the help of scholars from Mercia, the Continent, and Wales. Among them they made available works of Bede and Orosius, Gregory and Augustine, and the *De consolatione philosophiae* of Boethius. Compilation of the *Anglo-Saxon Chronicle* began in his reign. The effects of Alfred's educational reforms can be glimpsed in succeeding reigns, and his works continued to be copied. Only in his attempt to revive monasticism did he achieve little, for the monastic idea had lost its appeal—in England as well as on the Continent—during the Viking Age.

THE ACHIEVEMENT OF POLITICAL UNITY

Edward's successes in England

The reconquest of the Danelaw. When Alfred died in 899, his son Edward succeeded him. A large-scale incursion by the Danes of Northumbria ended in their crushing defeat at Tettenhall in 910. Edward completed his father's plan of building a ring of fortresses around Wessex, and his sister Aethelflaed took similar measures in Mercia. In 912 Edward was ready to begin the series of campaigns by which he relentlessly advanced into the Danelaw (Danish territory in England), securing each advance by a fortress, until he won back Essex, East Anglia, and the east-Midland Danish areas. Aethelflaed moved similarly against the Danish territory of the Five Boroughs (Derby, Leicester, Nottingham, Lincoln, and Stamford). She obtained Derby and Leicester and gained a promise of submission from the Northumbrian Danes before she died in 918. Edward had by then reached Stamford, but he broke off his advance to secure his acceptance by the Mercians at Tamworth and to prevent their setting up an independent kingdom. Then he took Nottingham, and all the Danes in Mercia submitted to him.

Meanwhile another danger had arisen: Norsemen from Ireland had been settling for some time west of the Pennines, and Northumbria was threatened by Raegnald, a Norse leader from Dublin, who made himself king at York in 919. Edward built fortresses at Thelwall and Manchester, and in 920 he received Raegnald's submission, along with that of the Scots, the Strathclyde Welsh, and all the Northumbrians. Yet Norse kings reigned at York intermittently until 954.

The kingdom of England. Athelstan succeeded his father Edward in 924. He made terms with Raegnald's successor Sihtric and gave him his sister in marriage. When Sihtric died in 927, Athelstan took possession of Northumbria, thus becoming the first king to have direct rule of all England. He received the submission of the kings of Wales and Scotland and of the English ruler of Northumbria beyond the Tyne.

Athelstan's ascendancy over all England

Athelstan was proud of his position, calling himself "king of all Britain" on some of his coins and using in his charters flamboyant rhetoric carrying the same message; he held great courts attended by dignitaries from all over England and by Welsh kings; he subjected the Welsh to tribute and quelled a revolt of the Britons of Cornwall. His sisters were married to continental princes—Charles the Simple, king of the Franks; Otto, son of Henry the Fowler; and Hugh, duke of the Franks. Among those brought up at his court were Louis, Charles's son; Alan of Brittany, Athelstan's godson; and Haakon, son of Harald Fairhair of Norway; they all returned to win their respective inheritances with his support. He was a generous donor to continental and English churches. But Athelstan is remembered chiefly as the victor at Brunanburh, against a combine of Olaf Guthfrithson, king of Dublin; Owain of Strathclyde; and Constantine, king of the Scots, whom Athelstan had defeated in 934. They invaded England in 937, and their defeat is celebrated by a poem in the *Anglo-Saxon Chronicle.*

Immediately after Athelstan's death in 939 Olaf seized not only Northumbria but also the Five Boroughs. By 944 Athelstan's successor, his younger brother Edmund, had regained control, and in 945 Edmund conquered Strathclyde and gave it to Malcolm of Scotland. But Edmund's successor, Eadred, lost control of Northumbria for part of his reign to the Norse kings Erik Bloodax (son of Harald Fairhair) and Olaf Sihtricson. When Erik was killed in 954, Northumbria became a permanent part of the kingdom of England.

By becoming rulers of all England, the West Saxon kings had to administer regions with variant customs, governed under West Saxon, Mercian, or Danish law. In some parts of the area of Danish occupation, especially in northern England and the district of the Five Boroughs, the evidence of place-names, personal names, and dialect seems to indicate dense Danish settlement, but this has been seriously questioned; many "Danish" features are also found in Anglo-Saxon areas, and Danish names do not always prove Danish institutions. Moreover, the older Anglo-Saxon regions, such as Mercia, which often cut across both Danish and English areas, were politically more significant. Money, however, was calculated in marks and ores instead of shillings in Danish areas, and arable land was divided into plowlands and oxgangs instead of hides and virgates in the northern and northeastern parts of the Danelaw. Most important was the presence in some areas of a number of small landholders with a much greater degree of independence than their counterparts elsewhere; many ceorls had so suffered under the Danish ravages that they had bought a lord's support by sacrificing some of their independence. Excavations (1976–81) have shown 10th-century Jorvik (York), a Danish settlement, to have been a centre of international trade, economic specialization, and town planning; it was on its way to becoming by 1086 (in the Domesday survey) one of Europe's largest cities, numbering at least 2,000 households.

The kings did not try to eradicate the local peculiarities. King Edgar (reigned 959–975) expressly granted local autonomy to the Danes. But from Athelstan's time it was decreed that there was to be one coinage for all the king's dominion, and a measure of uniformity in administrative divisions was gradually achieved. Mercia became divided into shires on the pattern of those of Wessex. It is uncertain how early the smaller divisions of the shires were called "hundreds," but they now became universal (except in the northern Danelaw, where an area called a wapentake carried on its fiscal and jurisdictional functions). An ordinance of the mid-10th century laid down that the court in each hundred (called "hundred courts") must meet every four weeks to handle local legal matters, and Edgar enjoined that the shire courts must meet twice a year and the borough courts three times. This pattern of local government survived the Norman Conquest.

The church and the monastic revival. To those who judged the church solely by the state of its monasteries, the first half of the 10th century seemed a period of inertia. In fact, the great tasks of converting the heathen settlers, restoring ecclesiastical organization in Danish areas, and repairing the damages of the invasions elsewhere must have absorbed much energy. Even so, learning and book production were not at so low an ebb as monastic reformers claimed. Moreover, new monasteries were founded and benefactions were made to older ones, even though, by post-revival standards, none of these monasteries was enforcing a strict monastic rule and several benefactions were held by secular priests. Alfred had failed to arouse much enthusiasm for monasticism. The movement for reform began in England about 940 and soon came under the influence of reforms in Fleury and Lorraine. King Edgar, an enthusiastic supporter, promoted the three chief reformers to important positions—Dunstan to Canterbury, Aethelwold to Winchester, and Oswald to Worcester and later to York. The secular clergy were violently ejected from

Movement for monastic reform

Winchester and some other places; Oswald gradually replaced them with monks at Worcester. All three reformers founded new houses, including the great monasteries in the Fenlands, where older houses had perished in the Danish invasion; but Oswald had no success in Northumbria. The reformers, however, were concerned with more than monasticism—they paid great attention to other needs of their dioceses; the scholars Abbot Aelfric and Archbishop Wulfstan, trained by the reformers, directed much of their writings to improving the education and morals of the parish clergy and, through them, of the people.

The monastic revival resulted in a great revival of both vernacular and Latin literature, of manuscript production and illumination, and of other forms of art. It reached its zenith in the troubled years of King Ethelred II (reigned 978–1016), after a brief, though violent, reaction to monasticism following Edgar's death. In the 11th century monasteries continued to be productive and new houses were founded; there was also a movement to impose a communal life on bodies of secular priests and to found houses of secular canons.

THE ANGLO-DANISH STATE

The Danish conquest and the reigns of the Danish kings. Ethelred succeeded as a child in 978, after the murder of his stepbrother Edward. He took the throne in an atmosphere of insecurity and distrust, which partly accounts for the incompetence and treachery rife in his reign. Viking raids began in 980 and steadily increased in intensity. They were led by formidable leaders: from 991 to 994 by Olaf Tryggvason, later king of Norway, and frequently from 994 by Sweyn, king of Denmark. Ethelred's massacre of the Danes in England on St. Brice's Day, 1002, called for vengeance by Sweyn and, from 1009 to 1012, by a famous Viking, Thorkell the Tall. In 1013 the English, worn out by continuous warfare and heavy tributes to buy off the invaders, accepted Sweyn as king. Ethelred, his wife Emma, and his younger sons sought asylum with Richard, duke of Normandy, brother of Emma. Ethelred was recalled to England after Sweyn's death in 1014; but Sweyn's son Canute (Cnut) renewed the invasions and, in spite of valiant resistance by Ethelred's son and successor, Edmund, obtained half of England after a victory at Ashingdon in October 1016 and the rest after Edmund's death that November.

Revival of Danish attacks

Canute rewarded some of his followers with English lands and ruthlessly got rid of some prominent Englishmen, among them Edmund's brother Edwy. (Edmund's infant sons, however, were carried away to safety in Hungary.) Yet Canute's rule was not tyrannical, and his reign was remembered as a time of good order. The Danish element in his entourage diminished; and the Englishmen Leofric, Earl of Mercia, and Godwine, Earl of Wessex, became the most powerful magnates. Canute married Ethelred's widow, Emma, thus removing the danger of Norman support for her sons by Ethelred. Canute fought a successful campaign in Scotland in 1031, and Englishmen were drawn into his wars in Scandinavia, which made him lord of Norway. But at home there was peace. Probably under the influence of Archbishop Wulfstan he became a stout supporter of the church, which in his reign had the vitality to engage in missionary work in Scandinavia. Religious as well as political motives may have caused his pilgrimage to Rome in 1027 to attend the coronation of the emperor Conrad; from the pope, the emperor, and the princes whom he met he obtained concessions for English pilgrims and traders going to Rome. Canute's laws, drafted by Archbishop Wulfstan, are mainly based on those of earlier kings, especially Edgar.

Already in 1018 the English and Danes had come to an agreement "according to Edgar's law." No important changes were made in the machinery of government except that small earldoms were combined to make great earldoms, a change that placed much power in the hands of their holders. No attempt was made to restore the English line when Canute died in 1035; he was followed by his sons Harold and Hardecanute, whose reigns were unpopular. Denmark passed to Sweyn, son of Canute's sister Estrith, in 1043. Meanwhile the Norwegians in 1035 had driven out another Sweyn, the son whom Canute had set to rule over them with his mother, Aelfgifu, and had elected Magnus.

Canute's successors

The close links with Scandinavia had benefited English trade, but they left one awkward heritage: Hardecanute and Magnus made an agreement that if either died without a son, the survivor was to succeed to both kingdoms. Hardecanute died without a son in 1042, but he was succeeded by Ethelred's son Edward, who was known as the Confessor or the Saint because of his reputation for chastity. Magnus was prevented by trouble with Denmark from invading England as he intended in 1046; but Harold Hardraada inherited Magnus' claim to the English throne, and he came to enforce it in 1066.

The reign of Edward the Confessor and the Norman Conquest. It is easy to regard the years of Edward's rule simply as a prelude to the catastrophe of 1066, yet there are other aspects of his reign. Harrying caused by political disturbances or by incursions of the Scots or Welsh was only occasional and localized; friendly relations were usually maintained with Malcolm of Scotland, whom Earl Siward of Northumbria had supported against Macbeth in 1054; and in 1063 the victories of Harold, Earl of Wessex, and his brother Tostig ended the trouble from Wales. The normal course of administration was maintained, with efficient mints, writing office, taxation system, and courts of justice. Trade was prosperous. The church contained several good and competent leaders, and bad appointments—like those of the Normans, Ulf to Dorchester and Robert to London and Canterbury, and of Stigand to Winchester—were the exception. Scholarship was not in decline, and manuscripts were produced in great number. English illumination and other forms of art were admired abroad.

The troubles of the reign came from the excessive power concentrated in the hands of the rival houses of Leofric of Mercia and Godwine of Wessex and from resentment caused by the king's introduction of Norman friends, though their influence has sometimes been exaggerated. A crisis arose in 1051 when Godwine defied the king's order to punish the men of Dover, who had resisted an attempt by Eustace of Boulogne to quarter his men on them by force. The support of Earl Leofric and Earl Siward enabled Edward to secure the outlawry of Godwine and his sons; and William of Normandy paid Edward a visit during which Edward may have promised William succession to the English throne, although this Norman claim may have been mere propaganda. Godwine and his sons came back the following year with a strong force, and the magnates were not prepared to engage them in civil war but forced the king to make terms. Some unpopular Normans were driven out, including Archbishop Robert, whose archbishopric was given to Stigand; this act supplied one excuse for the papal support of William's cause.

Resentment of Normans

Harold succeeded his father Godwine as earl of Wessex in 1053; Tostig was made earl of Northumbria in 1055; and their younger brothers were also provided with earldoms. To settle the question of succession, negotiations were begun in 1054 to bring Edward, Edmund's son (nephew of Edward the Confessor), from Hungary; but Edward died in 1057, leaving a son, Edgar Aetheling, then a child, who was passed over in 1066. In about 1064 Harold of Wessex, when visiting Normandy, swore to support William's claim. Only Norman versions of the incident survive and the true circumstances cannot be ascertained, but William used Harold's broken oath to help secure papal support later. In 1065 Harold acquiesced in the appointment of Morcar, brother of Edwin, Earl of Mercia, to replace Tostig when the Northumbrians revolted against him, and thus Harold turned his brother into an enemy. King Edward, when dying, named Harold to succeed him, and, after overcoming Northumbrian reluctance with the help of Bishop Wulfstan of Worcester, Harold was universally accepted.

Harold might have proved an effective ruler, but the forces against him were too strong. The papacy, without hearing the defense in favour of Harold's succession, gave its blessing to an invasion of a people who had always been distinguished for their loyalty to Rome, and this

papal support helped William to collect his army widely. The threat from Harold III Hardraade, who was joined by Tostig, prevented Harold from concentrating his forces in the south and took him north at a critical moment. He fought at Hastings only 24 days after the armies of Mercia and Northumbria had been put out of action by enormous losses at Fulford and only 19 days after he had defeated and killed Harold III Hardraade and Tostig at Stamford Bridge. Harold was slain at Hastings, and on Christmas Day, 1066, William of Normandy was crowned king of England. Although the Anglo-Saxon fighting force was perhaps the best in Europe and the defeat at Hastings due largely to a series of historical accidents, it is not difficult to understand the English chronicler's view that God was angry with the English people. (D.W./W.A.C.)

The crowning of William

The Normans (1066–1154)

WILLIAM I (1066–87)

The Norman Conquest has long been argued about. The question has been whether William I introduced fundamental changes in England or based his rule solidly on Anglo-Saxon foundations. A particularly controversial issue has been the introduction of feudalism. On balance, the debate has favoured dramatic change while also granting that in some respects the Normans learned much from the English past. Yet William replaced his initial policy of trying to govern through Englishmen with an increasingly thoroughgoing Normanization.

Resistance and rebellion. The Conquest was not achieved at a single stroke. In 1068 Exeter rose against the Normans, and a major rising began in the north. A savage campaign in 1069–70, the so-called harrying of the north, emphasized William's military supremacy and his brutality. A further English rising in the Fens achieved nothing. In 1075 William put down rebellion by the earls of Hereford, Norfolk, and Northumbria. The latter, the last surviving English earl, was executed for treason.

The introduction of feudalism. The Conquest resulted in the subordination of England to a Norman aristocracy. William probably distributed estates to his followers on a piecemeal basis as lands came into his hands. He granted lands directly to fewer than 180 men, making them his tenants in chief. Their estates were often well distributed, consisting of manors scattered through a number of shires. In vulnerable regions, however, compact blocks of land were formed, clustered around castles. The tenants in chief owed homage and fealty to the king and held their land in return for military service. They were under obligation to supply a certain number of knights for the royal feudal host—a number that was not necessarily related to the quantity or quality of land held. Early in the reign many tenants in chief provided knights from their own households to meet demands for service, but they soon began to grant some of their own lands to knights who would serve them just as they in turn served the king. They could not, however, use their knights for private warfare, which, in contrast to Normandy, was forbidden in England. In addition to drawing on the forces provided by feudal means, William made extensive use of mercenary troops to secure the military strength he needed. Castles, which were virtually unknown in pre-Conquest England and could only be built with royal permission, provided bases for administration and military organization. They were an essential element in the Norman settlement of England.

Government and justice. William hoped to be able to rule England in much the same way as his Anglo-Saxon predecessors had done, though in many respects the old institutions and practices had to be changed in response to the problems of ruling a conquered land. The Anglo-Saxon witan, or council, became the king's *curia regis*, a meeting of the royal tenants in chief, both lay and ecclesiastical. William was said by chroniclers to have held full courts three times a year, at Christmas, Easter, and Whitsuntide, to which all the great men of the realm were summoned and at which he wore his crown. These were similar to the great courts he held in Normandy. Inevitably there were many disputes over land, and the *curia regis* was where justice was done to the great tenants in chief. William himself is said to have sat one Sunday "from morn till eve" to hear a plea between William de Braose and the abbot of Fécamp.

Curia regis

William at first did little to change Anglo-Saxon administrative organization. The royal household was at the centre of royal government, and the system, such as it was, under Edward the Confessor had probably been quite similar to that which existed in Normandy at the same period, although the actual titles of the officers were not the same. Initially under William there also was little change in personnel. But, by the end of his reign, all important administrative officials were Norman, and their titles corresponded to those in use in Normandy. There were a steward, a butler, a chamberlain, a constable, a marshal, and a head of the royal scriptorium, or chancellor. This scriptorium was the source from which all writs (*i.e.*, written royal commands) were issued. At the start of William's reign the writs were in English, and by the end of it, in Latin.

In local government the Anglo-Saxon shire and hundred courts continued to function as units of administration and justice, but with important changes. Bishops and earls ceased to preside over the shire courts. Bishops now had their own ecclesiastical courts, while earls had their feudal courts. But although earls no longer presided over shire courts, they were entitled to take a third of the proceeds coming from them. The old Anglo-Saxon office of sheriff was transformed into a position resembling that of the Norman *vicomte*, as native sheriffs were replaced by Norman nobles. They controlled the shire and hundred courts, were responsible for collecting royal revenue, and controlled the royal castles that had been built both to subdue and protect the country.

William made the most of the financial system he had inherited. In addition to customary dues, such as revenues from justice and income from royal lands, his predecessors had been able to levy a geld, or tax, assessed on the value of land and originally intended to provide funds to buy off Danish invaders. The Confessor had abandoned this tax, but the Conqueror collected it at least four times. Profits from the ample royal estates must have been significant, along with those from royal mints and towns.

The Conqueror greatly strengthened the administration of justice in his new land. He occasionally appointed justiciars to preside over local cases and at times named commissioners to act as his deputies in the localities. There were a number of great trials during the reign. The most famous of them was the trial at Pinnenden Heath of a case between Lanfranc, archbishop of Canterbury, and the king's half brother, Odo, bishop of Bayeux and earl of Kent. Not only all the Normans of the shire but also many Englishmen, especially those learned in the customary law, attended. On occasion jurors were summoned to give a collective verdict under oath. Historians have debated as to whether juries were introduced as a result of the Viking conquests or were a Norman innovation, derived from Carolingian practice in France. Whichever argument is correct, it is evident that, under the Normans, juries came into more frequent use. William introduced one measure to protect his followers: he made the local community of the hundred responsible for the murder of any Norman.

Juries

Church–state relations. The upper ranks of the clergy were Normanized and feudalized, following the pattern of lay society. Bishops received their lands and the symbols of their spiritual office from the king. They owed knight service and were under firm royal control. Sees were reorganized, and most came to be held by continental clergy. In 1070 Lanfranc replaced Stigand as archbishop of Canterbury. An ecclesiastical lawyer, teacher, and church statesman, Lanfranc, a native of Italy, had been a monk at Bec and an abbot of Saint Stephen's at Caen. Lanfranc and William understood each other and worked together to introduce discipline and order into the English church. The see of York was subordinated to Canterbury, and efforts were made to bring the ecclesiastical affairs of Ireland and Scotland under Lanfranc's control. Several church councils were held in England to legislate for the English church, as similar councils did in Normandy. William

denied that he owed homage or fealty to the pope for his English lands, although he acknowledged papal support in winning the new realm. William and Lanfranc resisted Pope Gregory VII's claim to papal supremacy: the king decreed that without his consent no pope was to be recognized in England, no papal letter was to be received, no church council was to legislate, and no baron or royal official was to be excommunicated. During William's reign the controversy over the right of lay rulers to invest ecclesiastics with the symbols of their office did not affect England, in contrast to other parts of Latin Christendom.

William's accomplishments. At Christmas 1085 William had "deep speech" with his council and as a result ordered a general survey of the land to be made. Historians have debated the purpose of this "Domesday" survey, some seeing it as primarily a tax assessment, others emphasizing its importance as a basis for assignment of feudal rights and duties. Its form owed much to Anglo-Saxon precedent, but within each county section it was organized on a feudal basis. It was probably a multipurpose document with the main emphasis on resources for taxation. It was incomplete, for the far north of England, London, and Winchester were not included, while the returns for Essex, Norfolk, and Suffolk were not condensed into the same form as was used for the rest of the country. Domesday is a unique record and offers rich materials for research.

The Domesday survey

One policy that caused deep resentment under William I, and even hatred under his successor William II, was the taking over of vast tracts of land for the king's forest. In some areas whole villages were destroyed and the people driven out; elsewhere, people living in forest areas, though not necessarily removed, were subjected to a severe system of law with drastic penalties for poaching.

William the Conqueror is presented in contemporary chronicles as a ruthless tyrant who rigorously put down rebellion and devastated vast areas, especially in his pacification of the north in 1069–70. He was, however, an able administrator. Perhaps one of his greatest contributions to England's future was the linking up of England with continental affairs. If the country had been conquered again by the Danes, as seemed possible, it might have remained in a backwater of European development. In the event, England was linked, economically and culturally, to France and continental Europe. The aristocracy spoke French, while Latin was the language of the church and the administration.

THE SONS OF WILLIAM I

William II Rufus (1087–1100). Under William I's two sons William II Rufus and Henry I, strong, centralized government continued, and England's link with Normandy was strengthened. Rebellion by Norman barons, led by the king's half uncles, Odo of Bayeux and Robert of Mortain, was soon put down by William II, who made promises of good government and relief from taxation and the severity of the forest laws. Odo of Bayeux was banished, and William of St. Calais, bishop of Durham, tried for treason. As an ecclesiastic he rejected the jurisdiction of the king's court. But Lanfranc pointed out that it was not as a churchman but as lord of his temporal fiefs that he was being tried. He was finally allowed to leave the country, in return for surrender of his fiefs.

William II's main preoccupation was to win Normandy from his elder brother Robert. After some initial skirmishing, William's plans were furthered by Robert's decision to go on crusade in 1096. Robert mortgaged his lands to William for 10,000 marks, which was raised in England by drastic and unpopular means. In his last years William campaigned successfully in Maine and the French Vexin so as to extend the borders of Normandy. His death was the result of an "accident" possibly engineered by his younger brother Henry: he was shot with an arrow in the New Forest. Henry, who was conveniently with the hunting party, rode posthaste to Winchester, seized the treasury, and was chosen king the next day.

Henry I (1100–35). A good politician and administrator, Henry I was the ablest of the Conqueror's sons. At his coronation on Aug. 5, 1100, he issued a charter intended to win the support of the nation. This propaganda document, in which Henry promised to give up many practices of the past, demonstrates how oppressive Norman government had become. Henry promised not to exploit church vacancies, as his brother had done, and guaranteed that reliefs, sums paid by feudal vassals when they took over their fathers' estates, would be "just and legitimate." He also promised to return to the laws of Edward the Confessor, though this cannot have been intended literally.

Henry's coronation charter

Following the suppression of rebellion in England, the conquest of Normandy was an important priority for Henry. By 1105 he took the offensive, and in September 1106 he won a decisive battle at Tinchebray that gave him control of the whole of Normandy. Robert was captured and was to spend the rest of his 80 years in castle dungeons. His son, William Clito, escaped and remained until his death in 1128 a thorn in Henry's flesh. Success in Normandy was followed by wars against Louis VI of France, but by 1120 Henry was everywhere successful in both diplomacy and war. He had arranged a marriage for his only legitimate son, William, to Matilda, daughter of Fulk of Anjou, and had received Fulk's homage for Maine. Pope Calixtus II, his cousin, gave him full support for his control of Normandy on condition that his son William should do homage to the French king.

Relations with the church had not always been easy. Henry had inherited from William II a quarrel with the church that became part of the Europe-wide Investiture Controversy. After Lanfranc's death William had delayed appointing a successor, presumably for the privilege of exploiting the resources of the archbishopric. After four years, during a bout of illness, he appointed Anselm of Bec, one of the great scholars of his time (1093). Anselm did homage for his temporalities, but whether or not he was ever invested with the symbols of spiritual office by the king is not clear. Papal confirmation was complicated by the fact that there were two claimants to the papal throne. Anselm refused to accept a decision made by the king's supporters and insisted on receiving his pallium from Urban II, a reform pope in the tradition of Gregory VII, rather than from the imperial nominee, Clement III. Conflict between king and archbishop flared up again in 1097 over what William considered to be an inadequate Canterbury contingent for his Welsh war. The upshot was that Anselm went into exile until William's death. At Rome he heard new papal decrees against lay investiture.

Anselm of Bec

Anselm supported Henry's bid for the throne and returned from exile in 1100. Almost immediately he quarreled with Henry when the king asked him to do homage and to receive his archbishopric from his hands. After various ineffective appeals to Rome, Anselm again went into exile. A compromise was finally arranged in 1107, when it was agreed that the king would surrender investiture with the symbols of spiritual office in return for an agreement that he should supervise the election of the archbishop and take homage for the temporalities before investiture with the spiritual symbols took place. It was said that the concession cost the king "a little, perhaps, of his royal dignity, but nothing of his power to enthrone anyone he pleased."

Henry continued and extended the administrative work of his father. His frequent absences from England prompted the development of a system that could operate effectively in his absence, under the guidance of such men as Roger, bishop of Salisbury. The exchequer was developed as a department of government dealing with royal revenues, and the first record of the sheriffs' regular accounting at the exchequer, or Pipe Roll, to survive is that of 1129–30. Justices with wide-ranging commissions were sent out into the shires to reinforce local administration and to inquire into crown pleas, royal revenues, and other matters of interest and profit for the king. Henry's government was highly efficient, but it was also harsh and demanding.

During the last 15 years of his reign the succession was a major issue. William, Henry's only legitimate son, was drowned in 1120, leaving Henry's daughter Matilda, wife of the German emperor Henry V, as heir. When Henry V died in 1125, Matilda returned to England. Henry I persuaded his barons to swear an oath in her support but did

not consult them over her second marriage to Geoffrey of Anjou, who at 14 was 11 years her junior. Within a year Geoffrey repudiated Matilda, but during a temporary reconciliation, Matilda and Geoffrey had three children.

Henry was a skilled politician, adept at using the levers of patronage. Men such as Geoffrey de Clinton, the royal chamberlain, owed much to the favours they received from the king, and they served him well in return. There was tension between the established nobility and the "new men" raised to high office by the king, but Henry maintained control with great effect and distributed favours evenhandedly. In England his rule, particularly when seen in retrospect, was characterized by peace, order, and justice. He died, probably of a heart attack, on Dec. 1, 1135.

THE PERIOD OF ANARCHY (1135–54)

Matilda and Stephen. Henry I's death precipitated a 20-year crisis whose immediate cause was a succession dispute. But there has been much debate among historians as to whether the problems of these years were the result of some deeper malaise.

No one was enthusiastic about accepting Matilda as queen, especially as her husband, Geoffrey of Anjou, was actually at war with Henry at the time of his death. Robert, Earl of Gloucester, one of Henry's many illegitimate sons, was an impressive candidate for the throne, as were Henry's nephews, Theobald and Stephen of Blois. The outcome of the struggle in 1135 was unexpected: while Theobald, the elder brother, was receiving the homage of continental vassals for Normandy, Stephen took ship for England and claimed the throne. Having secured the treasury at Winchester, he was crowned on December 22.

Stephen had been quick and resolute in securing the crown. But after the first flush of victory he made concessions that, instead of winning him support, served to expose his weakness. One such concession was to King David of Scotland, who was also the Earl of Huntingdon in England. When David learned of Stephen's succession, he crossed the border by force. He was effectively bought off by Stephen's agreeing that David's son Henry should receive Carlisle, Doncaster, and the honour of Huntingdon. Stephen obtained the support of Robert of Gloucester by a lavish charter. He also granted a charter to the church forbidding simony and confirmed the rights of church courts to all jurisdiction over clerics. Stephen's lavish appointment of new earls (19 in the course of the reign) was intended in part as a way of undermining the power of the sheriffs and constituted a shift of power away from the centre. Expenditure in Stephen's early years was heavy and achievements few.

Stephen's clash with King David

Stephen soon alienated the church. Much power in central government had been concentrated in the hands of Roger, bishop of Salisbury, and his family. One of Roger's nephews was bishop of Ely, and another was bishop of Lincoln. This was resented by the Beaumont family, headed by the Earl of Leicester, and their allies, who formed a powerful court faction. They planned the downfall of the bishops, and, when a council meeting was held at Oxford in June 1139, they seized on the opportunity provided by a brawl in which some of Roger's men were involved. Rumours of treason were spread, and Stephen demanded that the bishops should make satisfaction. When they did not do so, he ordered their arrest. Thenceforth Stephen was in disfavour with the clergy; he had already forfeited the support of his brother Henry of Blois, bishop of Winchester, by failing to make him archbishop of Canterbury in 1137. As papal legate, Henry was to be the most influential member of the clergy in the realm.

Civil war. Matilda did not land in England until 1139. She and her half brother Robert of Gloucester established themselves in the southwest; Stephen's main strength lay in the east. In 1141 Stephen was defeated and taken prisoner at the battle of Lincoln, but Matilda alienated the Londoners and lost support. When Stephen was exchanged for Robert of Gloucester, who was captured at Winchester, Matilda's fortunes waned. The Angevin cause, however, prospered in Normandy. Although Matilda's son, Henry, mounted an unsuccessful invasion from Normandy in 1147, in 1153 he carried out a vigorous and effective campaign. Stephen, saddened by the death of his elder son Eustace, agreed to a compromise peace. He was to remain king, but he recognized Henry as his heir.

Capture of King Stephen

One chronicler said, "In this king's time there was nothing but disturbance and wickedness and robbery." Though this was an exaggeration, it is clear that disorder was widespread, with a great many adulterine castles built (that is, unlicensed castles). It was possible for the earls of Chester and Leicester to make a treaty without any reference to royal authority. Stephen's government lost control of much of England, and power was fragmented and decentralized.

There has been much debate as to why men fought in the civil war. It was much more than a simple succession dispute and can be seen as a natural reaction both to the strong, centralized government of Henry I and to the weak rule of Stephen. The aim of many magnates was to recover lands and offices to which they considered they had hereditary rights: much land had changed hands under Henry I. Men such as Ranulf de Gernons, 4th Earl of Chester, and Geoffrey de Mandeville, 1st Earl of Essex, changed sides frequently, obtaining fresh grants each time. Essex wanted to recover the lands and positions his grandfather had held. Most men, however, probably did not want to demolish royal government but rather wished to control and profit from it. The institutions of government did not disappear altogether. The period of the "anarchy" strengthened feudal principles of succession to land, but such offices as those of sheriff and castellan did not become hereditary.

ENGLAND IN THE NORMAN PERIOD

Despite, or perhaps in part because of, the political strains of this period, these were constructive years. The economy, for which Domesday Book is a magnificent source, was essentially agrarian, the main unit being the manor, where the lord's land (or demesne) was worked by unfree peasants who held their land in return for performing labour services. Towns, notably London, flourished, and many received new privileges based on continental practice. The church benefited from closer connections with the Continent in many ways. One such benefit was the arrival of new religious orders: the first Cluniac house was established at Lewes in 1077, and the Cistercians came to England in 1129. A great many Augustinian houses were founded in the first part of the 12th century. Imposing buildings such as Durham Cathedral and the Tower of London give eloquent testimony to the architectural achievement of the Normans, while the illuminated Winchester Bible and Psalter, made for Henry of Blois, bear witness to the artistic excellence of the age.

Social developments

The early Plantagenets

HENRY II (1154–89)

Matilda's son Henry Plantagenet, the first and greatest of three Angevin kings of England, succeeded Stephen in 1154. Aged 21, he already possessed a reputation for restless energy and decisive action. He was to inherit vast lands. As heir to his mother and to Stephen he held England and Normandy; as heir to his father he held Anjou (hence Angevin), Maine, and Touraine; as heir to his brother Geoffrey he obtained Brittany; as husband of Eleanor, the divorced wife of Louis VII of France, he held Aquitaine, the major part of southwestern France. Altogether his holdings in France were far larger than those of the French king. They have become known as the Angevin empire, although Henry never in fact claimed any imperial rights or used the title of emperor. From the beginning Henry showed himself determined to assert and maintain his rights in all his lands. In England this meant reasserting the centralized power of his grandfather, Henry I. His success in these aims is the measure of his greatness.

Government of England. In the first decade of his reign Henry was largely concerned with continental affairs, though he made sure that the adulterine castles in England were destroyed. Many of the earldoms created in the anarchy of Stephen's reign were allowed to lapse. Major change in England began in the mid-1160s. The Assize

of Clarendon of 1166, and that of Northampton 10 years later, promoted public order. Juries were used to provide evidence of what crimes had been committed and to bring accusations. New forms of legal action were introduced, notably the so-called possessory assizes, which determined who had the right to immediate possession of land, not who had the best fundamental right. That could be decided by the grand assize, by means of which a jury of 12 knights would decide the case. The use of standardized forms of writ greatly simplified judicial administration. "Returnable" writs, which had to be sent back by the sheriffs to the central administration, enabled the crown to check that its instructions were obeyed. An increasing number of cases came before royal courts rather than private feudal courts. Henry I's practice of sending out itinerant justices was extended and systematized. In 1170 a major inquiry into local administration, the Inquest of Sheriffs, was held, and many sheriffs were dismissed.

The Inquest of Sheriffs

There were important changes to the military system. In 1166 the tenants in chief were commanded to disclose the number of knights enfeoffed on their lands so that Henry could take proper financial advantage of changes that had taken place since his grandfather's day. Scutage (money payment in lieu of military service) was an important source of funds, and Henry preferred scutage to service because mercenaries were more efficient than feudal contingents. In the Assize of Arms of 1181 Henry determined the arms and equipment appropriate to every free man, based on his income from land. This measure, which could be seen as a revival of the principles of the Anglo-Saxon fyrd, was intended to provide for a local militia, which could be used against invasion, rebellion, or for peacekeeping.

Struggle with Thomas Becket. Henry attempted to restore the close relationship between church and state that had existed under the Norman kings. His first move was the appointment in 1162 of Thomas Becket as archbishop of Canterbury. Henry assumed that Becket, who had served efficiently as chancellor since 1155 and been a close companion to him, would continue to do so as archbishop. Becket, however, disappointed him. Once appointed archbishop, he became a militant defender of the church against royal encroachment and a champion of the papal ideology of ecclesiastical supremacy over the lay world. The struggle between Henry and Becket reached a crisis at the Council of Clarendon in 1164. In the Constitutions of Clarendon Henry tried to set down in writing the ancient customs of the land. The most controversial issue proved to be that of jurisdiction over "criminous clerks" (clerics who had committed crimes); the king demanded that such men should, after trial in church courts, be sent for punishment in royal courts.

Becket initially accepted the Constitutions but would not set his seal to them. Shortly thereafter, however, he suspended himself from office for the sin of yielding to the royal will in the matter. Although he failed to obtain full papal support at this stage, Alexander III ultimately came to his aid over the Constitutions. Later in 1164 Becket was charged with peculation of royal funds when chancellor. After Becket had taken flight for France, the king confiscated the revenues of his province, exiled his friends, and confiscated their revenues. In 1170 Henry had his eldest son crowned king by the archbishop of York, not Canterbury, as was traditional. Becket, in exile, appealed to Rome and excommunicated the clergy who had taken part in the ceremony. A reconciliation between Becket and Henry at the end of the same year settled none of the points at issue. When Becket returned to England, he took further measures against the clergy who had taken part in the coronation. In Normandy the enraged king, hearing the news, burst out with the fateful words that incited four of his knights to take ship for England and murder the archbishop in Canterbury Cathedral.

Becket's murder

Almost overnight the martyred Thomas became a saint in the eyes of the people. Henry repudiated responsibility for the murder and reconciled himself with the church. But despite various royal promises to abolish customs injurious to the church, royal control of the church was little affected. Henceforth criminous clerks were to be tried in church courts, save for offenses against the forest laws. Disputes over ecclesiastical patronage and church lands that were held on the same terms as lay estates were, however, to come under royal jurisdiction. Finally Henry did penance at Canterbury, allowing the monks to scourge him. But with Becket out of the way, it proved possible to negotiate most of the points at issue between church and state. The martyred archbishop, however, was to prove a potent example for future prelates.

Rebellion of Henry's sons and Eleanor of Aquitaine. Henry's sons, urged on by their mother and by a coalition of his enemies, raised a rebellion throughout his domains in 1173. King William I the Lion of Scotland joined the rebel coalition and invaded the north of England. Lack of cooperation among the rebels, however, enabled Henry to defeat them one at a time with a mercenary army. The Scottish king was taken prisoner at Alnwick. Queen Eleanor was retired to polite imprisonment for the rest of Henry's life. The king's sons and the baronial rebels were treated with leniency, but many baronial castles were destroyed following the rising. A brief period of amity between Henry and Louis of France followed, and the years between 1175 and 1182 marked the zenith of Henry's prestige and power. In 1183 the younger Henry again tried to organize opposition to his father, but he died in June of that year. Henry spent the last years of his life locked in combat with the new French king, Philip II Augustus, with whom his son Richard had entered into an alliance. Even his youngest son, John, deserted him at the end.

RICHARD I (1189–99)

Henry II died in 1189, an embittered old man. He was succeeded by his son Richard I, nicknamed the Lion-Heart. Richard, a renowned and skillful warrior, was mainly interested in the Crusade to recover Jerusalem and in the struggle to maintain his French holdings against Philip Augustus. He spent only about six months of his 10-year reign in England. During his frequent absences he left a committee in charge of the realm. The chancellor, William Longchamp, bishop of Ely, dominated the early part of the reign until forced into exile by baronial rebellion in 1191. Walter of Coutances, archbishop of Rouen, succeeded Longchamp, but the most important and able of Richard's ministers was Hubert Walter, archbishop of Canterbury, justiciar from 1193 to 1198, and chancellor from 1199 to 1205. With the king's mother, Eleanor, he put down a revolt by Richard's brother John in 1193 with strong and effective measures. But when Richard returned from abroad, he forgave John and promised him the succession.

This reign saw some important innovations in taxation and military organization. Warfare was expensive, and in addition Richard was captured on his return from the Crusade by Leopold V of Austria and held for a high ransom of 150,000 marks. Various methods of raising money were tried: an aid, or scutage; a carucage, or tax on plow lands; a general tax of a fourth of revenues and chattels (this was a development of the so-called Saladin Tithe raised earlier for the Crusade); and a seizure of the wool crop of Cistercian and Gilbertine houses. The ransom, although never paid in full, caused Richard's government to become highly unpopular. Richard also faced some unwillingness on the part of his English subjects to serve in France. A plan to raise a force of 300 knights who would serve for a whole year met with opposition led by the bishops of Lincoln and Salisbury. Richard was, however, remarkably successful in mustering the resources, financial and human, of his kingdom in support of his wars. It can also be argued that his demands on England weakened the realm unduly and that Richard left his successor a very difficult legacy.

Richard's capture and ransom

JOHN (1199–1216)

Richard, mortally wounded at a siege in France in 1199, was succeeded by his brother John, one of the most detested of English kings. John's reign was characterized by failure. Yet while he must bear a heavy responsibility for his misfortunes, it is only fair to recognize that he inherited the resentment that had built up against his brother

and father. Also, while his reign ended in disaster, some of his financial and military measures anticipated positive developments in Edward I's reign.

Loss of French possessions. John had nothing like the military ability or reputation of his brother. He could win a battle in a fit of energy, only to lose his advantage in a spell of indolence. After repudiating his first wife, Isabella of Gloucester, John married the fiancée of Hugh IX the Brown of the Lusignan family, one of his vassals in Poitou. For this offense he was summoned to answer to Philip II, his feudal overlord for his holdings in France. When John refused to attend, his lands in France were declared forfeit. In the subsequent war he succeeded in capturing his nephew Arthur of Brittany, whom many in Anjou and elsewhere regarded as Richard I's rightful heir. Arthur died in mysterious and suspicious circumstances. But once the great castle of Château Gaillard, Richard I's pride and joy, had fallen in March 1204, the collapse of Normandy followed swiftly. By 1206 all that was left of the inheritance of the Norman kings was the Channel Islands. John, however, was determined to recover his losses.

Struggle with the papacy. Upon his return to England John became involved in a conflict with Pope Innocent III over the choice of an archbishop. At Hubert Walter's death in 1205 the monks at Canterbury had secretly elected their subprior and sent him to Rome to receive the pallium from the pope. The secret got out, however, and John forced the election of one of his confidants, John de Grey, bishop of Norwich, who then was also sent to Rome. Innocent III was not a man to miss such a good opportunity to demonstrate the plenitude of papal power. He quashed both elections and engineered the election of the learned and talented cardinal Stephen Langton. John, however, refused to receive Stephen and seized the revenues of Canterbury. Since John had already quarreled with his half brother the archbishop of York, who had fled abroad, England was without either archbishop. In 1208 Innocent imposed an interdict on England, forbidding the administration of the sacraments and certain church rites. In the following year he excommunicated John. The bishops of Winchester and Norwich remained the sole support of John's power in the church. John made the most of the opportunity to collect the revenues of the sees vacated by bishops who had gone into exile.

John's excommunication

In theory John's excommunication freed his vassals from their oaths of fealty to him, but there was no immediate rebellion. John was able to conduct highly successful expeditions to Scotland, Wales, and Ireland, and it was not until 1212 that a plot, involving Robert Fitzwalter and Eustace de Vesci, was first hatched against the king. John's brilliant solution to the problem of multiple threats was to effect a reconciliation with the papacy. He agreed to accept Stephen Langton as archbishop, to reinstate the exiled clergy, and to compensate the church for his exactions. In addition he surrendered his kingdom to the pope, receiving it back as a fief from the pope. He now had an able ally at no great cost in terms of concessions on his part.

Revolt of the barons and Magna Carta. Ever since the loss of Normandy John had been building up a coalition of rulers in Germany and the Low Countries to assist him against the French king. His chief ally was Otto IV, king of Germany and Holy Roman emperor. Plans for a campaign in Poitou proved very unpopular in England, especially with the northern barons. In 1214 John's allies were defeated at Bouvines, and the king's own campaign in Poitou disintegrated. John had to withdraw and return home to face his disgruntled barons.

John's efforts had been very costly, and measures such as the tax of a 13th in 1207 (which raised about £60,000) were highly unpopular. In addition John levied massive reliefs (inheritance duties) on some barons: Nicholas de Stuteville, for example, was charged 10,000 marks (about £6,666) to inherit his brother's lands in 1205. The fact alone that John, unlike his predecessors on the throne, spent most of his time in England made his rule more oppressive. Resistance sprang chiefly from the northern barons who had opposed service in Poitou, but by the spring of 1215 many others had joined them in protest against John's abuse or disregard of law and custom.

On June 15, 1215, the rebellious barons met John at Runnymede on the Thames. The king was presented with a document known as the Articles of the Barons, on the basis of which Magna Carta was drawn up. For a document hallowed in history during more than 750 years and frequently cited as a forerunner of the Declaration of Independence and the Declaration of the Rights of Man and of the Citizen, Magna Carta is a singularly undramatic document. It is thorny with problems of feudal law and custom that are largely untranslatable into modern idiom. Still, it was remarkable in many ways, not least because it was not written in a purely baronial interest but aimed to provide protection for all freemen. It was an attempt to provide guarantees against the sort of arbitrary disregard of feudal right that the three Angevin kings had made familiar. The level of reliefs, for example, was set at £100 for a barony. Some clauses derived from concessions already offered by the king in efforts to divide opposition. The celebrated clause 39, which promised judgment by peers or by the law of the land to all freemen, had its origins in a letter sent by Innocent III to the king. The barons, however, were not attempting to dismantle royal government; in fact, many of the legal reforms of Henry II's day were reinforced. Nor did they seek to legitimate rebellion but rather they tried to ensure that the king was beneath rather than above the law. In immediate terms Magna Carta was a failure, for it was no more than a stage in ineffective negotiations to prevent civil war. John was released by the pope from his obligations under it. The document was, however, reissued with some changes under John's son, with papal approval, and so it became, in its 1225 version, a part of the permanent law of the land. John himself died in October 1216, with the civil war still at an inconclusive stage.

The Articles of the Barons

Economy and society. From about 1180 the pace of economic change quickened, with a shift to what is known as "high farming." The direct management of estates began to replace a rentier system. There was a marked price and wage inflation. Daily wages for a knight rose from eight pence a day early in Henry II's day to two shillings under John. Landlords who relied upon fixed rents found times difficult, but most responded by taking manors into their own hands and by profiting from direct sales of demesne produce at market. A new class of professional estate managers, or stewards, began to appear. Towns continued to prosper, and many bought privileges of self-government from Richard I and John. The weaving industry was important, and England was noted as a producer of very high quality woolen cloth.

England, notably under Henry II, participated in the cosmopolitan movement that has come to be called the "12th-century Renaissance." Scholars frequented the court, and works on law and administration, especially the *Dialogue of the Exchequer* and the law book attributed to Ranulf de Glanville, show how modern ideas were being applied to the arts of government. In ecclesiastical architecture new methods of vaulting gave builders greater freedom, as may be seen, for example, in the construction of the choir at Canterbury, rebuilt after a fire in 1174 by William of Sens. In military architecture, the traditional rectangular plan was abandoned in keeps such as those at Orford and Conisborough. It was a self-confident, innovative, and assertive age.

The 13th century

The 13th century saw England develop a much clearer identity. The loss of continental possessions under King John focused the attention of the monarchy on England in a way that had not happened since 1066. Not only did the concept of the community of the realm develop—used both by the crown and its opponents—but the period was also notable in constitutional terms, seeing the beginning of Parliament.

The notion that the realm was a community and that it should be governed by representatives of that community perhaps found its first practical expression in the period following the issue of Magna Carta in which a council of regency ruled on behalf of a child king not yet able

The notion of the community of the realm

to govern in his own right. The phrase "community of the land" initially meant little more than the totality of the baronage. But the need to obtain a wider degree of consent to taxation, and perhaps also the impact of new ideas derived from Roman law, led to change. In addition the county communities exerted some pressure. Knights were being asked to play an increasingly important part in local government, and soon they made their voice heard at a national level. In the conflict that broke out between Henry III and the barons in the latter part of that king's reign, political terms acquired some sophistication, and under Edward I the concept of representation was further developed.

HENRY III (1216–72)

Minority. The years until his death in 1219 were dominated by William Marshal, 1st Earl of Pembroke. As regent in all but name he achieved success in the civil war and, assisted by the papal legate Guala, did much to restore royal government in its aftermath. After Marshal's death there was a struggle for political power between Hubert de Burgh, the justiciar, and Peter des Roches, bishop of Winchester. Despite factional disputes, by the time Henry III declared himself to be of age in 1227, the minority government had achieved much. To have retained control of royal castles was a notable achievement, while the seizure of Bedford Castle from Fawkes de Breauté, a former protégé of King John, was a spectacular triumph.

Early reign. Henry came under increasing foreign dominance. His marriage in 1236 to Eleanor of Provence was followed by an influx of her Savoyard relations, while the other significant group of foreigners was headed by the king's half brothers, the Lusignans (children of his mother, Isabella, by her second marriage). Attempts to recover the lost lands in France with expeditions in 1230 and 1242 were unsuccessful. Only in Wales did he achieve limited military success. In the 1250s plans, backed by the papacy, were made to place Henry's second son Edmund on the Sicilian throne; by 1258 these plans had involved the crown in an impossibly heavy financial commitment of 135,000 marks. A lenient policy toward the magnates did not yield much support for the king, and after 1237 it proved impossible to negotiate the grant of direct taxes with unwilling subjects.

Foreign influences

Henry, moreover, faced a series of political crises. A baronial revolt in 1233, led by Richard, son of William Marshal, ended in tragedy. Richard was killed in Ireland, to the king's great grief: there were allegations that the king had been tricked into agreeing to the earl's destruction. Further political crises in 1238 and 1244 did nothing to resolve tensions. In 1238 the king's brother, Richard, Earl of Cornwall, rebelled, and leading advisers such as William of Savoy left the royal council. In 1244 Henry III faced opposition in Parliament from both lay and ecclesiastical magnates. A draft proposal suggested a complex system for adding four men to the council, who were to be "conservators of liberties" as well as overseers of royal finance. The king was able, however, to exploit the differences between his opponents, and their campaign achieved little. Henry was naive; he was, on the one hand, overly trustful and, on the other, bitter against those who betrayed his trust. There was growing discontent at a local level with the conduct of royal government.

The county communities. The society of the period should not be seen solely in terms of the feudal hierarchy. There are indications that the community of the county, dominated by local knights and the stewards of the magnates, was of growing importance in this period. Although the crown could and did rely extensively on the knights in local government and administration, the knights were resentful of any intrusion of royal officers from outside and determined to defend local rights and privileges. Incidents such as that in Lincolnshire in 1226, when the county community protested against innovations in the holding of the county court and appealed to Magna Carta, show a new political awareness at a local level. The localities resented the increased burdens placed on them by Henry III's government, and tension between court and country was evident.

Simon de Montfort and the Barons' War. The main crisis of the reign came in 1258 and was brought on by a cluster of causes. The Savoyard and Lusignan court factions were divided; there were reverses in Wales; the costs of the Sicilian affair were mounting; and there was perceived to be a crisis in local government. In May 1258 the king was compelled to agree to a meeting of Parliament and to the appointment of a joint committee of dissident barons and his own supporters, 12 from each side, which was to recommend measures for the reform of the kingdom. In the Provisions of Oxford, drawn up in June, a scheme was set out for the creation of a council of 15 to supervise royal government. Parliament was to be held three times a year, at which the 15 would meet with 12 barons representing "the community" (*le commun* in the original French). The office of justiciar was to be revived, and he, with the chancellor and treasurer, was to account annually before the council. The new justiciar was to hear complaints throughout the country against royal officials. Sheriffs were to be local men, appointed for one year. The households of the king and queen were to be reformed. The drafting of further measures took time. In October 1259 a group calling itself the Community of Bachelors, which seems to have claimed to represent the lesser vassals and knights, petitioned for the fulfillment of the promises of the magnates and king to remedy its grievances. As a result the Provisions of Westminster were duly published, comprising detailed legal measures that in many cases were in the interests of the knightly class.

The Provisions of Oxford

The Provisions of Oxford led to two years in which the king was under tutelage; he was less even than the first among equals because he was not free to choose his own councillors. The Oxford settlement, however, began to break down in 1260. There were divisions among the king's opponents, notably between the Earl of Gloucester and the ambitious Simon de Montfort, Earl of Leicester, Henry's brother-in-law. The king's eldest son, Edward, at first backed the unpopular Lusignans, whose exile had been demanded, but then came to an agreement with Simon de Montfort before being reconciled to his father. In 1261, when a papal bull released Henry from his oath to support the Provisions of Oxford, he dismissed the baronial sheriffs, castellans, and other officials imposed on him. Simon de Montfort, by now the undisputed leader of the opposition, raised rebellion, but an agreement was reached to submit the dispute to the arbitration of Louis IX of France. The verdict of the Mise of Amiens in 1264, however, was so favourable to Henry III that Simon de Montfort could not accept it.

Civil war was inevitable. In May 1264 Simon won a resounding victory at Lewes, and a new form of government was set up. Representatives of the boroughs were summoned to Parliament for the first time early in 1265, along with knights of the shire. Simon's motive for summoning Parliament was undoubtedly political: he needed support from many elements of society. In May 1265 the young Edward, held hostage since 1264 to ensure fulfillment of the terms of the peace of Lewes, escaped and rallied the royalist forces, notably the Welsh marcher lords who played a decisive part throughout these conflicts. In August, Simon was defeated and slain at Evesham.

Later reign. Henry spent the remainder of his reign settling the problems created by the rebellion. He deprived Simon's supporters of their lands, but "the Disinherited" fought back from redoubts in forests or fens. The garrison of Kenilworth Castle carried on a notable resistance. Terms were set in 1266 for former rebels to buy back their lands, and with the issue of the Statute of Marlborough, which renewed some of the reform measures of the Provisions of Westminster, the process of reconstruction began. By 1270 the country was sufficiently settled for Edward to be able to set off on crusade, from which he did not return until two years after his father's death. By then the community of the realm was ready to begin working with, not against, the crown.

"The Disinherited"

EDWARD I (1272–1307)

Edward was in many ways the ideal medieval king. He went through a difficult apprenticeship, was a good fighter,

and was a man who enjoyed both war and statecraft. His crusading reputation gave him prestige, and his chivalric qualities were admired. Although he had a gift for leadership, he lacked sympathy for others and had an obstinacy that led to inflexibility.

Law and government. In the 13th century the development of law became a dominant concern, as is shown by the great treatise *On the Laws and Customs of England,* attributed to the royal judge Bracton but probably put together in the 1220s and '30s under one of his predecessors on the King's Bench. Soon after Edward's return to England in 1274, a major inquiry into government in the localities took place that yielded the so-called Hundred Rolls, a heterogeneous group of records, and brought home the need for changes in the law. In 1275 the First Statute of Westminster was issued. A succession of other statutes followed in later years, providing a kind of supplement to the common law. Some measures protected the king's rights; others remedied the grievances of his subjects. In the quo warranto proceedings set up under the Statute of Gloucester of 1278 the magnates were asked by what warrant they claimed rights of jurisdiction and other franchises. This created much argument, which was resolved in the Statute of Quo Warranto of 1290. By the Statute of Mortmain of 1279 it was provided that no more land was to be given to the church without royal license. The Statute of Quia Emptores of 1290 had the effect of preventing further subinfeudation of land. In the first and second statutes of Westminster, of 1275 and 1285, many deficiencies in the law were corrected, such as those concerning the relationship between lords and tenants and the way in which the system of distraint was operated. Merchants benefited from the Statute of Acton Burnell of 1283 and the Statute of Merchants of 1285, which facilitated debt collection. Problems of law and order were tackled in the Statute of Winchester of 1285.

Hundred Rolls

Finance. Edward began his reign with heavy debts incurred on crusade, and his various wars also were costly. In 1275 Edward gained a secure financial basis when he negotiated a grant of export duties on wool, woolfells, and hides that brought in an average of £10,000 a year. He borrowed extensively from Italian bankers on the security of these customs revenues. The system of levying taxes on an assessment of the value of movable goods was also of great value. Successive profitable taxes were granted, mostly in Parliament. It was partly in return for one such tax, in 1290, that Edward expelled the Jews from England. Their moneylending activities had made them unpopular, and royal exploitation had so impoverished the Jews that there was no longer an advantage for Edward in keeping them in England.

The growth of Parliament. Edward fostered the concept of the community of the realm and the practice of calling representative knights of the shire and burgesses from the towns to Parliament. Representatives were needed to give consent to taxation, as well as to enhance communication between the king and his subjects. The process of petitioning the king and his council in Parliament was greatly encouraged. Historians have argued much about the nature of Edward's Parliament, some seeing the dispensation of justice as the central element, others emphasizing the multifaceted character of an increasingly complex institution. Some see Edward as responding to the dictates of Roman law, while others interpret the development of Parliament in terms of the practical solution of financial and political problems. Historians used to refer to the 1295 assembly as the Model Parliament because it contained all the elements later associated with the word parliament, but in fact these can all be found earlier. The writs to the sheriffs asking them to call knights and burgesses did, however, reach a more or less final form in 1295. They were to be summoned "with full and sufficient authority on behalf of themselves and the community . . . to do whatever shall be ordained by common counsel." Representatives of the lower clergy were also summoned. This Parliament was fully representative of local communities and of the whole community of the realm, but many Parliaments were attended solely by the magnates with no representatives present.

The Model Parliament

Edward's wars. In the first half of his reign Edward was thoroughly successful in Wales. Llywelyn ap Gruffudd, prince of Gwynedd, had taken advantage of the Barons' War to try to expand his authority throughout Wales. He refused to do homage to Edward, and in 1277 the English king conducted a short and methodical campaign against him. Using a partly feudal, partly paid army, the core of which was provided by the royal household knights, and a fleet from the Cinque Ports, Edward won a quick victory and exacted from Llywelyn the Treaty of Conway. Llywelyn agreed to perform fealty and homage, to pay a large indemnity (from which he was soon excused), and to surrender certain districts of North Wales. There was considerable Welsh resentment after 1277 at the manner in which Edward imposed his jurisdiction in Wales.

David, Llywelyn's younger brother, was responsible for a renewal of war in 1282. He was soon joined by Llywelyn, who was killed in battle late in the year. David was captured and executed as a traitor in 1283. This second Welsh war proved much longer, more costly, and more difficult for the English than the first. In the succeeding peace North Wales was organized into counties, and law was revised along English lines. Major castles, notably Flint and Rhuddlan, had been built after the first Welsh war; now Conway, Caernarvon, and Harlech were started, designed by a Savoyard expert, Master James of St. George. Merchant settlements, colonized with English craftsmen and merchants, were founded. Archbishop Pecham reorganized the Welsh church and brought it more fully under the sway of Canterbury. A brief revolt in 1287 was soon quelled, but Edward faced a major rebellion in 1294–95, after which he founded the last of his Welsh castles, Beaumaris in Anglesey.

Edward devoted much attention to Gascony, the land he held in southwestern France. He went there prior to returning to England at the start of the reign and spent the period 1286–89 there. In 1294 he had to undertake a costly defense of his French lands, when war began with Philip IV, king of France. Open hostilities lasted until 1297. In this case the French were the aggressors. Following private naval warfare between Gascon and Norman sailors, Philip summoned Edward (who, as Duke of Aquitaine, was his vassal) to his court and, having deceived English negotiators, decreed Gascony confiscate. Edward built up a grand alliance against the French, but the war proved costly and inconclusive.

French war

Edward intervened in Scotland in 1291, when he claimed jurisdiction over a complex succession dispute. King Alexander III had been killed when his horse fell one stormy night in 1286. His heiress was his three-year-old granddaughter, Margaret, the Maid of Norway. Arrangements were made for her to marry Edward's son Edward, but these plans were thwarted by Margaret's death in 1290. There were 13 claimants to the Scottish throne, the two main candidates being John de Balliol and Robert de Bruce, both descendants of David, 8th Earl of Huntingdon, brother of William I the Lion. Balliol was the grandson of David's eldest daughter, and Bruce was the son of his second daughter. A court of 104 auditors, of whom 40 were chosen by Balliol and 40 by Bruce, was set up. Balliol was designated king and performed fealty and homage to Edward.

Edward did all he could to emphasize his own claims to feudal suzerainty over Scotland, and his efforts to put these into effect provoked Scottish resistance. In 1295 the Scots, having imposed a baronial council on Balliol, made a treaty with the French. War was inevitable, and in a swift and successful campaign Edward defeated Balliol in 1296, forcing him to abdicate. The victory, however, had been too easy. Revolt against the inept officials Edward had appointed to rule in Scotland came in 1297, headed by William Wallace and Andrew Moray. Victory for Edward at the battle of Falkirk in 1298, however, did not win the war. A lengthy series of costly campaigns appeared to have brought success by 1304, and in the next year Edward set up a scheme for governing Scotland, by now termed by the English a land, not a kingdom. But in 1306 Robert de Bruce, grandson of the earlier claimant to the throne, a man who had fought on both sides in the war,

seized the Scottish throne and reopened the conflict, which continued into the reign of Edward II, who succeeded his father in 1307.

It has been claimed that during his wars Edward I transformed the traditional feudal host into an efficient, paid army. In fact, feudal summonses continued throughout his reign, though only providing a proportion of the army. The paid forces of the royal household were a very important element, but it is clear that the magnates also provided substantial unpaid forces for campaigns of which they approved. The scale of infantry recruitment increased notably, enabling Edward to muster armies up to 30,000 strong. The king's military successes were primarily due to the skill of his government in mobilizing resources, in terms of men, money, and supplies, on an unprecedented scale.

Domestic difficulties. The wars in the 1290s against the Welsh, French, and Scots imposed an immense burden on England. The character of the king's rule changed as the preoccupation with war put an end to further reform of government and law. Edward's subjects resented the heavy taxation, large-scale recruitment, and seizures of food supplies and wool crops. Pope Boniface VIII forbade the clergy to pay taxes to the king. A political crisis ensued in 1297, which was only partly resolved by the reissue of Magna Carta and some additional concessions. Argument continued for much of the rest of the reign, while the king's debts mounted. The Riccardi, Edward's bankers in the first part of the reign, were effectively bankrupted in 1294, and their eventual successors, the Frescobaldi, were unable to give the king the same level of support as their predecessors.

SOCIAL, ECONOMIC, AND CULTURAL CHANGE

The population expanded rapidly in the 13th century, reaching a level of about five million. Great landlords prospered with the system of high farming, but the average size of small peasant holdings fell, with no compensating rise in productivity. There has been debate about the fate of the knightly class: some historians have argued that lesser landowners suffered a decline in wealth and numbers, while others have pointed to their increased political importance as evidence of their prosperity. Although there were probably both gainers and losers, the overall number of knights in England almost certainly fell to less than 2,000. Ties between magnates and their feudal tenants slackened as the relationship became increasingly a legal rather than a personal one. Lords began to adopt new methods of recruiting their retinues, using contracts demanding service either for life or for a short term, in exchange for fees, robes, and wages. Towns continued to grow, with many new ones being founded, but the weaving industry suffered a decline, in part because of competition from rural areas and in part as a result of restrictive guild practices. In trade, England became increasingly dependent on exports of raw wool.

Changes in feudal ties

The advent of the friars introduced a new element to the church. The universities of Oxford and Cambridge were developing rapidly, and in Robert Grosseteste and Roger Bacon, England produced two major, if somewhat eccentric, intellectual figures. Ecclesiastical architecture flourished, showing a strong French influence: Henry III's patronage of the new Westminster Abbey was particularly notable. Edward I's castles in North Wales rank high among the finest examples of medieval military architecture.

The 14th century

The 14th century, despite some gains, was a bleak age. At its beginning and close were kings whose reigns ended in failure. In between, however, came the 50-year reign of the popular and successful Edward III. During the century the importance of the Commons in Parliament continued to grow. But dominant factors of the age were war and plague. The increased scale, cost, and frequency of wars from the 1290s onward imposed heavy burdens on state and society. Conflicts between England and France continued intermittently throughout the century, those from 1337 onward being called the Hundred Years' War. The Black Death struck in 1348–49; it became endemic, recurring several times in the second half of the century, and brought with it profound economic and social change.

EDWARD II (1307–27)

Edward II's reign was an almost unmitigated disaster. He inherited some of his problems from his father, the most significant being a treasury deficit of some £200,000, and the Scottish war. He inherited none of his father's strengths. He was a good horseman but did not enjoy swordplay or tournaments, preferring swimming, ditch digging, thatching, and theatricals. Although surrounded by a ruling class strongly tied to his family by blood and service, Edward rejected the company of his peers, preferring that of Piers Gaveston, son of a Gascon knight, with whom he probably had a homosexual relationship. Edward's father had exiled Gaveston in an attempt to quash the friendship. Edward the son recalled him and conferred on him the highest honours he had to bestow: the earldom of Cornwall and marriage to his niece Margaret de Clare, sister of the Earl of Gloucester. Edward also recalled Archbishop Winchelsey and Bishop Bek of Durham, both of whom had gone into exile under Edward I. He dismissed and put on trial one of his father's most trusted servants, the treasurer, Walter Langton.

Historians used to emphasize the constitutional struggle that took place in this reign, seeing a conflict between a baronial ideal of government conducted with the advice of the magnates and based on the great offices of state, the Chancery and the Exchequer, on the one hand, and a royal policy of reliance upon the departments of the royal household, notably the wardrobe and chamber, on the other. More recent interpretations have shifted the emphasis to personal rivalries and ambitions.

Opposition to Edward began to build as early as January 1308. At the coronation in February a new clause was added to the king's oath that obligated him to promise that he would keep such laws "as the community of the realm shall have chosen." In April the barons came armed to Parliament and warned the king that "homage and the oath of allegiance are stronger and bind more by reason of the crown than by reason of the person of the king." The first phase of the reign culminated in the production of the Ordinances in 1311. They were in part directed against Gaveston—who was again to be exiled—and other royal favourites, but much of the document looked back to the grievances of Edward I's later years, echoing concessions made by the king in 1300. Hostility was expressed to the practice of prise (compulsory purchase of foodstuffs for royal armies). Baronial consent was required for foreign war (possibly in remembrance of Edward I's Flanders campaign of 1297). The privy seal was not to be used to interfere in justice. A long list of officials were to be chosen with the advice and consent of the barons in Parliament. All revenues were to be paid into the Exchequer. The king's bankers, the Frescobaldi, who had also served Edward I, were to be expelled from the realm. Royal grants of land made since the appointment of the Ordainers in 1310 were annulled. It is noteworthy that the first clear statement that consent should be given in Parliament is to be found in the Ordinances. No explicit role, however, was given to the Commons, the representative element in Parliament.

The Ordinances of 1311

The middle years of Edward's reign were dominated by the enigmatic figure of Thomas, 2nd Earl of Lancaster, the king's cousin and chief opponent, whose surly inactivity for long periods blocked effective political initiatives. His political program never amounted to much more than enforcement of the Ordinances. He supervised the capture and execution of Gaveston in 1312 and came to dominance after the disastrous defeat of a royal army at the hands of the Scottish pikemen and bowmen at Bannockburn in 1314. At the Lincoln Parliament of 1316 he was named chief councillor, but he soon withdrew from active government.

A conciliar regime was set up with the Treaty of Leake of 1318. This was once thought to have been the work of a "middle party," but the political alliances of this period cannot be categorized in such a manner. New

royal favourites emerged, and in 1321 the peace was broken when the Welsh marcher lords moved against two of them, a father and son, both called Hugh Despenser. When Parliament met, the two were exiled, but they soon returned. In this brief civil war, which ended in 1322, Edward was victorious. He had Lancaster executed for treason after his ignominious defeat at Boroughbridge in 1322. In death Lancaster attracted a popular sympathy he had rarely received in life, with many rumours of miracles at his tomb. Edward had many of Lancaster's followers executed in a horrific bloodbath. In the same year the Ordinances were repealed in Parliament at York, and in the Statute of York the intention of returning to the constitutional practices of the past was announced. But in specifying that the "consent of the prelates, earls, and barons, and of the community of the realm" was required for legislation, the Statute of York provided much scope for historical argument; some historians have made claims for a narrow baronial interpretation of what is meant by "community of the realm," while others have seen the terminology as giving the representative element in Parliament a new role. A tract written in this period, the *Modus tenendi parliamentum,* certainly placed a new emphasis on the representatives of shire, borough, and lower clergy. In terms of practical politics, however, the Statute of York permitted the fullest resumption of royal authority.

The Statute of York

The final period of the reign saw the Despensers restored to power. They carried out various administrative reforms, ably assisted by the treasurer, Walter Stapledon. For the first time in many years, a substantial treasury of about £60,000 was built up. At the same time, crude blackmail and blatant corruption characterized this regime. A brief war against the French was unsuccessful. The reign ended with the invasion of Edward's estranged queen, Isabella, assisted by Roger Mortimer, soon to be Earl of March. With the support above all of the Londoners, the government was overthrown, the Despensers executed, and the king imprisoned. Parliament was called in his name, and he was simultaneously deposed and persuaded to abdicate in favour of his son, Edward III. After two conspiracies to release him, he was almost certainly killed in Berkeley Castle.

EDWARD III (1327–77)

The Hundred Years' War, to 1360. Edward III achieved personal power when he overthrew his mother's and Mortimer's dominance in 1330 at the age of 17. Their regime had been just as corrupt as that of the Despensers but less constructive. The young king had been sadly disappointed by an unsuccessful campaign against the Scots in 1327; in 1333 the tide turned when he achieved victory at Halidon Hill. Edward gave his support to Edward Balliol as claimant to the Scottish throne, rather than to Robert I's son David II. But as long as the Scots had the support of the French king Philip VI, final success proved impossible, and this was one of the causes for the outbreak of the French war in 1337. Another was the long-standing friction over Gascony, chronic since 1294 and stemming ultimately from the Treaty of Paris of 1259. By establishing that the kings of England owed homage to the kings of France for Gascony the treaty had created an awkward relationship. The building of bastides (fortified towns) by each side contributed to friction, as did piracy by English and French sailors. The English resented any appeals to the French court by Gascons. English-French rivalry also extended into the Netherlands, which was dependent on English wool for industrial prosperity but some of whose states, including Flanders, were subject to French claims of suzerainty. Finally, there was the matter of the French throne itself. Edward, through his mother, was closer in blood to the last ruler of the Capetian dynasty than was the Valois Philip VI. The claim was of great propaganda value to Edward, for it meant that he did not appear as simply a rebellious vassal of the French king. His allies could fight for him without dishonour.

Edward's claim to the French throne

The initial phase of the war was inconclusive. Edward won a naval victory at Sluys in 1340, but he lacked the resources to follow it up. Although intervention in a succession dispute in Brittany saw the English register successes, stalemate came in 1343. The first great triumph came with the invasion of Normandy in 1346. As Edward was retreating northward, he defeated the French at Crécy and then settled to the siege of Calais, which fell in 1347. The French allies, the Scots, were also defeated in 1346 at Neville's Cross, where their king, David II, was taken prisoner. The focus of the war moved south in 1355, when the king's son, the Black Prince, was sent to Gascony. He launched a successful raid in 1355 and another in 1356, and at Poitiers he defeated and captured the French king John, for whom a heavy ransom was charged. As at Crécy, English archery proved decisive. A major campaign in 1359–60, planned as the decisive blow, proved unsatisfactory to the English. Rheims did not open its gates to Edward as he had hoped, and a storm caused severe damage to the army and its baggage in April 1360. Negotiations led to a truce at Brétigny, and in the subsequent negotiations Edward agreed to drop his claim to the French throne. In return, English possessions in France would be held in full sovereignty. The terms, particularly those involving the exchange of territory, were not carried out in full, but neither side wished to reopen the war immediately. War was costly, and Edward III's armies were no longer recruited by feudal means. Most were formed by contract, and all who fought received wages as well as a share of the profits of campaigning. These could be substantial if wealthy nobles were captured and ransomed.

Domestic achievements. The war, and the need to finance it, dominated domestic affairs under Edward III. The king faced a crisis in 1340–41 because he found himself disastrously indebted by 1339, even though he had received generous grants from Parliament since 1336. It was estimated that he owed £300,000. He had seized wool exports and had borrowed recklessly from Italian, English, and Flemish bankers and merchants. A grant in 1340 of a ninth of all produce failed to yield the expected financial return. In the autumn of 1340 Edward returned from abroad and charged John Stratford, archbishop of Canterbury, the man who had been in charge in his absence, with working against him. He also engaged in a widespread purge of royal ministers. Stratford whipped up opposition to the king, and in Parliament in 1341 statutes were passed that were reminiscent of the kind of restraints put on earlier and less popular kings. Officers of state and of the king's household were to be appointed and sworn in Parliament. Commissioners were to be sworn in Parliament to audit the royal accounts. Peers were to be entitled to trial before their peers in Parliament. Breaches of the Charters were to be reported in Parliament. Charges were brought against Stratford, only to be dropped. But in 1343 Edward III was able to repudiate the statutes. The crisis had little permanent effect, though it did demonstrate the king's dependence on Parliament, and within it on the Commons, for supply.

In the following years the country was well governed, with William Edington and John Thoresby serving the king loyally and well. Edward's compliance toward the requests of the Commons made it relatively easy for him to obtain the grants he needed. Discontent in 1346–47 was overcome by the good news from France. Much of the legislation passed at this time was in the popular interest. In 1352 the king agreed that no one should be bound to find soldiers for the war save by common consent in Parliament, and demands for purveyance were moderated. The Statute of Provisors of 1351 set up statutory procedures against the unpopular papal practice of making appointments to church benefices in England, and the Statute of Praemunire two years later forbade appeals to Rome in patronage disputes. The crown in practice had sufficient weapons available to it to deal with these matters, but Edward was ready to accept the views of his subjects, even though he did little about them later. Much attention was given to the organization of the wool trade because it was intimately bound up with the finance of war. In 1363 the Calais staple was set up, under which all English exports of raw wool were channeled through Calais. The currency was reformed very effectively with the introduction in the 1340s of a gold coinage alongside the traditional silver pennies.

Legislation

Law and order. The maintenance of law and order, a prime duty for a medieval king, had reached a point of crisis by the end of Edward I's reign when special commissions, known as commissions of trailbaston, were set up to try to deal with the problem. Matters became worse under Edward II, from whose reign there is much evidence of gang warfare, often involving men of knightly status. Maintaining law and order was also an urgent issue in Edward III's reign. In the early years there was conflict between the magnates, who wanted to be given full authority in the localities, and the county knights and gentry, who favoured locally appointed keepers of the peace. A possible solution, favoured by the chief justice, Geoffrey Scrope, was to extend the jurisdiction of the king's bench into the localities. There was a major crime wave in 1346 and 1347, intensified by the activities of soldiers returning from France. The justices reacted by greatly extending the use of accusations of treason, but the Commons protested against procedures they claimed did little to promote order and much to impoverish the people. In 1352 the crown gave way, producing in the Statute of Treason a narrow definition of great treason that made it impossible to threaten common criminals with the harsh penalties which followed conviction for treason. The concern of the Commons had been that in cases of treason goods and land forfeited by those found guilty went to the crown, not to the overlord. In 1361 the position of justice of the peace was established by statute, marking another success for the Commons.

The crises of Edward's later years. The war with France was reopened in 1369 and went badly. The king was in his dotage and, since the death of Queen Philippa in 1369, in the clutches of his unscrupulous mistress Alice Perrers. The heir to the throne, Edward the Black Prince, was ill and died in 1376. Lionel of Antwerp, Duke of Clarence, the next son, had died in 1368, and John of Gaunt, Duke of Lancaster, the third surviving son, was largely occupied with his claims to Castile, his inheritance through his second wife, Constance. Edmund of Langley, the fourth surviving son, was a nonentity, and the youngest, Thomas of Woodstock, was not yet of age. In 1371 Parliament demanded the dismissal of William of Wykeham, the chancellor, and the appointment of laymen to state offices. The new government, dominated by men such as William Latimer, the chamberlain, proved unpopular and ineffective. When the so-called Good Parliament met in 1376, grievances had accumulated and needed to be dealt with. As in previous crises, a committee consisting of four bishops, four earls, and four barons was set up to take responsibility for the reforms. Then, under the leadership of Peter de la Mare, who may be termed the first Speaker, the Commons impeached Latimer, Alice Perrers, and a number of ministers and officials, some of whom had profited personally from the administration of the royal finances. The Commons took the role of prosecutors before the Lords in what amounted to a new procedure.

The rise of John of Gaunt

John of Gaunt, an unpopular figure at this time, had, as a result of the king's illness, presided uncomfortably over the Good Parliament. He ensured that the achievement of Peter de la Mare and his colleagues was ephemeral, taking charge of the government at the end of the reign. De la Mare was jailed in Nottingham. William of Wykeham was attacked for alleged peculation as chancellor, and Alice Perrers was restored to court. The Parliament of 1377 reversed all important acts of the Good Parliament. There were rumours in London that Gaunt aimed at the throne. But the Black Prince's widow made peace between Gaunt and the Londoners, and Wykeham's temporalities were restored. The reign ended in truce, if not peace.

RICHARD II (1377–99)

Richard II's reign was fraught with crises—economic, social, political, and constitutional. He was 10 years old when his grandfather died, and the first problem the country faced was having to deal with his minority. A "continual council" was set up to "govern the king and his kingdom." Although John of Gaunt was still the dominant figure in the royal family, neither he nor his brothers were included.

The Peasants' Revolt (1381). Financing the increasingly expensive and unsuccessful war with France was a major preoccupation. At the end of Edward III's reign a new device, a poll tax of four pence a head, had been introduced. A similar but graduated tax followed in 1379, and in 1380 another set at one shilling a head was granted. It proved inequitable and impractical, and, when the government tried to speed up collection in the spring of 1381, a popular rebellion—the Peasants' Revolt—ensued. Although the poll tax was the spark that set it off, there were also deeper causes related to changes in the economy and to political developments. The government, in particular, engendered hostility to the legal system by its policies of expanding the powers of the justices of the peace at the expense of local and manorial courts. In addition, popular poor preachers spread subversive ideas with slogans such as: "When Adam delved and Eve span / Who was then the gentleman?" The Peasants' Revolt began in Essex and Kent. Widespread outbreaks occurred through the southeast of England, taking the form of assaults on tax collectors, attacks on landlords and their manor houses, destruction of documentary evidence of villein status, and attacks on lawyers. Attacks on religious houses, such as that at St. Albans, were particularly severe, perhaps because they had been among the most conservative of landlords in commuting labour services.

Imposition of the poll tax

The men of Essex and Kent moved on London to attack the king's councillors. Admitted to the city by sympathizers, they attacked John of Gaunt's palace of the Savoy as well as the Fleet prison. On June 14 the young king made them various promises at Mile End; on the same day they broke into the Tower and killed Sudbury, the chancellor, Hales, the treasurer, and other officials. On the next day Richard met the rebels again at Smithfield, and their main leader, Wat Tyler, presented their demands. But during the negotiations Tyler was attacked and slain by the mayor of London. The young king rode forward and reassured the rebels, asking them to follow him to Clerkenwell. This proved to be a turning point, and the rebels, their supplies exhausted, began to make their way home. Richard went back on the promises he had made, saying, "Villeins ye are and villeins ye shall remain." In October Parliament confirmed the king's revocation of charters but demanded amnesty save for a few special offenders.

Wat Tyler

The events of the Peasants' Revolt may have given Richard an exalted idea of his own powers and prerogative as a result of his success at Smithfield, but for the rebels the gains of the rising amounted to no more than the abolition of the poll taxes. Improvements in the social position of the peasantry did occur, but not so much as a consequence of the revolt as of changes in the economy that would have occurred anyhow.

John Wycliffe. Religious unrest was another subversive factor under Richard II. England had been virtually free from heresy until John Wycliffe, a priest and an Oxford scholar, began his career as a religious reformer with two treatises in 1375–76. He argued that the exercise of lordship depended on grace and that, therefore, a sinful man had no right to authority. Priests and even the pope himself, Wycliffe went on to argue, might not necessarily be in a state of grace and thus would lack authority. Such doctrines appealed to anticlerical sentiments and brought Wycliffe into direct conflict with the church hierarchy, although he received protection from John of Gaunt. The beginning of the Great Schism in 1378 gave Wycliffe fresh opportunities to attack the papacy, and in a treatise of 1379 on the Eucharist he openly denied the doctrine of transubstantiation. He was ordered before a church court at Lambeth in 1378. In 1380 his views were condemned by a commission of theologians at Oxford, and he was forced to leave the university. At Lutterworth he continued to write voluminously until his death in 1384. The movement he inspired was known as Lollardy. Two of his followers translated the Bible into English, and others went out to spread Wycliffe's doctrines, which soon became debased and popularized. The movement continued to expand despite the death of its founder and the government's attempts to destroy it.

Political struggles and Richard's deposition. Soon after

putting down the Peasants' Revolt, Richard began to build up a court party, partly in opposition to Gaunt. A crisis was precipitated in 1386 when the king asked Parliament for a grant to meet the French threat. Parliament responded by demanding the dismissal of the king's favourites, but Richard insisted that he would not dismiss so much as a scullion in his kitchen at the request of Parliament. In the end he was forced by the impeachment of the chancellor, Michael de la Pole, to agree to the appointment of a reforming commission. Richard withdrew from London and went on a "gyration" of the country. He called the judges before him at Shrewsbury and asked them to pronounce the actions of Parliament illegal. An engagement at Radcot Bridge, at which Richard's favourite, Robert de Vere, 9th Earl of Oxford, was defeated, settled the matter of ascendancy. In the Merciless Parliament of 1388 five lords accused the king's friends of treason under an expansive definition of the crime.

The Merciless Parliament

Richard was chastened, but he began to recover his authority as early as the autumn of 1388 at the Cambridge Parliament. Declaring himself to be of age in 1389, Richard announced that he was taking over the government. He pardoned the Lords Appellant and ruled with some moderation until 1394, when his queen, Anne of Bohemia, died. After putting down a rebellion in Ireland, he was, for a time, almost popular. He began to implement his personal policy once more and rebuilt a royal party with the help of a group of young nobles. He made a 28-year truce with France and married the French king's seven-year-old daughter. He built up a household of faithful servants, including the notorious Sir John Bushy, Sir William Bagot, and Sir Henry Green. He enlisted household troops and built a wide network of "king's knights" in the counties, distributing to them his personal badge, the White Hart.

The first sign of renewed crisis emerged in January 1397, when complaints were put forward in Parliament and their author, Thomas Haxey, was adjudged a traitor. Richard's rule, based on fear rather than consent, became increasingly tyrannical. Three of the Lords Appellant of 1388 were arrested in July and tried in Parliament. The Earl of Arundel was executed and Warwick exiled. Gloucester, whose death was reported to Parliament, had probably been murdered. The acts of the 1388 Parliament were repealed. Richard was granted the customs revenues for life, and the powers of Parliament were delegated to a committee after the assembly was dissolved. Richard also built up a power base in Cheshire.

Events leading to Richard's downfall followed quickly. The Duke of Norfolk and Henry Bolingbroke, John of Gaunt's son, accused each other of treason and were banished, the former for life, the latter for 10 years. When Gaunt himself died early in 1399, Richard confiscated his estates instead of allowing his son to claim them. Richard, seemingly secure, went off to Ireland. Henry, however, landed at Ravenspur in Yorkshire to claim, as he said, his father's estates and the hereditary stewardship. The Percys, the chief lords in the north, welcomed him. Popular support was widespread, and when Richard returned from Ireland his cause was lost.

Richard's deposition

The precise course of events is hard to reconstruct, in view of subsequent alterations to the records. A Parliament was called in Richard's name, but before it was fully assembled at the end of September, its members were presented with Richard's alleged abdication and Henry's claim to the throne as legitimate descendant of Henry III as well as by right of conquest. Thirty-three articles of deposition were set forth against Richard, and his abdication and deposition were duly accepted. Richard died at Pontefract Castle, either of self-starvation or by smothering. Thus ended the last attempt of a medieval king to exercise arbitrary power. Whether or not Richard had been motivated by new theories about the nature of monarchy, as some have claimed, he had failed in the practical measures necessary to sustain his power. He had tried to rule through fear and mistrust in his final years, but he had neither gained sufficient support among the magnates by means of patronage nor created a popular basis of support in the shires.

ECONOMIC CRISIS AND CULTURAL CHANGE

Although the outbreak of the Black Death in 1348 dominated the economy of the 14th century, a number of adversities had already occurred in the preceding decades. Severe rains in 1315 and 1316 caused famine, which led to the spread of disease. Animal epidemics in succeeding years added to the problems, as did an increasing shortage of currency in the 1330s. Economic expansion, which had been characteristic of the 13th century, had slowed to a halt. The Black Death, possibly a combination of bubonic and pneumonic plagues, carried off from one-third to one-half of the population. In some respects it took time for its effects to become detrimental to the economy, but with subsequent outbreaks, as in 1361 and 1369, the population declined further, causing a severe labour shortage. By the 1370s wages had risen dramatically and prices of foodstuffs fallen. Hired labourers, being fewer, asked for higher wages and better food, and peasant tenants, also fewer, asked for better conditions of tenure when they took up land. Some landlords responded by trying to reassert labour services where they had been commuted. The Ordinance (1349) and Statute (1351) of Labourers tried to set maximum wages at the levels of the pre-Black Death years, but strict enforcement proved impossible. The Peasants' Revolt of 1381 was one result of the social tension caused by the adjustments needed after the epidemic. Great landlords saw their revenues fall as a result of the Black Death, although probably by only about 10 percent, whereas for the lower orders of society real wages rose sharply by the last quarter of the 14th century because of low grain prices and high wages.

The Black Death

Edward III ruined the major Italian banking companies in England by failing to repay loans early in the Hundred Years' War. This provided openings for English merchants, who were given monopolies of wool exports by the crown in return for their support. The most notable was William de la Pole of Hull, whose family rose to noble status. Heavy taxation of wool exports was one reason for the growth of the cloth industry and cloth exports in the 14th century. The wine trade from Gascony was also important. In contrast to the 13th century, no new towns were founded, but London in particular continued to prosper despite the ravages of plague.

In cultural terms, a striking change in the 14th century was the increasing use of English. Although an attempt to make the use of English mandatory in the law courts failed because lawyers claimed that they could not plead accurately in the language, the vernacular began to creep into public documents and records. Henry of Lancaster even used English when he claimed the throne in 1399. Chaucer wrote in both French and English, but his important poetry is in the latter. The early 14th century was an impressive age for manuscript illumination in England, with the so-called East Anglian school, of which the celebrated Luttrell Psalter represents a late example. In ecclesiastical architecture the development of the Perpendicular style, largely in the second half of the 14th century, was particularly notable.

Lancaster and York

Recent scholarship has done much to transform the view that the 15th century was a period dominated by a factious nobility, when constructive achievements were few. In particular, the character of the nobility has been reconceived, and the century has emerged in a more positive light. It appears that even in politics and administration much was done that anticipated the achievements of the Tudors, while in the economy the foundations for future growth and prosperity were laid.

HENRY IV (1399–1413)

Henry of Lancaster gave promise of being able to develop a better rapport with his people than his predecessor, Richard II. He was a warrior of great renown who had traveled to Jerusalem and had fought in Prussia against infidels. He also had a reputation for affability and for statesmanlike self-control, and he had won his crown with the support of "the estates of the realm." It did not matter

much whether that meant Parliament or something more vague and symbolic. Henry, however, intended to rule as a true king, with the prerogatives of the crown unimpaired, whereas his Parliaments, from the first, expected him to govern with the advice and consent of his council, and to listen to Parliament regarding requests for money. Thus although Archbishop Arundel stressed in 1399 that Henry wished to be properly advised and that he intended to be governed by common advice and counsel, some argument and conflict was inevitable.

The rebellions. Henry's immediate task after his accession was to put down a rebellion threatening to restore Richard. The earls of Rutland, Kent, and Huntingdon, supported by the bishop of Carlisle, conspired against the king. The rising was unexpected, but Henry won support in London and defeated the rebels near Cirencester. More

Glyn Dwr's rebellion

significant was the revolt of Owain Glyn Dwr that broke out in 1399 and became serious in 1402. Glyn Dwr sought a French alliance and captured Edmund Mortimer, uncle of the Earl of March, Richard II's legitimate heir. Mortimer was persuaded to join the rebellion, which now aimed to make March king. In 1403 the Welsh rebels joined the Percys of Northumberland in a powerful coalition. The younger Percy, "Hotspur," was killed at Shrewsbury in 1403. The elder was pardoned, only to rebel once more in 1405, again in conjunction with Glyn Dwr. Henry broke the alliance with a victory at Shipton Moor. Percy was finally killed in 1408, but Glyn Dwr, driven into the mountains of North Wales, was never captured.

Henry and Parliament. Henry's relations with his Parliaments were uneasy. The main problem, of course, was money. Henry, as Duke of Lancaster, was a wealthy man, but as king he had forfeited some of his income by repudiating Richard II's tactics, though he also avoided Richard's extravagance. His needs were still great, threatened as he was by rebellion in England and war in France. A central issue was Parliament's demand, as in 1404, that the king take back all royal land that had been granted and leased out since 1366. This was so that he might "live of his own." The king could hardly adopt a measure that would cause much upheaval. Arguments in 1406 were so protracted that the Parliament met for 159 days, becoming the longest Parliament of the medieval period. On several occasions the Commons insisted on taxes being spent in the way that they wished, primarily on the defense of the realm.

The later Parliaments of Henry's reign brought no new problems, but the king became less active in government as he was more and more incapacitated by illness. From 1408 to 1411 the government was dominated first by Archbishop Arundel and then by the king's son Henry, who, with the support of the Beaufort brothers, sons of John of Gaunt by Katherine Swynford, attempted to win control over the council. There was much argument over the best political strategy to adopt in France, where civil war was raging; young Henry wanted to resume the war in France, but the king favoured peace. In 1411 the king recovered his authority, and the Prince of Wales was dismissed from the council. Uneasy relations between the prince and his father lasted until Henry IV's death in 1413.

HENRY V (1413–22)

Henry V's brief reign is important mainly for the glorious victories in France, which visited on his infant son the enormous and not-so-glorious burden of governing both France and England. Two rebellions undermined the security of the realm in the first two years of the reign.

Sir John Oldcastle

The first was organized by Sir John Oldcastle, a Lollard and former confidant of the king. Though Oldcastle was not arrested until 1417, little came of his rising. Another plot gathered around Richard, 5th Earl of Cambridge, a younger brother of the Duke of York. The aim was to place the Earl of March on the throne, but March himself gave the plot away, and the leading conspirators were tried and executed on the eve of the king's departure for France.

The French war. Henry invaded France in 1415 with a small army of some 9,000 men. The siege of Harfleur was followed by a march toward Calais. At Agincourt the English were forced to fight because their route onward was blocked; they won an astonishing victory. Between 1417 and 1419 Henry followed up this success with the conquest of Normandy and the grant of Norman lands to English nobles and lesser men. This was a new strategy for the English to adopt, replacing the plundering raids of the past. In 1420 in the Treaty of Troyes it was agreed that Henry would marry Catherine, Charles VI's daughter. He was to be heir to the French throne, and that throne was to descend to his heirs in perpetuity. But Charles VI's son, the Dauphin, was not a party to the treaty, and so the war continued. Henry, still wanting money but reluctant to ask for subsidies at a time when he needed all the support he could get for the treaty, obtained forced loans. There were increasing indications of unease in England. In 1422 Henry contracted dysentery and died at the siege of Meaux in August, leaving as his heir a son less than a year old.

Domestic affairs. England was competently governed under Henry V. Problems of law and order were dealt with by reviving the use of the King's Bench as a traveling court; central and local administration operated smoothly. Henry proved adept at persuading men to serve him energetically for limited rewards. Parliament, well-satisfied with the course of events in France, gave the king all the support he needed. War finance was efficiently managed, and although Henry died in debt, the level was a manageable one. His was a most successful reign.

HENRY VI (1422–61 AND 1470–71)

Henry VI was a pious and generous man, but he lacked the attributes needed for effective kingship. Above all he lacked political sense and was no judge of men. Until 1437 he was a child, under the regency of a council of nobles dominated by his uncles and his Beaufort kin. When he was declared of age, the Beauforts were the real rulers of England. In 1445, through the initiative of the Earl (later Duke) of Suffolk, he married Margaret of Anjou, who with Suffolk dominated the king. Finally, in the period from 1450 to 1461 he suffered two bouts of mental illness. During these crises Richard, 3rd Duke of York, ruled the kingdom as protector.

Domestic rivalries and the loss of France. In the first period of the reign John, Duke of Bedford, proved to be as able a commander in the French war as had his brother Henry V. But in 1429 Joan of Arc stepped forth and

Joan of Arc

rallied French resistance. Bedford died in 1435, and the Congress of Arras, an effort at a general peace settlement, failed. When Philip of Burgundy deserted the English alliance and came to terms with Charles VII, the conflict became a war of attrition. By 1453 the English had lost all their overseas possessions save Calais.

Despite the factional nature of politics, there was no breakdown at home. The country was ruled by a magnate council with the increasingly reluctant financial support of Parliament. Humphrey, Duke of Gloucester, and Henry Beaufort, bishop of Winchester (cardinal from 1426), were the dominant figures. The main problem was financing the war. The bishop had great wealth, which he increased by lending to the crown, receiving repayment out of the customs. Divisions in the council became more acute after 1435, with Gloucester advocating an aggressive war policy. He was, however, discredited when his wife was accused of witchcraft in 1441.

In 1447 both Cardinal Beaufort and Gloucester died, the latter in suspicious circumstances. The Duke of Suffolk was in the ascendant; he had negotiated a peace with France in 1444 and arranged the king's marriage to Margaret of Anjou in 1445. When war was renewed in 1446, the English position in Normandy collapsed. Becoming the scapegoat for the English failure, Suffolk was impeached in the Parliament of March 1450. As he was fleeing into exile, he was slain by English sailors from a ship called the *Nicholas of the Tower.* Edmund Beaufort, 2nd Duke of Somerset, succeeded him as leader of the court party.

Cade's rebellion. Less than three months later Jack Cade, a man of obscure origins, led a popular rebellion in southeastern England. In contrast to the rising of 1381, this was not a peasant movement; Cade's followers included many gentry, whose complaints were mainly about lack of government rather than economic repression. Thus

the remedies they proposed were political, such as the resumption of royal estates that had been granted out, the removal of corrupt councillors, and improved methods of collecting taxes. The rebels demanded that the king accept the counsel of Henry's rival, the Duke of York. They executed Lord Saye and Sele, the treasurer, and the sheriff of Kent, but the rising was soon put down.

The beginning of the Wars of the Roses. The so-called Wars of the Roses was the struggle between the Yorkist and Lancastrian descendants of Edward III for control of the throne and of local government. The origins of the conflict have been the subject of much debate. It can be seen as brought about as a result of Henry VI's inadequacy and the opposition of his dynastic rival Richard, Duke of York, but local feuds between magnates added a further dimension. Because of the crown's failure to control these disputes, they acquired national significance. Attempts have been made to link these civil conflicts to what is known as "bastard feudalism," the system that allowed magnates to retain men in their service by granting them fees and livery and made possible the recruiting of private armies. Yet this system can be seen as promoting stability in periods of strong rule as well as undermining weak rule such as that of Henry VI. Many nobles sought good government, rather than being factious, and were only forced into war by the king's incompetence. The outbreak of civil war in England was indirectly linked to the failure in France, for Henry VI's government had suffered a disastrous loss of prestige and, with it, authority.

The Duke of York's claim to the throne

The Duke of York had a claim to the throne in two lines of descent. One was through his mother, great-granddaughter of Lionel of Antwerp, Duke of Clarence, second surviving son of Edward III, and the other was through his father, son of Edmund of Langley, 1st Duke of York, fourth surviving son of Edward III. According to feudal principles he had a better hereditary right than anyone of the Lancastrian line. He had been sent as royal lieutenant to Ireland in 1446, but he returned from there with 4,000 men in 1450 to reassert his right to participate in the king's council and to counter Somerset's machinations. In 1454 York was made protector of the king, who had become insane in 1453, even though the queen and court party had tried to disguise the king's illness. Early in 1455 Henry recovered his wits. During his spell of insanity his queen had a son, Edward, which changed the balance of politics. York was no longer the heir apparent, and the country was faced with the prospect, should the king die, of another lengthy minority.

In 1455 York gathered forces in the north, alleging that he could not safely attend a council called to meet at Leicester without the support of his troops. He met the king at St. Albans. Negotiations were unsuccessful, and in the ensuing battle York's forces, larger than the king's, won a decisive victory. Somerset was slain and the king captured. A Yorkist regime was set up, with York as constable and the Earl of Warwick, emerging as the strong support of the Yorkist cause, as captain of Calais. The king fell ill again in the autumn of 1455, and York was again protector for a brief period; the king, however, recovered early in 1456.

Hostilities were renewed in 1459. The Yorkists fled without fighting before a royal force at Ludford Bridge, but the Lancastrians failed to make the most of the opportunity. Demands for money, purveyances, and commissions of array increased the burdens but not the benefits of Lancastrian rule. The earls of Warwick and Salisbury, with York's son Edward, used Calais as a base from which to invade England, landing at Sandwich in 1460. A brief battle at Northampton in July went overwhelmingly for the Yorkists, and the king was captured. At Parliament the Duke of York claimed the throne as heir to Richard II. The Commons and judges refused to consider a matter so high, leaving it to the Lords' decision. During the fortnight of debate the Lancastrians regrouped, and when Richard met them at Wakefield, he was defeated and killed. Warwick, somewhat later, was defeated at St. Albans.

The death of York

The Yorkist cause would have been lost if it had not been for Richard's son, Edward, Earl of March, who defeated the Lancastrians first at Mortimer's Cross and then at Towton Moor early in 1461. He was crowned king on June 28, but dated his reign from March 4, the day the London citizens and soldiers recognized his right as king..

EDWARD IV (1461–70 AND 1471–83)

During the early years of his reign, from 1461 to 1470, Edward was chiefly concerned with putting down opposition to his rule. Lancastrian resistance in the northeast and in Wales caused problems. France and Burgundy were also of concern because Margaret of Anjou's chief hope of recovering Lancastrian fortunes lay in French support; but Louis XI was miserly in his aid. Edward's main internal problem lay in his relations with Warwick, who had been his chief supporter in 1461. Richard Neville, 1st (or 16th) Earl of Warwick, called "the Kingmaker," was cousin to the king and related to much of the English nobility. Edward, however, refused to be dominated by him, particularly with respect to his marriage. When the crucial moment came in Warwick's negotiations for the king to marry the French king's sister-in-law, Edward disclosed his secret marriage in 1464 to a commoner, Elizabeth Woodville. The marriage of the king's sister to Charles the Bold of Burgundy was a success for the Woodvilles, for Warwick was not involved in the negotiations. Warwick allied himself to Edward's younger brother George, Duke of Clarence, and ultimately, through the machinations of Louis XI, joined forces with Margaret of Anjou, deposed Edward in 1470, and brought back Henry VI. The old king, dressed in worn and unregal clothing, was from time to time exhibited to the London citizens, while Warwick conducted the government. Edward IV went into brief exile in the Netherlands. But with the help of his brother-in-law, Charles the Bold, he recovered his throne in the spring of 1471 after a rapid campaign with successes at Barnet and Tewkesbury. Henry VI was put to death in the Tower, and his son was killed in battle.

Return of Edward IV

The second half of Edward's reign, 1471–83, was a period of relative order, peace, and security. The one event reminiscent of the politics of the early reign was the trial of the Duke of Clarence, who was attainted in Parliament in 1478 and put to death, reputedly by drowning in a butt of Malmsey wine. But Edward was popular. Because his personal resources from the duchy of York were considerable and because he agreed early in his reign to acts of resumption whereby former royal estates were taken back into royal hands, Edward had a large personal income and was less in need of parliamentary grants than his predecessors had been. Thus he levied few subsidies and called Parliament only six times. Among the few subsidies Edward did levy were benevolences, supposedly voluntary gifts, from his subjects primarily to defray the expenses of war. In 1475 Edward took an army to France but accepted a pension from the French king for not fighting, thereby increasing his financial independence still further. Councils were set up to govern in the Marches of Wales and in the north, where Edward's brother Richard presided efficiently. Edward's rule was characterized by the use of his household, its servants, and its departments, such as the chamber. He was a pragmatic ruler, whose greatest achievement was to restore the prestige of the monarchy. Where he failed was to make proper provision for the succession after his death.

Edward died in 1483, at age 40, worn out, it was said, by sexual excesses and by debauchery. He left two sons, Edward and Richard, to the protection of his brother Richard, Duke of Gloucester. After skirmishes with the queen's party Richard placed both of the boys in the Tower of London, then a royal residence as well as a prison. He proceeded to eliminate those who opposed his function as protector and defender of the realm and guardian to the young Edward V. Even Lord Hastings, who had sent word to Richard of Edward IV's death and who had warned him against the queen's party, was accused of treachery and was executed. On the day after the date originally set for Edward V's coronation the Lords and Commons summoned to Parliament unanimously adopted a petition requesting Richard to take over the throne. He accepted and was duly crowned king on July 6, taking the oath in English.

Accession of Richard III

RICHARD III (1483–85)

Richard was readily accepted no doubt because of his reputed ability and because people feared the insecurity of a long minority. The tide began to turn against him in October 1483, when it began to be rumoured that he had murdered or connived at the murder of his nephews. Whether this was true or not matters less than the fact that it was thought to be true and that it obscured the king's able government during his brief reign. Legislation against benevolences and protection for English merchants and craftsmen did little to counteract his reputation as a treacherous friend and a wicked uncle. Rebellion failed in 1483. But in the summer of 1485, when Henry Tudor, sole male claimant to Lancastrian ancestry and the throne, landed at Milford Haven, Richard's supporters widely deserted him, and he was defeated and killed at the Battle of Bosworth Field.

ENGLAND IN THE 15TH CENTURY

Central to all social change in the 15th century was change in the economy. Although plague remained endemic in England, there was little change in the level of population. Villein labour service largely disappeared, to be replaced by copyhold tenure (tenure by copy of the record of the manorial court). The period has been considered a golden age for the English labourer, but individual prosperity varied widely. There was a well-developed land market among peasants, some of whom managed to rise above their neighbours and began to constitute a class called yeomen. Large landlords entirely abandoned direct management of their estates in favour of a leasehold system. In many cases they faced growing arrears of rent and found it difficult to maintain their income levels. Because many landholders solved the problem of labour shortage by converting their holdings to sheep pasture, much land enclosure took place. As a result a great many villages were abandoned by their inhabitants.

Economic changes

Though England remained a predominantly agrarian society, significant development and change occurred in the towns. London continued to grow, dominating the southeast. Elsewhere the development of the woolen industry brought major changes. Halifax and Leeds grew at the expense of York, and the West Riding at the expense of the eastern part of Yorkshire. Suffolk and the Cotswold region became important in the national economy. As the cloth trade grew in importance, so did the association of the Merchant Adventurers. The merchants of the Staple, who had a monopoly on the export of raw wool, did less well. Italian merchants prospered in 15th-century England, and important privileges were accorded to the German Hanseatic merchants by Edward IV.

Culturally the 15th century was a period of sterility. Monastic chronicles came to an end, and the writing of history declined. Thomas Walsingham (d. *c.* 1422) was the last of a distinguished line of St. Albans chroniclers. Although there were some chronicles written by citizens of London as well as two lives of Henry V, distinguished works of history did not come until later. Neither were there any superior works of philosophy or theology. Reginald Pecock, an arid Scholastic philosopher, wrote an English treatise against the Lollards and various other works emphasizing the rational element in the Christian faith; he was judged guilty of heresy for his pains. No noteworthy poets succeeded Chaucer, though a considerable quantity of English poetry was written in this period. John Lydgate produced much verse in the Lancastrian interest. The printer William Caxton set up his press in 1476 to publish English works for the growing reading public. The first great collections of family correspondence, those of the Pastons, Stonors, and Celys, survive from this period.

The 15th century, however, was an important age in the foundation of schools and colleges. Some schools were set up as adjuncts to chantries, some by guilds, and some by collegiate churches. Henry VI founded Eton College in 1440 and King's College, Cambridge, in 1441. Other colleges at Oxford and Cambridge were also founded in this period. The Inns of Court expanded their membership and systematized their teaching of law. Many gentlemen's sons became members of the Inns, though not necessarily lawyers: they needed the rudiments of law to be able to defend and extend their estates. The influence of the Italian Renaissance in learning and culture was very limited before 1485, although there were some notable patrons, such as Humphrey, Duke of Gloucester, who collected books and supported scholars interested in the new learning.

Eton College

Only in architecture did England show great originality. Large churches were built in English Perpendicular style, especially in regions made rich by the woolen industry. The tomb of Richard Beauchamp at Warwick and King's College Chapel in Cambridge show the quality of English architecture and sculpture in the period. (M.Has./M.C.P.)

England under the Tudors

HENRY VII (1485–1509)

When Henry Tudor, earl of Richmond, seized the throne on Aug. 22, 1485, leaving the Yorkist Richard III dead upon the field of battle, few Englishmen would have predicted that 118 years of Tudor rule had begun. Six sovereigns had come and gone, and at least 15 major battles had been fought between rival contenders to the throne since that moment in 1399 when the divinity that "doth hedge a king" was violated and Richard II was forced to abdicate. Simple arithmetic forecast that Henry VII would last no more than a decade and that Bosworth Field was nothing more than another of the erratic swings of the military pendulum in the struggle between the houses of York and Lancaster. What gave Henry Tudor victory in 1485 was not so much personal charisma as the fact that key noblemen deserted Richard III at the moment of his greatest need, that Thomas Stanley, 2nd Baron Stanley (later 1st Earl of Derby), and his brother, Sir William, stood aside during most of the battle in order to be on the winning team, and that Louis XI of France supplied the Lancastrian forces with 1,000 mercenary troops.

The desperateness of the new monarch's gamble was equalled only by the doubtfulness of his claim. Henry VII's Lancastrian blood was tainted by bastardy twice over. He was descended on his mother's side from the Beaufort family, the offspring of John of Gaunt and his mistress Katherine Swynford, and, though their children had been legitimized by act of Parliament, they had been specifically barred from the succession. His father's genealogy was equally suspect: Edmund Tudor, Earl of Richmond, was born to Catherine of Valois, widowed queen of Henry V, by her clerk of the wardrobe, Owen Tudor; and the precise marital status of their relationship has never been established. Had quality of Plantagenet blood, not military conquest, been the essential condition of monarchy, Edward, Earl of Warwick, the 10-year-old nephew of Edward IV, would have sat upon the throne. Might, not soiled right, had won out on the high ground at Bosworth Field, and Henry VII claimed his title by conquest. The new king, however, wisely sought to fortify his doubtful genealogical pretension first by parliamentary acclamation and then by royal marriage. The Parliament of November 1485 did not confer regal power on the first Tudor monarch—victory in war had already done that—but it did acknowledge Henry as "our new sovereign lord." Then, on Jan. 18, 1486, Henry VII married Elizabeth of York, the eldest daughter of Edward IV, thereby uniting "the white rose and the red" and launching England upon a century of "smooth-fac'd peace with smiling plenty."

The weakness of Henry's claim to the throne

"God's fair ordinance," which Shakespeare and later generations so clearly observed in the events of 1485–86, was not limited to military victory, parliamentary sanction, and a fruitful marriage; the hidden hand of economic, social, and intellectual change was also on Henry's side. The day was coming when the successful prince would be more praised than the heroic monarch and the solvent sovereign more admired than the pious one. Henry Tudor was probably no better or worse than the first Lancastrian, Henry IV; they both worked diligently at their royal craft and had to fight hard to keep their crowns; but the seventh Henry achieved what the fourth had not—a secure and permanent dynasty—because England in 1485 was moving into a period of unprecedented economic growth and social change.

Economy and society. Long before 1485 the kingdom had begun to recover from the demographic catastrophe of the Black Death and the agricultural depression of the late 14th century; but as the 15th century came to a close, the rate of revival increased. The population in 1400 may have dropped as low as 2.5 million; by 1500 it was back up to at least three million, and a century later it was more than four million. More people meant more mouths to feed, more backs to cover, and more vanity to satisfy. In response, yeoman farmers, gentleman sheep growers, urban cloth manufacturers, and merchant adventurers produced a social and economic revolution. With extraordinary speed the export of raw wool gave way to the export of woolen cloth manufactured at home, and the wool clothier or entrepreneur was soon buying fleece from sheep raisers, transporting the wool to cottagers for spinning and weaving, paying the farmer's wife and children by the piece, and collecting the finished article for shipment to Bristol, London, and eventually Europe. By the time Henry VII seized the throne, the Merchant Adventurers, an association of London cloth exporters, were controlling the London–Antwerp market, replacing in economic importance the Staplers who dealt in raw wool. By 1496 they were a chartered organization with a legal monopoly of the woolen cloth trade, and largely as a consequence of their political and international importance, Henry successfully negotiated the Intercursus Magnus, a highly favourable commercial treaty between England and the Low Countries.

Revolution in the wool industry

As landlords increased the size of their flocks to the point that ruminants outnumbered human beings 3 to 1, and as clothiers grew rich on the wool trade, inflation injected new life into the economy. England was caught up in a vast European spiral of rising prices, declining real wages, and cheap money. Between 1500 and 1540, prices in England doubled, and they doubled again in the next generation. In 1450 the cost of wheat was what it had been in 1300; by 1550 it had tripled. Contemporaries blamed inflation on human greed and only slowly began to perceive that rising prices were the result of inflationary pressures brought on by the increase in population, international war, and the flood of gold and silver arriving from the New World.

Inflation and the wool trade together created an economic and social upheaval. Land plenty, labour shortage, low rents, and high wages, which had prevailed throughout the early 15th century as a consequence of economic depression and reduced population, were replaced by land shortage, labour surplus, high rents, and declining wages. The landlord, who a century before could find neither tenants nor labourers for his land and had left his fields fallow, could now convert his meadows into sheep runs. His rents and profits soared; his need for labour declined, for one shepherd and his dog could do the work of half a dozen men who had previously tilled the same field. Slowly the medieval system of land tenure and communal farming broke down. The common land of the manor was divided up and fenced in, and the peasant farmer who held his tenure either by copy (a document recorded in the manor court) or by unwritten custom was evicted. The total extent of enclosure and eviction is difficult to assess, but between 1455 and 1607 in 34 counties 516,573 acres (208,954 hectares), or 2.76 percent of the total, were enclosed, and some 50,000 persons were forced off the land. Statistics, however, are deceptive both as to emotional impact and extent of change. The most disturbing aspect of the land revolution was not the emergence of a vagrant and unemployable labour force for whom society felt no social responsibility but an unprecedented increase in what men feared most—change. Farming techniques were transformed, the gap between rich and poor increased, the timeless quality of village life was upset, and on all levels of society old families were being replaced by new.

Enclosures

The beneficiaries of change, as always, were the most grasping, the most ruthless, and the best educated segments of the population: the landed country gentlemen and their socially inferior cousins, the merchants and lawyers. By 1500 the essential economic basis for the landed country gentleman's future political and social ascendancy was being formed: the 15th-century knight of the shire was changing from a desperate and irresponsible land proprietor, ready to support the baronial feuding of the Wars of the Roses, into a respectable landowner desiring strong, practical government and the rule of law. The gentry did not care whether Henry VII's royal pedigree could bear close inspection; their own lineage was not above suspicion, and they were willing to serve the prince "in parliament, in council, in commission and other offices of the commonwealth."

Dynastic threats. It is no longer fashionable to call Henry VII a "new monarch," and indeed if the first Tudor had a model for reconstructing the monarchy, it was the example of the great medieval kings. Newness, however, should not be totally denied Henry Tudor; his royal blood was very "new," and the extraordinary efficiency of his regime introduced a spirit into government that had rarely been present in the medieval past. It was, in fact, "newness" that governed the early policy of the reign, for the Tudor dynasty had to be secured and all those with a better or older claim to the throne liquidated. Elizabeth of York was deftly handled by marriage; the sons of Edward IV had already been removed from the list, presumably murdered by their uncle Richard III; the Earl of Warwick was promptly imprisoned; but the descendants of Edward IV's sister and daughters remained a threat to the new government. Equally dangerous was the persistent myth that the younger of the two princes murdered in the Tower had escaped his assassin and that the Earl of Warwick had escaped his jailers. The existence of pretenders acted as a catalyst for further baronial discontent and Yorkist aspirations, and in 1487 John de la Pole, a nephew of Edward IV by his sister Elizabeth, with the support of 2,000 mercenary troops paid for with Burgundian gold, landed in England to support the pretensions of Lambert Simnel, who passed himself off as the authentic Earl of Warwick. Again Henry Tudor was triumphant in war; at the Battle of Stoke, de la Pole was killed and Simnel captured and demoted to a scullery boy in the royal kitchen. Ten years later Henry had to do it all over again, this time with a handsome Flemish lad named Perkin Warbeck, who for six years was accepted in Yorkist circles in Europe as the real Richard IV, brother of the murdered Edward V. Warbeck tried to take advantage of Cornish anger against heavy royal taxation and increased government efficiency and sought to lead a Cornish army of social malcontents against the Tudor throne. It was a measure of the new vigour and popularity of the Tudor monarchy, as well as the support of the gentry, that social revolution and further dynastic war were total failures, and Warbeck found himself in the Tower along with the Earl of Warwick. In the end both men proved too dangerous to live, even in captivity, and in 1499 they were executed.

The threats of pretenders

The policy of dynastic extermination did not cease with the new century. Under Henry VIII, the Duke of Buckingham, who was descended from the youngest son of Edward III, was destroyed in 1521; the Earl of Warwick's sister, the Countess of Salisbury, was beheaded in 1541 and her descendants harried out of the land; and in 1546 the poet Henry Howard, Earl of Surrey, the grandson of Buckingham, was put to death. By the end of Henry VIII's reign the job had been so well done that the curse of Edward III's fecundity had been replaced by the opposite problem—the Tudor line proved to be infertile when it came to producing healthy male heirs. Henry VII sired Arthur, who died in 1502, and Henry VIII in turn produced only one legitimate son, Edward VI, who died at the age of 16, thereby ending the direct male descent.

Financial policy. It was not enough for Henry VII to secure his dynasty; he also had to reestablish the financial credit of his crown and reassert the authority of royal law. Feudal kings had traditionally lived off four sources of nonparliamentary income: rents from the royal estates, revenues from import and export taxes, fees from the administration of justice, and moneys extracted on the basis of a vassal's duty to his overlord. The first Tudor was no different from his Yorkist or medieval predecessors; he was simply more ruthless and successful in demanding every penny that was owed him. Henry's first move was

Henry's sources of revenue

to confiscate all the estates of Yorkist adherents and to restore all property over which the crown had lost control since 1455 (in some cases as far back as 1377). To these essentially statutory steps he added efficiency of rent collection. In 1485 income from crown lands had totalled £29,000; by 1509 land revenues had risen to £42,000 and the profits from the Duchy of Lancaster had jumped from £650 to £6,500. At the same time, the Tudors profited from the growing economic prosperity of the realm, and custom receipts rose from over £20,000 to an average of £40,000 by the time Henry died.

The increase in custom and land revenues was applauded, for it meant fewer parliamentary subsidies and fitted the medieval formula that kings should live on their own, not parliamentary, income. But the collection of revenues from feudal sources and from the administration of justice caused great discontent and earned Henry his reputation as a miser and extortionist. Generally Henry demanded no more than his due as the highest feudal overlord, and a year after he became sovereign, he established a commission to look into land tenure to discover who held property by knight's fee—that is, by obligation to perform military services. Occasionally he overstepped the bounds of feudal decency and abused his rights. In 1504, for instance, he levied a feudal aid (tax) to pay for the knighting of his son—who had been knighted 15 years before and had been dead for two. Henry VIII continued his father's policy of fiscal feudalism, forcing through Parliament in 1536 the Statute of Uses to prevent landowners from escaping "relief" and wardship (feudal inheritance taxes) by legal trickery and establishing the Court of Wards and Liveries in 1540 to handle the profits of feudal wardship. The howl of protest was so great that in 1540 Henry VIII had to compromise, and by the Statute of Wills a subject who held his property by knight's fee was permitted to bequeath two-thirds of his land without feudal obligation.

Income from the law courts and from marriage arrangements

To fiscal feudalism Henry VII added rigorous administration of justice. As law became more effective, it also became more profitable, and the policy of levying heavy fines as punishment upon those who dared break the king's peace proved to be a useful whip over the mighty magnate and a welcome addition to the king's exchequer. Even war and diplomacy were sources of revenue; one of the major reasons Henry VII wanted his second son, Henry, to marry his brother's widow was that the king was reluctant to return the dowry of 200,000 crowns that Ferdinand and Isabella of Spain had given for the marriage of their daughter, Catherine of Aragon. Generally Henry believed in a good-neighbour policy—alliance with Spain by the marriage of Arthur and Catherine in 1501 and peace with Scotland by the marriage of his daughter Margaret to James IV in 1503—on the grounds that peace was cheap and trade profitable. In 1489, however, he was faced with the threat of the union of the Duchy of Brittany with the French crown; and England, Spain, the empire, and Burgundy went to war to stop it. Nevertheless, as soon as it became clear that nothing could prevent France from absorbing the duchy, Henry negotiated the unheroic but financially rewarding Treaty of Étaples in 1492, whereby he disclaimed all historic rights to French territory (except Calais) in return for an indemnity of £159,000. By fair means or foul, when the first Tudor died, his total non-parliamentary annual income had risen at least twofold and stood in the neighbourhood of £113,000 (some estimates are as high as £142,000). From land alone the king received £42,000, while the greatest landlord in the realm had to make do with less than £5,000; economically speaking, there were no longer any overmighty magnates.

The administration of justice. Money could buy power, but respect could only be won by law enforcement. The problem for Henry VII was not to replace an old system of government with a new—no Tudor was consciously a revolutionary—but to make the ancient system work tolerably well. He had to tame but not destroy the nobility, develop organs of administration directly under his control, and wipe out provincialism and privilege wherever they appeared. In the task of curbing the old nobility, the king was immeasurably helped by the high aristocratic death rate during the Wars of the Roses; but where war left off, policy took over. Commissions of Array composed of local notables were appointed by the crown for each county in order to make use of the power of the aristocracy in raising troops but to prevent them from maintaining private armies (livery) with which to intimidate justice (maintenance) or threaten the throne.

Previous monarchs had sought to enforce the laws against livery and maintenance, but the first two Tudors, though they never totally abolished such evils, built up a reasonably efficient machine for enforcing the law, based on the historic premise that the king in the midst of his council was the fountain of justice. Traditionally the royal council had heard all sorts of cases, and its members rapidly began to specialize. The Court of Chancery had for years dealt with civil offenses, and the Court of Star Chamber evolved to handle criminal cases, the Court of Requests poor men's suits, and the Court of Admiralty piracy. The process by which the conciliar courts developed was largely accidental, and the Court of Star Chamber acquired its name from the star-painted ceiling of the room in which the councillors sat, not from the statute of 1487 that recognized its existence. Conciliar justice was popular because the ordinary courts where common law prevailed were slow and cumbersome, favoured the rich and mighty, and tended to break down when asked to deal with riot, maintenance, livery, perjury, and fraud. The same search for efficiency applied to matters of finance. The traditional fiscal agency of the crown, the exchequer, was burdened down with archaic procedures and restrictions, and Henry VII turned to the more intimate and flexible departments of his personal household—specifically to the treasurer of the chamber, whom he could supervise directly—as the central tax-raising, rent-collecting, and money-disbursing segment of government.

The Tudors sought to enforce law in every corner of their kingdom, and step by step the blurred medieval profile of a realm shattered by semiautonomous franchises, in which local law and custom were obeyed more than the king's law, was transformed into the clear outline of a single state filled with loyal subjects obeying the king's decrees. By 1500 royal government had been extended into the northern counties and Wales by the creation of a Council of the North and a Council for the Welsh Marches. The Welsh principalities had always been difficult to control, and it was not until 1536 that Henry VIII brought royal law directly into Wales and incorporated the 136 self-governing lordships into a greater England with five new shires.

Henry VII's achievements

If the term "new monarchy" was inappropriate in 1485, the same cannot be said for the year of Henry VII's death, for when he died in 1509, after 24 years of reign, he bequeathed to his son something quite new in English history: a safe throne, a solvent government, a prosperous land, and a reasonably united kingdom. Only one vital aspect of the past remained untouched, the independent Roman Catholic church, and it was left to the second Tudor to destroy this remaining vestige of medievalism.

HENRY VIII (1509–47)

Cardinal Wolsey. A prince of 18 inherited his father's throne, but the son of an Ipswich butcher carried on the first Tudor's administrative policies. While the young sovereign enjoyed his inheritance, Thomas Wolsey collected titles—archbishop of York in 1514, lord chancellor and cardinal legate in 1515, and papal legate for life in 1524. He exercised a degree of power never before wielded by king or minister, for as lord chancellor and cardinal legate he united in his portly person the authority of church and state. He sought to tame both the lords temporal and spiritual, administering to the nobility the "new law of the Star Chamber," protecting the rights of the underprivileged in the poor men's Court of Requests, and teaching the abbots and bishops that they were subjects as well as ecclesiastical princes. Long before Henry assumed full power over his subjects' souls as well as their bodies, his servant had marked the way. The cardinal's administration, however, was stronger on promise than performance, and for all his fine qualities and many talents he exposed himself to the accusation that he prostituted policy for pecuniary gain and personal pride.

Foreign policy

Together, the king and cardinal plunged the kingdom into international politics and war and helped make England one of the centres of Renaissance learning and brilliance. But the sovereign and his chief servant overestimated England's international position in the continental struggle between Francis I of France and the emperor Charles V. Militarily, the kingdom was of the same magnitude as the papacy—the English king had about the same revenues and could field about the same size army—and, as one contemporary noted, England with its back door constantly exposed to Scotland and with its economy dependent upon the Flanders wool trade was a mere "morsel among those choppers" of Europe. Nevertheless, Wolsey's diplomacy was based on the expectation that England could swing the balance of power either to France or to the empire and by holding that position could maintain the peace of Europe. The hollowness of the cardinal's policy was revealed in 1525 when Charles disastrously defeated and captured Francis at the Battle of Pavia. Italy was overrun with the emperor's troops, the pope became an imperial chaplain, all of Europe bowed before the conqueror, and England sank from being the fulcrum of continental diplomacy to the level of a second-rate power just at the moment when Henry had decided to rid himself of his wife, the 42-year-old Catherine of Aragon.

The king's "Great Matter." It is still a subject of debate whether Henry's decision to seek an annulment of his marriage and wed Anne Boleyn was a matter of state, of love, or of conscience. Quite possibly all three operated; Catherine was fat, seven years her husband's senior, incapable of bearing further children, and Anne was everything that the queen was not—pretty, vivacious, and fruitful. Catherine had produced only one child to live past infancy and that was a girl, Princess Mary; it seemed ironic indeed that the first Tudor should have solved the question of the succession only to expose the kingdom to an even greater peril in the second generation: a female ruler. The need for a male heir was paramount, for the last queen of England, Matilda, in the 12th century, had been a disaster, and there was no reason to believe that another would be any better. Finally, there was the question of the king's conscience. Henry had married his brother's widow, and though the pope had granted a dispensation, the fact of the matter remained that every male child born to Henry and Catherine had died, and it was clearly written in Leviticus: "If a man takes his brother's wife, it is impurity; he has uncovered his brother's nakedness, they shall be childless" (20:21).

Unfortunately, Henry's annulment was not destined to stand or fall upon the theological issue of whether a papal dispensation could set aside such a prohibition, for Catherine was not simply the king's wife, she was also the aunt of the emperor Charles V, the most powerful sovereign in Europe. Both Henry and his cardinal knew that the annulment would never be granted unless the emperor's power in Italy could be overthrown by an Anglo-French military alliance and the pope rescued from imperial domination, and for three years Wolsey worked desperately to achieve this diplomatic and military end. Caught between an all-powerful emperor and a truculent English king, Clement VII procrastinated and offered all sorts of doubtful solutions short of annulment, including the marriage of Princess Mary and the king's illegitimate son, Henry Fitzroy, Duke of Richmond; the legitimizing of all children begotten of Anne Boleyn; and the suggestion that Catherine go into a nunnery so that the king could be given permission to remarry. Wolsey's purpose was to have the marriage annuled and the trial held in London, but in 1529, despite the arrival of Lorenzo Cardinal Campeggio to set up the machinery for a hearing, Wolsey's plans exploded. In July the pope ordered Campeggio to transfer the case to Rome, where a decision against the king was a foregone conclusion; and in August Francis and the emperor made peace at the Treaty of Cambrai. Wolsey's policies were a failure, and he was dismissed from office in October 1529. He died on November 29, just in time to escape trial for treason.

The fall of Wolsey

The Reformation background. Henry now began groping for new means to achieve his purpose. At first he contemplated little more than blackmail to frighten the pope into submission; but slowly, reluctantly, and not realizing the full consequences of his actions, he moved step by step to open defiance and a total break with Rome. Wolsey in his person and his policies had represented the past. He was the last of the great ecclesiastical statesmen who had been as much at home in the cosmopolitan world of European Christendom, with its spiritual centre in Rome, as in a provincial capital such as London. By the time of Henry's matrimonial crisis Christendom was dissolving. Not only were feudal kingdoms assuming the character of independent nation-states, but the spiritual unity of Christ's seamless cloak was also being torn apart by heresy. Possibly Henry would never have won his annulment had there not existed in England men who desired a break with Rome, not because it was dynastically expedient but because they regarded the pope as the "whore of Babylon."

The medieval church had become an anachronism out of touch with the 16th-century reality of changing economic practices, governmental structure, and social values. More and more God was French or German or English, and his representative in Rome was having ever greater difficulty in speaking so many languages and in persuading his international flock that he was the spiritual leader of all Christians and not simply a petty Italian potentate motivated by family ambition and political aggrandizement. The church was also withering from within. Historically it was a state within a state—an independent clerical body possessed of special rights and privileges because of the fundamental division of man into body and soul. In the eyes of many, however, the church's duties in matters spiritual had been superseded by matters temporal. Absenteeism and pluralism were rife, and by 1520 in Oxfordshire alone 58 percent of the county's 192 parish priests were absentees. Bishops and high ecclesiastics were meant to tend to the cure of souls, but in fact they were engrossed in worldly affairs. Wolsey himself, as the greatest and richest clerical statesman, seemed to epitomize the worst aspects of that worldliness and corruption. Men continued to go to church, but it was increasingly difficult, especially for the landed gentleman and the wealthy merchant, to respect the old church. A sure sign that zeal for the ancient structure was flagging was the economic decline of the monasteries: in Norfolk, Yorkshire, and Buckinghamshire the capital wealth of the religious foundations rose only 1.13 percent between 1480 and 1540, which was not enough to offset normal depreciation, let alone keep up with inflation. More and more surplus wealth was being directed into other than religious channels; in the 15th century the wool merchant Thomas Paycocke of Coggeshall had used the proceeds of trade to found a chantry to sing masses for his soul; a century later William Sanderson of London invested the profits of fishmongering into two small ships to carry Captain John Davis over the top of the world in search of the Northwest Passage to Cathay.

Abuses within the church

As the old church lived on in a fossilized condition, Christians looked elsewhere for inner contentment, and all over Europe men like Martin Luther, the German monk in Saxony, and Thomas Bilney, the Cambridge scholar in England, sought spiritual meaning and relief from ritualism, worldliness, and religious apathy. Luther in his monastery and Bilney in his college turned to the Bible, and each stumbled across the knowledge that even in the midst of despair faith in God's mercy could save sinners. The new religious ideas flowed into England largely in the form of Lutheran doctrines, but they found a receptive audience not only because there were upper-class individuals who could find no spiritual satisfaction in the old religious formulas and who were looking for exactly what Luther and Bilney had to offer but also because there existed in England a religious subculture in the form of Lollardy. Its existence had always been officially denied by the established church, but the ideas of John Wycliffe (d. 1384) had never been exterminated. They lived on just below the surface, and by the time of the Reformation Lollardy was once again becoming respectable. Though Henry himself was never a Protestant and even during the first 20 years of his reign was a zealous persecutor of religious

nonconformity, be it Lutheran or Lollard, he would never have been able to push through the break with Rome simply on the basis of anticlericalism or apathy within the existing church. If his headship of an independent English church was to live in "the hearts of his subjects" and not "post alone hidden in acts of parliament," he had to call upon the support of the "zely people" (Protestant zealots), who viewed the political and constitutional steps by which Henry's marriage to Anne Boleyn was legalized as being the prelude to a thorough spiritual reformation.

The break with Rome. With Wolsey and his papal authority gone, Henry turned to the authority of the state to obtain his annulment, and the so-called Reformation Parliament that first met in November 1529 was unprecedented—it lasted seven years, enacted 137 statutes (32 of which were of vital importance), and legislated in areas that no feudal Parliament had ever dreamed of entering. "King in Parliament" became the revolutionary instrument by which the medieval church was destroyed. The first step was to intimidate the church, and in 1531 Convocation was forced under threat of praemunire (a statute prohibiting the operation of the legal and financial jurisdiction of the pope without royal consent) to grant the sovereign a gift of £119,000 and to acknowledge him supreme head of the church "as far as the law of Christ allows." Then the government struck at the papacy, threatening to cut off its revenues; the Annates Statute of 1532 empowered Henry, if he saw fit, to abolish payment to Rome of the first year's income of all newly installed bishops. The implied threat had little effect on the pope; and time was running out, for by December 1532 Anne Boleyn was pregnant, and on Jan. 25, 1533, she was secretly married to Henry. If the king was to be saved from bigamy and his child born in wedlock, he had less than eight months to get rid of Catherine of Aragon. Archbishop William Warham conveniently died in August 1532, and in March 1533 a demoralized and frightened pontiff sanctioned the installation of Thomas Cranmer as primate of the English church. Cranmer was a friend of the annulment, but before he could oblige his sovereign, the queen's right of appeal from the archbishop's court to Rome had to be destroyed; and this could be done only by cutting the constitutional cords holding England to the papacy. Consequently, in April 1533 the crucial statute was enacted; the Act of Restraint of Appeals boldly decreed that "this realm of England is an empire." A month later an obliging archbishop heard the case and adjudged the king's marriage to be null and void. On June 1 Anne was crowned rightful queen of England, and three months and a week later, on Sept. 7, 1533, the royal child was born. To "the great shame and confusion" of astrologers, it turned out to be Elizabeth Tudor.

Thomas Cranmer and the annulment

Henry was mortified; he had risked his soul and his crown for yet another girl. But Anne had proved her fertility, and it was hoped that a male heir would shortly follow. In the meantime it was necessary to complete the break with Rome and rebuild the Church of England. By the Act of Succession of March 1534 subjects were ordered to accept the king's marriage to Anne as "undoubted, true, sincere and perfect." A second Annate's Statute severed most of the financial ties with Rome, and in November the constitutional revolution was solemnized in the Act of Supremacy, which announced that Henry Tudor was and always had been "Supreme Head of the Church of England"; not even the qualifying phrase "as far as the law of Christ allows" was retained.

The consolidation of the Reformation. The medieval tenet that church and state were separate entities with divine law standing higher than human law had been legislated out of existence; the new English church was in effect a department of the Tudor state. The destruction of the Roman Catholic church led inevitably to the dissolution of the monasteries. As monastic religious fervour and economic resources began to dry up, it was easy enough for the government to build a case that monasteries were centres of vice and corruption. In the end, however, what destroyed them was neither apathy nor abuse but the fact that they were contradictions within a national church, for religious foundations by definition were international, supranational organizations that traditionally supported papal authority. Though they bowed to the royal supremacy, the government continued to view them with suspicion, arguing that they had obeyed only out of fear, and their destruction got underway early in 1536. In the name of fiscal reform and efficiency, foundations with endowments of under £200 a year (nearly 400 of them) were dissolved on the grounds that they were too small to do their job effectively. By late 1536 confiscation had become state policy, for the Pilgrimage of Grace, a Roman Catholic-inspired uprising in the north, seemed to be clear evidence that all monasteries were potential nests of traitors. By 1539 the foundations, both great and small, were gone, and property worth possibly £2,000,000 was nationalized and incorporated into the crown lands, thereby almost doubling the government's normal peacetime, nonparliamentary income. Had those estates remained in the possession of the crown, English history might have been very different, for the kings of England would have been able to rule without calling upon Parliament, and the constitutional authority that evolved out of the crown's fiscal dependence on Parliament would never have developed. For better or for worse, Henry and his descendants had to sell the profits of the Reformation; and by 1603 three-fourths of the monastic loot had passed into the hands of the landed gentry. The legend of a "golden shower" is false: monastic property was never given away at bargain prices, nor was it consciously presented to the kingdom in order to win the support of the ruling elite. Instead, most of the land was sold at its fair market value to pay for Henry's wars and foreign policy. The effect, however, was crucial—the most powerful elements within Tudor society now had a vested interest in protecting their property against papal Catholicism.

The dissolution of the monasteries

The marriage to Anne, the break with Rome, and even the destruction of the monasteries went through with surprisingly little opposition. It had been foreseen that the royal supremacy might have to be enacted in blood, and the Act of Supremacy (March 1534) and the Act of Treason (December) were designed to root out and liquidate the dissent. The former was a loyalty test requiring subjects to take an oath swearing to accept not only the matrimonial results of the break with Rome but also the principles on which it stood; the latter extended the meaning of treason to include all those who did "maliciously wish, will or desire, by words or writing or by craft imagine" the king's death or slander his marriage. Sir Thomas More (who had succeeded Wolsey as lord chancellor), Bishop John Fisher (who almost alone among the episcopate had defended Catherine during her trial), and a handful of monks suffered death for their refusal to accept the concept of a national church. Even the Pilgrimage of Grace of 1536–37 was a short-lived eruption. The uprisings in Lincolnshire in October and in Yorkshire during the winter were without doubt religiously motivated, but they were also as much feudal and social rebellions as revolts in support of Rome. Peasants, landed country gentlemen, and feudal barons could unite in defense of the monasteries and the old religion, and for a moment the rebels seemed on the verge of toppling the Tudor state. The nobility were angered that they had been excluded from the king's government by men of inferior social status, and they resented the encroachment of bureaucracy into the northern shires. The gentry were concerned by rising taxes and the peasants by threatened enclosure; but the three elements had little in common outside religion, and the uprisings fell apart from within. The rebels were soon crushed and their leaders—including Robert Aske, one of the more pleasing figures of the century—brutally executed. The Reformation came to England piecemeal, which goes far to explain the government's success. Had the drift toward Protestantism, the royal supremacy, and the destruction of the monasteries come as a single religious revolution, it would have produced a violent reaction. As it was, the Roman Catholic opposition could always argue that each step along the way to Reformation would be the last.

The Acts of Supremacy and Treason

Henry's last years. Henry was so securely seated upon his throne that the French ambassador announced that he was more an idol to be worshiped than a king to be

obeyed. The king successfully survived four more matrimonial experiments, the enmity of every major power in Europe, and an international war. On May 19, 1536, Anne Boleyn's career was terminated by the executioner's ax. She had failed in her promise to produce further children to secure the succession. The king's love had turned to hatred, but what sealed the queen's fate was the death of her rival, Catherine of Aragon, on Jan. 8, 1536. From that moment it was clear that should Henry again marry, whoever was his wife, the children she might bear would be legitimate in the eyes of Roman Catholics and Protestants alike. How much policy, how much revulsion for Anne, how much attraction for Jane Seymour played in the final tragedy is beyond analysis, but 11 days after Anne's execution Henry married Jane. Sixteen months later the future Edward VI was born. The mother died as a consequence, but the father finally had what it had taken a revolution to achieve, a legitimate male heir.

Birth of Edward VI

Henry married thrice more, once for reasons of diplomacy, once for love, and once for peace and quiet. Anne of Cleves, his fourth wife, was the product of Reformation international politics. For a time in 1539 it looked as if Charles V and Francis would come to terms and unite against the schismatic king of England, and the only allies Henry possessed were the Lutheran princes of Germany. In something close to panic he was stampeded into marriage with Anne of Cleves. But the following year, the moment the diplomatic scene changed, he dropped both his wife and the man who had engineered the marriage, his vicar general in matters spiritual, Thomas Cromwell. Anne was divorced July 12, Cromwell was executed July 28, and Henry married Catherine Howard the same day. The second Catherine did not do as well as her cousin, the first Anne; she lasted only 18 months. Catherine proved to be neither a virgin before her wedding nor a particularly faithful damsel after her marriage. With the execution of his fifth wife, Henry turned into a sick old man, and he took as his last spouse Catherine Parr, who was as much a nursemaid as a wife. During those final years the king's interests turned to international affairs. Henry's last war (1543–46) was fought not to defend his church against resurgent European Catholicism but to renew a much older policy of military conquest in France. Though he enlarged the English Pale at Calais by seizing the small French port of Boulogne, the war had no lasting diplomatic or international effects except to assure that the monastic lands would pass into the hands of the gentry.

The achievements of Henry VIII

By the time Henry died (Jan. 28, 1547) medievalism had nearly vanished. The crown stood at the pinnacle of its power, able to demand and receive a degree of obedience from both great and small that no feudal monarch had been able to achieve. The measure of that authority was threefold: (1) the extent to which Henry had been able to thrust a very unpopular annulment and supremacy legislation down the throat of Parliament; (2) his success in raising unprecedented sums of money through taxation; and (3) his ability to establish a new church on the ashes of the old. It is difficult to say whether these feats were the work of the king or his chief minister, Thomas Cromwell. The will was probably Henry's, the parliamentary means his minister's, but whoever was responsible, by 1547 England was a long way along the road of Reformation. The crown had assumed the authority of the papacy without as yet fundamentally changing the old creed, but the ancient structure was severely shaken. Throughout England men were arguing that because the pontiff had been proved false, the entire Roman Catholic creed was suspect; and the cry went up to "get rid of the poison with the author." It was not long before every aspect of Roman Catholicism was under attack—the miracle of the mass whereby the bread and wine were converted into the body and blood of Christ, the doctrine of purgatory, the efficacy of saints and images, the concept of an ordained priesthood with miraculous powers, and the doctrine of the celibacy of the clergy. The time had come for Parliament and the supreme head to decide what constituted the "true" faith for Englishmen. Henry never worked out a consistent religious policy: the Ten Articles of 1536 and the *Bishop's Book* of the following year tended to be somewhat Lutheran in tone; the Six Articles of 1539, or the Act for Abolishing Diversity of Opinion, and the *King's Book* of 1543 were mildly Roman Catholic. Whatever the religious colouring, Henry's ecclesiastical *via media* was based on obedience to an authoritarian old king and on subjects who were expected to live "soberly, justly and devoutly." Unfortunately for the religious, social, and political peace of the kingdom, both these conditions disappeared the moment Henry died and a nine-year-old boy sat upon the throne.

EDWARD VI (1547–53)

Henry was legally succeeded by his son Edward VI, but power passed to his brother-in-law, Edward Seymour, Earl of Hertford, who became duke of Somerset and lord protector shortly after the new reign began. Seymour ruled in loco parentis; the divinity of the crown resided in the boy king, but authority was exercised by an uncle who proved himself to be more merciful than tactful, more idealistic than practical. Sweet reason and tolerance were substituted for the old king's brutal laws. The treason and heresy acts were repealed or modified, and the result came close to destroying the Tudor state. The moment idle tongues could speak with impunity, the kingdom broke into a chorus of religious and social discord. To stem religious dissent, the lord protector introduced the Prayer Book of 1549 and an act of uniformity to enforce it. Written by Thomas Cranmer, the Prayer Book was a literary masterpiece but a political flop, for it failed in its purpose. It sought to bring into a single Protestant fold all varieties of middle-of-the-road religious beliefs by deliberately obscuring the central issue of the exact nature of the mass—whether it was a miraculous sacrament or a commemorative service. The Prayer Book succeeded only in antagonizing Protestants and Roman Catholics alike.

Somerset was no more successful in solving the economic and social difficulties of the reign. Rising prices, debasement of the currency, and the cost of war had produced an inflationary crisis in which prices doubled between 1547 and 1549. A false prosperity ensued in which the wool trade boomed, but so also did enclosures with all their explosive potential. The result was social revolution. Whether Somerset deserved his title of "the good duke" is a matter of opinion. Certainly the peasants thought that he favoured the element in the House of Commons that was anxious to tax sheep raisers and to curb enclosures and that section of the clergy that was lashing out at economic inequality. In the summer of 1549 the peasantry in Cornwall and Devonshire revolted against the Prayer Book in the name of the good old religious days under Henry VIII, and almost simultaneously the humble folk in Norfolk rose up against the economic and social injustices of the century. At the same time that domestic rebellion was stirring, the protector had to face a political and international crisis, and he proved himself to be neither a farsighted statesman nor a shrewd politician. He embroiled the country in war with Scotland that soon involved France and ended in an inconclusive defeat, and he earned the enmity and disrespect of the members of his own council. In the eyes of the ruling elite he was responsible for governmental ineptitude and social and religious revolution. The result was inevitable: a palace revolution in October 1549 ensued in which Seymour was arrested and deprived of office, and two and a half years later he was executed on trumped-up charges of treason.

The fall of Seymour

The protector's successor and the man largely responsible for his fall was John Dudley, Earl of Warwick, who became duke of Northumberland. The duke was a man of action who represented most of the acquisitive aspects of the landed elements in society and who allied himself with the extreme section of the Protestant reformers. Under Northumberland, England pulled out of Scotland and in 1550 returned Boulogne to France; social order was ruthlessly reestablished in the countryside, the more conservative of the Henrician bishops were imprisoned, the wealth of the church was systematically looted, and uncompromising Protestantism was officially sanctioned. The Ordinal of 1550 transformed the divinely ordained priest into a governmental appointee, the new Prayer Book of 1552 was avowedly Protestant, altars were turned

into tables, clerical vestments gave way to plain surplices, and religious orthodoxy was enforced by a new and more stringent Act of Uniformity. How long a kingdom still attached to the outward trappings of Roman Catholicism would have tolerated doctrinal radicalism and the plundering of chantry lands and episcopal revenues under Somerset and Northumberland is difficult to say, but in 1553 the ground upon which Northumberland had built his power crumbled: Edward was dying of consumption. To save the kingdom from Roman Catholicism and himself from Roman Catholic Mary, who was Edward's legal heir, Northumberland, with the support, perhaps even the encouragement, of the dying king, tried his hand at kingmaking. Together they devised a new order of succession in which Mary was declared illegitimate and the crown passed to Lady Jane Grey, the granddaughter of Henry VIII's sister (Mary, Duchess of Suffolk), and incidentally Northumberland's daughter-in-law. The gamble failed, for when Edward died on July 6, 1553, the kingdom rallied to the daughter of Catherine of Aragon. Whatever their religious inclinations, Englishmen preferred a Tudor on the throne. In nine days the interlude was over, and Northumberland and his daughter-in-law were in the Tower of London.

Lady Jane Grey

MARY I (1553–58)

The new Roman Catholic queen had many fine qualities, and contemporaries announced that she was "a prince of heart and courage more than commonly is in womanhood"; but she was hopelessly outdated. She envisioned the return of a Roman Catholic church that had long since ceased to exist anywhere in Europe. The worldly and pliable church of pre-Reformation days had been destroyed by the fire of religious war and extremism, and both Catholic and Protestant now denied the tolerant humanistic principle that "men who live according to equity and justice shall be saved" no matter what their creed. For Mary it was a sacred obligation to return England to the Roman Catholic fold, and it was almost as great a duty to marry Philip of Spain, her Habsburg cousin and the son of Charles V, the man who had defended her mother's marital rights. She married Philip on July 25, 1554, and six months later, after the landed elements had been assured that their monastic property would not be taken from them, Parliament repealed the Act of Supremacy, reinstated the heresy laws, and petitioned for reunion with Rome. In the end both achievements proved sterile. Her marriage was without love or children, and, by associating Roman Catholicism in the popular mind with Spanish arrogance, it triggered a rebellion that almost overthrew the Tudor throne. In January 1554, under the leadership of Sir Thomas Wyat, the peasants of Kent rose up against the queen's Roman Catholic and Spanish policies, and 3,000 men marched on London. The rebellion was crushed, but it revealed to Mary and her chief minister, Reginald Cardinal Pole, that the kingdom was filled with disloyal hearts who placed Protestantism and nationalism higher than their obedience to the throne.

The tragedy of Mary's reign was the belief not only that the old church of her mother's day could be restored but also that it could be best served by fire and blood. Some 300 men and women were martyred in the Smithfield Fires during the last three years of her reign; compared to events on the Continent, the numbers were not large, but the emotional impact was great. Among the first half-dozen martyrs were the Protestant leaders Cranmer, Ridley, Latimer, and Hooper, who were burned to strike terror into the hearts of lesser men. Their deaths, however, had the opposite effect; their bravery encouraged others to withstand the flames, and the Smithfield Fires continued to burn because nobody could think of what to do with heretics except to put them to death. The law required it, the prisons were overflowing, and the martyrs themselves offered the government no way out except to enforce the grisly laws.

The Smithfield martyrs

Mary's reign was a study in failure. Her husband, who was 10 years her junior, remained in England as little as possible; the war between France and the Habsburg Empire, into which her Spanish marriage had dragged the kingdom, was a disaster and resulted in the loss of England's last continental outpost, Calais; her subjects learned to call her "bloody," and Englishmen greeted the news of her death and the succession of her sister Elizabeth on Nov. 17, 1558, with ringing bells and bonfires.

ELIZABETH I (1558–1603)

No one in 1558, any more than in 1485, would have predicted that despite the social discord, political floundering, and international humiliation of the past decade, the kingdom again stood on the threshold of an extraordinary reign. To make matters worse the new monarch was the wrong sex. Englishmen knew that it was unholy and unnatural that "a woman should reign and have empire above men." At 25, however, Elizabeth was better prepared than most women to have empire over men. She had survived the palace revolutions of her brother's reign and the Roman Catholicism of her sister's; she was the product of a fine Renaissance education, and she had learned the need for strong secular leadership devoid of religious bigotry. Moreover, she possessed her father's magnetism without his egotism or ruthlessness. She was also her mother's daughter, and the offspring of Anne Boleyn had no choice but to reestablish the royal supremacy and once again sever the ties with Rome.

Religious policy

Elizabeth's religious settlement was constructed on the doctrine of adiaphora, the belief that, except for a few fundamentals, there existed in religion a wide area of "things indifferent" that could be decided by the government on the basis of expediency. Conservative opposition was blunted by entitling the queen "supreme governor," not "head," of the church and by amending the Edwardian Prayer Book of 1552 to make it somewhat more acceptable to Roman Catholics. At the same time many of the old papal trappings of the church were retained. Protestant radicals went along with this compromise in the expectation that the principle of "things indifferent" meant that Elizabeth would, when the political dust had settled, rid her church of the "livery of Antichrist" and discard its "papal rags." In this they were badly mistaken, for the queen was determined to keep her religious settlement exactly as it had been negotiated in 1559. As it turned out, Roman Catholics proved to be better losers than Protestants: of the 900 parish clergy only 189 refused to accept Elizabeth as supreme governor, but the Protestant radicals—the future Puritans—were soon at loggerheads with their new sovereign.

The Tudor ideal of government. The religious settlement was part of a larger social arrangement that was authoritarian to its core. Elizabeth was determined to be queen in fact as well as in name. She tamed the House of Commons with tact combined with firmness, and she carried on a love affair with her kingdom in which womanhood, instead of being a disadvantage, became her greatest asset. The men she appointed to help her run and stage-manage the government were *politiques* like herself: William Cecil (later Lord Burghley), her principal secretary and in 1572 her lord treasurer; Matthew Parker, her archbishop of Canterbury; and a small group of other moderate and secular men.

In setting her house in order, the queen followed the hierarchical assumptions of her day. All creation was presumed to be a great chain of being, running from the tiniest insect to the godhead itself, and the universe was seen as an organic whole in which each part played a divinely prescribed role. In politics every element was expected to obey "one head, one governor, one law" in exactly the same way as all parts of the human body obeyed the brain. The crown was divine and gave leadership, but it did not exist alone, nor could it claim a monopoly of divinity, for all parts of the body politic had been created by God. The organ that spoke for the entire kingdom was not the king alone, but "King in Parliament," and, when Elizabeth sat in the midst of her Lords and Commons, it was said that "every Englishman is intended to be there present from the prince to the lowest person in England." The Tudors needed no standing army in "the French fashion" because God's will and the monarch's decrees were enshrined in acts of Parliament,

and this was society's greatest defense against rebellion. The controlling mind within this mystical union of crown and Parliament belonged to the queen. The Privy Council, acting as the spokesman of royalty, planned and initiated all legislation, and Parliament was expected to turn that legislation into law. Inside and outside Parliament the goal of Tudor government was benevolent paternalism in which the strong hand of authoritarianism was masked by the careful shaping of public opinion, the artistry of pomp and ceremony, and the deliberate effort to tie the ruling elite to the crown by catering to the financial and social aspirations of the landed country gentleman. Every aspect of government was intimate because it was small and rested on the support of probably no more than 5,000 key persons. The bureaucracy consisted of a handful of privy councillors at the top and at the bottom possibly 500 paid civil servants—the 15 members of the secretariat, the 265 clerks and custom officials of the treasury, a staff of 50 in the judiciary, and approximately 150 more scattered in other departments. Tudor government was not predominantly professional. Most of the work was done by unpaid amateurs: the sheriffs of the shires, the lord lieutenants of the counties, and above all the Tudor maids of all work—the 1,500 or so justices of the peace.

Benevolent paternalism

Smallness did not mean lack of government, for the 16th-century state was conceived of as an organic totality in which the possession of land carried with it duties of leadership and service to the throne, and the inferior part of society was obligated to accept the decisions of its elders and betters. The Tudors were essentially medieval in their economic and social philosophy. The aim of government was to curb competition and regulate life so as to attain an ordered and stable society in which all could share according to status. The Statute of Apprentices of 1563 embodied this concept, for it assumed the moral obligation of all men to work, the existence of divinely ordered social distinctions, and the need for the state to define and control all occupations in terms of their utility to society. The same assumption operated in the famous Elizabethan Poor Law of 1601—the need to assure a minimum standard of living to all men within an organic and noncompetitive society. By 1600 poverty, unemployment, and vagrancy had become too widespread for the church to handle, and the state had to take over, instructing each parish to levy taxes to pay for poor relief and to provide work for the able-bodied, punishment for the indolent, and charity for the sick, the aged, and the disabled. The Tudor social ideal was to achieve a static class structure by guaranteeing a fixed labour supply, restricting social mobility, curbing economic freedom, and creating a kingdom in which subjects could fulfill their ultimate purpose in life—spiritual salvation, not material well-being.

The Elizabethan Poor Law

Elizabethan society. Social reality, at least for the poor and powerless, was probably a far cry from the ideal, but for a few years Elizabethan England seemed to possess an extraordinary internal balance and external dynamism. In part the queen herself was responsible. She demanded no windows into men's souls, and she charmed both great and small with her artistry and tact. In part, however, the Elizabethan Age was a success because men had at their disposal new and exciting areas, both of mind and geography, into which to channel their energies. A revolution in reading and writing was taking place, and by 1640 nearly 100 percent of the gentry and merchant elements were literate. Wealth and literacy were directly related. Possibly 50 percent of the yeomanry but only 10 percent of the husbandry and none of the peasantry were able to read or write. Although literacy among townspeople was higher, the proportions relative to wealth still held true. The years between 1560 and 1650 were an age of school-building and educational endowment; by then 142 new schools had been founded and £293,000 given to grammar (secondary) school education. Oxford and Cambridge also reflected the new literacy, increasing from 800 students in 1560 to 1,200 in 1630. The aim of Tudor education was less to teach the "three Rs" than to establish mind control: to drill children "in the knowledge of their duty toward God, their prince and all other[s] in their degree." A knowledge of Latin and a smattering of Greek became, even more than elegant clothing, the mark of the social elite. The educated Englishman was no longer a cleric but a J.P. or M.P. (justice of the peace or member of Parliament), a merchant or a landed gentleman who for the first time was able to express his economic, political, and religious dreams and grievances in terms of abstract principles that were capable of galvanizing people into religious and political parties. Without literacy the spiritual impact of the Puritans or, later, the formation of parties based on ideologies that engulfed the kingdom in civil war would have been impossible. So also would have been the cultural explosion that produced Shakespeare, Marlowe, Spenser, Bacon, and Donne.

Poets, scholars, and playwrights dreamed and put pen to paper. Adventurers responded differently; they went "a-voyaging." From a kingdom that had once been known for its "sluggish security," Englishmen suddenly turned to the sea and the world that was opening up around them. The first hesitant steps had been taken under Henry VII when John Cabot in 1497 sailed in search of a Northwest Passage to China and as a consequence discovered Cape Breton Island. The search for Cathay became an economic necessity in 1550 when the wool trade collapsed and merchants had to find new markets for their cloth. In response, the Muscovy Company was established to trade with Russia, and by 1588, 100 vessels a year were visiting the Baltic. Martin Frobisher during the 1570s made a series of voyages to northern Canada in the hope of finding gold and a shortcut to the Orient; John Hawkins encroached upon Spanish and Portuguese preserves and sailed in 1562 for Africa in quest of slaves to sell to West Indian plantation owners; and Sir Francis Drake circumnavigated the globe (Dec. 13, 1577–Sept. 26, 1580) in search not only of the riches of the East Indies but also of Terra Australis, the great southern continent. Suddenly Englishmen were on the move: Sir Humphrey Gilbert and his band of settlers set forth for Newfoundland (1583); Sir Walter Raleigh organized the equally ill-fated "lost colony" at Roanoke (1587–91); John Davis in his two small ships, the *Moonshine* and the *Sunshine,* reached 72° north, the farthest north any Englishman had ever been (1585–87); and the honourable East India Company was founded to organize the silk and spice trade with the Orient on a permanent basis. The outpouring was inspired not only by the urge for riches but also by religion—the desire to labour in the Lord's vineyard and to found in the wilderness a new and better nation. As it was said, Englishmen went forth "to seek new worlds for gold, for praise, for glory." Even the dangers of the reign—the precariousness of Elizabeth's throne and the struggle with Roman Catholic Spain—somehow contrived to generate a self-confidence that had been lacking under "the little Tudors."

Mary, Queen of Scots. The first decade of Elizabeth's reign was relatively quiet, but after 1568 three interrelated matters set the stage for the crisis of the century: the queen's refusal to marry, the various plots to replace her with Mary of Scotland, and the religious and economic clash with Spain. Elizabeth Tudor's virginity was the cause of great international discussion, for every bachelor prince of Europe hoped to win a throne through marriage with Gloriana, and the source of even greater domestic concern, for everyone except the queen herself was convinced that Elizabeth should marry and produce heirs. The issue was the cause of her first major confrontation with the House of Commons, which was informed that royal matrimony was not a subject for commoners to discuss. Elizabeth preferred maidenhood—it was politically safer and her most useful diplomatic weapon—but it gave poignancy to the intrigues of her cousin Mary, Queen of Scots. Mary had been an unwanted visitor-prisoner in England ever since 1568, after she had been forced to abdicate her Scottish throne in favour of her 13-month-old son, James VI. She was Henry VIII's grandniece and, in the eyes of many Roman Catholics and a number of political malcontents, the rightful ruler of England, for Mary of Scotland was a Roman Catholic. As the religious hysteria mounted, there was steady pressure put on Elizabeth to rid England of this dangerous threat, but the queen delayed a final decision for almost 19 years. In the end, however, she

The question of the queen's marriage

had little choice. Jesuit priests were entering the kingdom to harden the hearts of the queen's subjects against her, forcing the government to introduce harsher and harsher recusancy laws (the fine for failure to attend Anglican service on Sundays was raised from one shilling a week to £20 a month). Puritans were thundering for even stiffer penalties, and Mary played into the hands of her religious and political enemies by involving herself in a series of schemes to unseat her cousin. One plot helped to trigger the rebellion of the northern earls in 1569. Another, the Ridolfi plot of 1571, called for an invasion by Spanish troops stationed in the Netherlands and resulted in the execution in 1572 of the Duke of Norfolk, the ranking peer of the realm. Yet another, the Babington plot of 1586, was in fact a carefully arranged government trap to gain sufficient evidence to have Mary tried and executed for high treason.

The clash with Spain. Mary was executed on Feb. 8, 1587; by then England had moved from cold war to open war against Spain. Philip II was the colossus of Europe and leader of resurgent Roman Catholicism. His kingdom was strong; Spanish troops were the best in Europe, Spain itself had been carved out of territory held by the infidel and still retained its crusading zeal, and the wealth of the New World poured into the treasury at Madrid. Spanish preeminence was directly related to the weakness of France, which ever since the accidental death of Henry II in 1559 had been torn by factional strife and civil and religious war. In response to this diplomatic and military imbalance, English foreign policy underwent a fundamental change. By the Treaty of Blois in 1572 England gave up its historic enmity with France, accepting by implication that Spain was the greater danger. It is difficult to say at what point a showdown between Elizabeth and her former brother-in-law became unavoidable—there were so many areas of disagreement—but the two chief points were the refusal of English merchants-cum-buccaneers to recognize Philip's claims to a monopoly of trade wherever the Spanish flag flew throughout the world and the military and financial support given by the English to Philip's rebellious and heretical subjects in the Netherlands.

The harassment of Spain at sea

The most blatant act of English poaching in Spanish imperial waters was Drake's circumnavigation of the Earth, during which Spanish shipping was looted, Spanish claims to California ignored, and Spanish world dominion proven to be a paper empire. But the encounter that really poisoned Anglo-Iberian relations was the Battle of San Juan de Ulúa in September 1568 where a small fleet captained by John Hawkins and Francis Drake was ambushed and almost annihilated through Spanish perfidy. Only Hawkins in the *Minion* and Drake in the *Judith* escaped. The English cried foul treachery, the Spanish dismissed the action as sensible tactics when dealing with pirates. Drake and Hawkins never forgot or forgave, and it was Hawkins who, as treasurer of the navy, began to build the revolutionary ships that destroyed the old-fashioned galleons of the Spanish Armada.

If the English never forgave Philip's treachery at San Juan de Ulúa, the Spanish never forgot Elizabeth's interference in the Netherlands, where Dutch Protestants were in full revolt. At first, aid had been limited to money and the harbouring of Dutch ships in English ports; but after the assassination of the Protestant leader, William of Orange, in 1584, the position of the rebels became so desperate that Elizabeth in August 1585 sent over an army of 6,000 under the command of the Earl of Leicester. Reluctantly, Philip decided on war against England as the only way of exterminating heresy and disciplining his subjects in the Netherlands. Methodically, he began to build a fleet of 130 vessels, 31,000 men, and 2,431 cannons to hold naval supremacy in the Channel long enough for the Duke of Parma's army, stationed at Dunkirk, to cross over to England. Nothing Elizabeth could do seemed to be able to stop the Armada Catholica. She sent Drake to Spain in April 1587 in a spectacular strike at that portion of the fleet forming at Cádiz, but it succeeded only in delaying the sailing date. That delay, however, was important, for Philip's Admiral of the Ocean Seas, the veteran Marquess de Santa Cruz, died, and the job of sailing the Armada was given to the Duke de Medina-Sidonia, who was invariably seasick and confessed that he knew more about gardening than war. What ensued was not the new commander's fault. He did the best he could in an impossible situation, for Philip's Armada was invincible in name only. It was technologically and numerically outclassed by an English fleet of close to 200. Worse, its strategic purpose was grounded on a fallacy: that Parma's troops could be conveyed to England. The Spanish controlled no deep-water port in the Netherlands in which the Armada's great galleons and Parma's light troop-carrying barges could rendezvous. Even the deity seemed to be more English than Spanish, and in the end the fleet, buffeted by gales, was dashed to pieces as it sought to escape home via the northern route around Scotland and Ireland. Of the 130 ships that left Spain, perhaps 85 crept home; 10 had been captured, sunk, or driven aground by English guns, 23 were sacrificed to wind and storm, and 12 others were "lost, fate unknown."

Defeat of the Armada

Internal discontent. When the Armada died during the first weeks of August 1588, the crisis of Elizabeth's reign was reached and successfully passed. The last years were an anticlimax, for the moment the international danger was surmounted, domestic strife ensued. There were moments of great heroism and success—as when Essex, Raleigh, and Howard made a second descent on Cádiz in 1596, seized the city, and burned the entire West Indian treasure fleet—but the war so gloriously begun deteriorated into a costly campaign in the Netherlands and France and an endless guerrilla action in Ireland, where Philip discovered he could do to Elizabeth what she had been doing to him in the Low Countries. Even on the high seas the days of fabulous victories were over, for the king of Spain soon learned to defend his empire and his treasure fleets. Both Drake and Hawkins died in 1596 on the same ill-conceived expedition into Spanish Caribbean waters—symbolic proof that the good old days of buccaneering were gone forever. At home the cost of almost two decades of war (£4 million) raised havoc with the queen's finances. It forced her to sell her capital (about £800,000, or roughly one-fourth of all crown lands) and increased her dependence upon parliamentary sources of income, which rose from an annual average of £35,000 to over £112,000 a year.

Elizabeth's financial difficulties were a symptom of a mounting political crisis that under her successors would destroy the entire Tudor system of government. The 1590s were years of depression—bad harvests, soaring prices, peasant unrest, high taxes, and increasing parliamentary criticism of the queen's economic policies and political leadership. Imperceptibly, the House of Commons was becoming the instrument through which the will of the landed classes could be heard and not an obliging organ of royal control. In Tudor political theory this was a distortion of the proper function of Parliament, which was meant to beseech and petition, never to command or initiate. Three things, however, forced theory to make way for reality. First was the government's financial dependence on the Commons, for the organ that paid the royal piper eventually demanded that it also call the governmental tune. Second, under the Tudors, Parliament had been summoned so often and forced to legislate on such crucial matters of church and state—legitimizing and bastardizing monarchs, breaking with Rome, proclaiming the supreme headship (governorship under Elizabeth), establishing the royal succession, and legislating in areas that no Parliament had ever dared enter before—that the Commons got into the habit of being consulted. Inevitably a different constitutional question emerged: if Parliament is asked to give authority to the crown, can it also take away that authority? Finally, there was the growth of a vocal, politically conscious, and economically dominant gentry; and the increase in the size of the House of Commons reflected the activity and importance of that class. In Henry VIII's first Parliament there were 74 knights who sat for 37 shires, and 224 burgesses who represented the chartered boroughs of the kingdom. By the end of Elizabeth's reign, borough representation had been increased by 135 seats. The Commons was replacing the Lords in importance because the

Elizabeth and Parliament

social element it represented had become economically and politically more important than the nobility. Should the crown's leadership falter, there existed by the end of the century an organization that was quite capable of seizing the political initiative, for as one disgruntled contemporary noted: "the foot taketh upon him the part of the head and commons is become a king." Elizabeth had sense enough to avoid a showdown with the Commons, and she retreated under parliamentary attack on the issue of her prerogative rights to grant monopolies regulating and licensing the economic life of the kingdom, but on the subject of her religious settlement she refused to budge.

By the last decade of the reign Puritanism was on the increase. During the 1570s and '80s "cells" had sprung up to spread God's word and rejuvenate the land, and Puritan strength was centred in exactly that segment of society that had the economic and social means to control the realm—the gentry and merchant classes. What set a Puritan off from other Protestants was the literalness with which he held to his creed, the discipline with which he watched his soul's health, the militancy of his faith, and the sense that he was somehow apart from the rest of corrupt humanity. This disciplined spiritual elite clashed with the queen over the purification of the church and the stamping out of the last vestiges of Roman Catholicism. The controversy went to the root of society: was the purpose of life spiritual or political, was the role of the church to serve God or the crown? In 1576 two brothers, Paul and Peter Wentworth, led the Puritan attack in the Commons, criticizing the queen for her refusal to allow Parliament to debate religious issues. The crisis came to a head in 1586, when Puritans called for legislation to abolish the episcopacy and the Anglican Prayer Book. Elizabeth ordered the bills to be withdrawn, and when Peter Wentworth raised the issue of freedom of speech in the Commons, she answered by clapping him in the Tower of London. There was emerging in England a group of religious idealists who derived their spiritual authority from a source that stood higher than the crown and who thereby violated the concept of the organic society and endangered the very existence of the Tudor paternalistic monarchy. As early as 1573 the threat had been recognized:

Paul and Peter Wentworth

> At the beginning it was but a cap, a surplice, and a tippet [over which Puritans complained]; now, it is grown to bishops, archbishops, and cathedral churches, to the overthrow of the established order, and to the Queen's authority in causes ecclesiastical.

James I later reduced the problem to one of his usual bons mots—"no bishop, no king." Elizabeth's answer was less catchy but more effective; she appointed as archbishop John Whitgift, who was determined to destroy Puritanism as a politically organized sect. Whitgift was only partially successful, but the queen was correct: the moment the international crisis was over and a premium was no longer placed on loyalty, Puritans were potential security risks.

The final years of Gloriana's life were difficult both for the theory of Tudor kingship and for Elizabeth herself. She began to lose hold over the imaginations of her subjects, and she faced the only palace revolution of her reign when her favourite, Robert Devereux, Earl of Essex, sought to touch her crown. There was still fight in the old queen, and Essex ended on the scaffold in 1601, but his angry demand could not be ignored:

> What! Cannot princes err? Cannot subjects receive wrong? Is an earthly power or authority infinite? Pardon me, pardon me, my good Lord, I can never subscribe to these principles.

When the queen died on March 24, 1603, it was as if the critics of her style of rule and her concept of government had been waiting patiently for her to step down. It was almost with relief that men looked forward to the problems of a new dynasty and a new century, as well as to a man, not a woman, upon the throne. (La.B.S.)

The early Stuarts and the Commonwealth

ENGLAND IN 1603

Economy and society. At the beginning of the 17th century, England and Wales contained more than four million people. The population had nearly doubled over the previous century, and it continued to grow for another 50 years. The heaviest concentrations of population were in the southeast and along the coasts. Population increase created severe social and economic problems, not the least of which was a long-term price inflation. English society was predominantly rural, with as much as 85 percent of its people living on the land. Small market towns of several hundred inhabitants facilitated local exchange, and in contrast to most of western Europe there were few large urban areas. Norwich and Bristol were the biggest provincial cities with populations of around 15,000. Exeter, York, and Newcastle were important regional centres, though they each had about 10,000 inhabitants. Only London could be ranked with the great continental cities. Its growth had outstripped even the doubling of the general population. By the beginning of the 17th century it contained more than a quarter of a million people and by the end nearly half a million, most of them poor migrants who flocked to the capital in search of work or charity. London was the centre of government, of overseas trade and finance, of fashion, taste, and culture. It was ruled by a merchant oligarchy, whose wealth increased tremendously over the course of the century as international trade expanded.

London not only ruled the English mercantile world, but it also dominated the rural economy of the southeast by its insatiable demand for food and clothing. The rural economy was predominately agricultural, with mixed animal and grain husbandry practiced wherever the land allowed. The population increase, however, placed great pressure upon the resources of local communities, and efforts by landlords and tenants to raise productivity for either profit or survival were the key feature of agricultural development. Systematic efforts to grow luxury market crops like wheat, especially in the environs of London, drove many smaller tenants from the land. So, too, did the practice of enclosure, which allowed for more productive land use by large holders at the expense of their poorer neighbours. There is evidence of a rural subsistence crisis lasting throughout the first two decades of the century. Marginally productive land came under the plow, rural revolts became more common, and harvest failures resulted in starvation rather than hunger. It was not until the middle of the century that the rural economy fully recovered and entered a period of sustained growth. A nation that could barely feed itself in 1600 was an exporter of grain by 1700.

Rural economy

In the northeast and southwest the harsher climate and poorer soils were more suited for sheep raising than for large-scale cereal production. The northeast and southwest were the location of the only significant manufacturing activity in England, the woolen cloth industry. Wool was spun into large cloths for export to Holland, where the highly technical finishing processes were performed before it was sold commercially. Because spinning and weaving provided employment for thousands of families, the downturn of the cloth trade at the beginning of the 17th century compounded the economic problems brought about by population increase. This situation worsened considerably after the opening of the Thirty Years' War (1618–48), as trade routes became disrupted and as new and cheaper sources of wool were developed. But the transformation of the English mercantile economy from its previous dependence upon a single commodity into a diversified entrepôt that transshipped dozens of domestic and colonial products was one of the most significant developments of the century.

The economic divide between rich and poor, between surplus and subsistence producers, was a principal determinant of rank and status. English society was organized hierarchically with a tightly defined ascending order of privileges and responsibilities. This hierarchy was as apparent in the family as in the state. There, as elsewhere, male domination was the rule; husbands ruled their wives, masters their servants, parents their children. But if hierarchy was stratified, it was not ossified; those who attained wealth could achieve status. The social hierarchy reflected gradations of wealth and responded to changes in the economic fortunes of individuals. In this sense it was more

Social hierarchy

open than most European societies; old wealth was not preferred to new, ancient title conferred no greater privileges than recent elevation; the humble could rise to become gentle, and the gentle could fall to become humble.

During the early 17th century a small titular aristocracy composed of between 75 and 100 peers formed the apex of the social structure. Their titles were hereditary, passed from father to eldest son, and they were among the wealthiest subjects of the state. Most were local magnates, inheriting vast county estates and occupying honorific positions in local government. The peerage was the military class of the nation, and in the counties peers held the office of lord lieutenant. Most were also called to serve at court, but at the beginning of the century their power was still local rather than central.

Gentry

Below them were the gentry, who probably composed only about 5 percent of the rural population but who were rising in importance and prestige. The gentry were not distinguished by title, though many were knights and several hundred purchased the rank of baronet after it was created in 1611. Sir Thomas Smith defined a member of the gentry as "he that can bear the port and charge of a gentleman." The gentry were expected to provide hospitality for their neighbours, treat their tenants paternally, and govern their counties. They served as deputy lieutenants, militia captains, and most importantly, as justices of the peace. To the justices fell the responsibility of enforcing the king's law and keeping the king's peace. They worked individually to mediate local disputes and collectively at quarter sessions to try petty crimes. As the magistracy the gentry were the backbone of county governance, and they maintained a fierce local independence even while enforcing the edicts of the crown.

Beneath the gentry were those who laboured for their survival. There were many prosperous tenants who were styled yeomen to denote their economic independence and the social gulf between them and those who eked out a bare existence. Some were the younger sons of gentlemen; others aspired to enter the ranks of the gentry, having amassed sufficient wealth to be secure against the fluctuations of the early modern economy. Like the gentry, the yeomanry were involved in local government, performing most of the day-to-day, face-to-face tasks. Yeomen were village elders, constables, and tax collectors, and they composed the juries that heard cases at quarter sessions. Most owned sufficient freehold land to be politically enfranchised and to participate in parliamentary selections. Filling out the ranks of rural society were husbandmen, cottagers, and labourers. They were the vast majority of local inhabitants, and their lives were bound up in the struggle for survival.

In towns, tradesmen and shopkeepers occupied the ranks below the ruling elites, but their occupational status clearly separated them from artisans, apprentices, and labourers. They were called the middling sort and were active in both civic and church affairs, holding the same minor offices as yeomen or husbandmen. Because of the greater concentrations of wealth and educational opportunities, the urban middling sort were active participants in urban politics.

Government and society. Seventeenth-century government was inextricably bound together with the social hierarchy that dominated local communities. Rank, status, and reputation were the criteria that enabled members of the local elite to serve the crown either in the counties or at court. Political theory stressed hierarchy, patriarchy, and deference in describing the natural order of English society. The most common visual description of this political community was the metaphor of the body politic. Like the human body, government and society were organic and their parts interdependent. Each element had its special and essential tasks to perform, without which the body could not function. At the head was the king, whose rule was based upon divine right and whose conception of his role in the state came closer to personal ownership than corporate management. Most of the aristocracy and gentry were the king's own tenants, whose obligations to him included military service, taxes, and local office holding. The monarch's claim to be God's vice-regent on earth was relatively uncontroversial, especially since his obligations to God included good governance. Except in dire emergency, the monarch could not abridge the laws and customs of England nor seize the persons or property of his subjects.

The body politic

The monarch ruled personally, and the permanent institutions of government were constantly being reshaped. Around the king was the court, a floating body of royal servants, officeholders, and place seekers. Personal service to the king was considered a social honour and thus fitting to those who already enjoyed rank and privilege. Most of the aristocracy and many gentlemen were in constant attendance at court, some with lucrative offices to defray their expenses, others extravagantly running through their fortunes. There was no essential preparation for royal service, no necessary skills or experiences. Commonly, members of the elite were educated at universities and the law courts, and most made a grand tour of Europe where they studied languages and culture. But their entry into royal service was normally through the patronage of family members and connections rather than through ability.

The privy council

From among his court the monarch chose a privy council. Its size and composition remained fluid, but it was largely composed of the chief officers of state: the lord treasurer, who oversaw revenue; the lord chancellor, who was the crown's chief legal officer; and the lord chamberlain, who was in charge of the king's household. The archbishop of Canterbury was the leading churchman of the realm, and he advised the king, who was the head of the established church. The privy council advised the king on foreign and domestic policy and was charged with the administration of government. It communicated with the host of unpaid local officials who governed in the communities, ordering the justices to enforce statutes or the deputy lieutenants to raise forces. In these tasks the privy councillors relied not only upon the king's warrant but upon their own local power and prestige as well. Thus, while the king was free to choose his own councillors, he was constrained to pick those who were capable of commanding respect. The advice that he received at the council table was from men who kept one eye on their localities and the other on the needs of central policy.

This interconnection between the centre and the localities was also seen in the composition of Parliament. Parliament was another of the king's councils, though its role in government was less well defined than the privy council's and its summoning was intermittent. In the early 17th century Parliament was less an institution than an event; it was convened when the king sought the aid of his subjects in the process of creating new laws or to provide extraordinary revenue. Like everything else in English society, Parliament was constituted in a hierarchy, composed of king, lords, and commons. Every peer of the realm was personally summoned to sit in the House of Lords, which was dominated by the greatest of the king's officers. The lower house was composed of representatives selected from the counties and boroughs of the nation. The House of Commons was growing as local communities petitioned for the right to be represented in Parliament and local gentry scrambled for the prestige of being chosen. It had 464 members in 1604 and 507 forty years later. Selection to the House of Commons was a mark of distinction, and many communities rotated the honour among their most important citizens and neighbours. Although there were elaborate regulations governing who could choose and who could be chosen, in fact very few members of the House of Commons were selected competitively. Contests for places were uncommon, and elections in which individual votes were cast were extremely rare.

Members of Parliament served the dual function of representing the views of the localities to the king and of representing the views of the king to the localities. Most were members of royal government, either at court or in their local communities, and nearly all had responsibility for enforcing the laws that were created at Westminster. Most parliaments were summoned to provide revenue in times of emergency, usually for defense, and most members were willing to provide it within appropriate limits. They came to Parliament to do the king's business, the business of their communities, and their own personal

business in London. Such conflicting obligations were not always easily resolved, but Parliament was not perceived as an institution in opposition to the king any more than the stomach was seen as opposing the head of the body. Upsets there were, and increasingly during the 17th century king and Parliament clashed over specific issues, but until the middle of the century they were part of one system of royal government.

JAMES I (1603–25)

James VI of Scotland (1567–1625) was the most experienced monarch to accede to the English throne since William the Conqueror as well as one of the greatest of all Scottish kings. A model of the philosopher prince, James wrote political treatises like *The Trew Law of a Free Monarchy* (1598), debated theology with learned divines, and reflected continually on the art of statecraft. He governed his poor nation by balancing its factions of clans and by restraining the enthusiastic leaders of its Presbyterian church. In Scotland, James was described as pleasing to look at and pleasing to hear. He was sober in habit, enjoyed vigorous exercise, and doted on his Danish wife, Anne, who had borne him two male heirs.

But for all of these qualities, James I was viewed with suspicion by his new subjects. Centuries of hostility between the two nations had created deep enmities, and these could be seen in English descriptions of the king. There he was characterized as hunchbacked and ugly, with a tongue too large for his mouth and a speech impediment that obscured his words. It was said that he drank to excess and spewed upon his filthy clothing. It was also rumoured that he was homosexual and preyed upon the young boys brought to service at court. This caricature, which has long dominated the popular view of James I, was largely the work of disappointed English office seekers whose pique clouded their observations and the judgments of generations of historians.

In fact, James showed his abilities from the first. In the counties through which he passed on his way to London he lavished royal bounty upon the elites who had been starved for honours during Elizabeth's parsimonious reign. He knighted hundreds as he went, enjoying the bountiful entertainments that formed such a contrast with his indigent homeland. He would never forget these first encounters with his English subjects, "their eyes flaming nothing but sparkles of affection." On his progress James also received a petition, putatively signed by a thousand ministers, calling his attention to the unfinished business of church reform.

Religious policy. The Millenary Petition (1603) initiated a debate over the religious establishment that James intended to defend. The king called a number of his leading bishops to hold a formal disputation with the reformers. The Hampton Court Conference (1604) saw the king in his element. He took a personal role in the debate and made clear that he hoped to find a place in his church for moderates of all stripes. It was only extremists that he intended to "harry from the land," those who, unlike the supporters of the Millenary Petition, sought to tear down the established church. The king responded favourably to the call for creating a better educated and better paid clergy and referred several doctrinal matters to the consideration of convocation. But only a few of the points raised by the petitioners found their way into the revised Canons of 1604. In fact, the most important result of the conference was the establishment of a commission to provide an authorized English translation of the Bible, the King James Version (1611).

Hampton Court Conference

Indeed, James's hope was that moderates of all persuasions, Roman Catholic and Protestant alike, might dwell together in his church. But his plan to find a formula to encompass Catholics within the Calvinist English church was overthrown by the hotheadedness of Guy Fawkes, a convert to Roman Catholicism, and his confederates, who conspired to assassinate the king, lords, and commons by blowing up the Houses of Parliament. The failure of the Gunpowder Plot (1605) led to reprisals against Catholics and prevented James from going any further than exhibiting humane leniency toward them in the later years of his reign. Nevertheless, James's ecumenical outlook did much to defuse religious conflict and led to 20 years of relative peace within the English church.

Finance and politics. To a king whose annual budget in Scotland was barely £50,000, England looked like the land of milk and honey. But in fact, James I inherited serious financial problems, which his own liberality quickly compounded. Elizabeth had left a debt of more than £400,000 and James, with a wife and two sons, had much larger household expenses than the unmarried queen. Land and duties from customs were the major sources of royal revenue, and it was James's good fortune that the latter increased dramatically after the judges ruled in Bate's case (1606) that the king could make impositions on imported commodities without the consent of Parliament. Two years later, under the direction of James's able minister Robert Cecil, Earl of Salisbury, impositions were levied on an expanded list of goods, and a revised book of rates (1608) was issued that increased the level of duties. By these measures customs revenues grew by £70,000 a year.

But even this windfall was not enough to stem the effects of inflation on the one hand and James's own free spending on the other. By 1606 royal debt was more than £600,000, and the crown's financial ministers had turned their attention to prerogative income from wardships, purveyance, and the discovery of concealed lands (*i.e.,* crown lands on which rents and dues were not being paid). The revival and rationalization of these ancient rights created an outcry. As early as 1604 Salisbury was examining proposals to commute these fiscal rights into an annual sum to be raised by a land tax. By 1610 negotiations began for the Great Contract between the king and his taxpaying subjects that aimed to raise £200,000 a year. But at the last moment both royal officials and leaders of the House of Commons backed away from the deal, the government believing that the sum was too low, the leaders of the Commons that a land tax was too unpopular. The failure of the Great Contract drove Salisbury to squeeze even more revenue out of the king's feudal rights, including the sale of titles. This policy violated the spirit of principles about property and personal liberty held by the governing classes and, along with impositions, was identified as a grievance during James's first parliaments.

There was much suspicion that the Scottish king would not understand the procedures and privileges of an English Parliament, and this was in evidence at the opening of the first session of the Parliament of 1604–10. The conventional ban upon the selection of outlaws to the Commons led to the Buckinghamshire Election Case (1604). The Commons reversed a decision by the lord chancellor and ordered Francis Goodwin, an outlaw, to be seated in the House. James clumsily intervened in the proceedings, stating that the privileges of the Commons had been granted by the grace of the monarch, a pronouncement that stirred the embers of Elizabethan disputes over parliamentary privilege. Although a compromise solution to the case was found, from this time forward the Commons took an active role in scrutinizing the returns of its members. A standing committee on elections was formed, and the freedom of members from arrest during sessions was reasserted. Some wanted to go even further and present the king with a defense of the ancient rights of their House. But this so-called apology was the work of a minority and was never accepted by the whole House or presented to the king.

Buckinghamshire Election Case

Factions and favourites. As in the previous reign, court politics were factionalized around noble groups tied together by kinship and interest. James had promoted members of the Howard family to places of leadership in his government; Henry Howard, Earl of Northampton, adeptly led a family group that included Thomas Howard, Earl of Suffolk, and Thomas Howard, Earl of Arundel. All managed to enrich themselves at the expense of the king, whose debts reached £900,000 by 1618. A stink of corruption pervaded the court during these years. The Howards formed the core of a pro-Spanish faction that desired better relations with Spain and better treatment of English Catholics. They also played upon the king's desire for peace in Europe.

The Howards were opposed by an anti-Spanish group

that included the queen, George Abbot, Archbishop of Canterbury, and William Herbert, Earl of Pembroke. This group wished to pursue an aggressively Protestant foreign policy and, after the opening of the Thirty Years' War, to support James's son-in-law, the elector Frederick of the Palatinate. It was the anti-Spanish group that introduced the king to George Villiers, reputedly one of the handsomest men in Europe. Through Villiers they sought a conduit to power.

George Villiers, Duke of Buckingham

Even at the time it was thought unseemly that a lover should be provided for the king at the connivance of the queen and the archbishop. But Villiers was nobody's fool, and, while he succeeded spectacularly in gaining James's confidence, he refused to be a cipher for those who had advanced him. Soon he had risen to the pinnacle of the aristocracy. First knighted in 1615, he was created duke of Buckingham in 1623, the first nonroyal duke in half a century. Buckingham proved an able politician. He supported the movement for fiscal reform that led to the disgrace of Lord Treasurer Suffolk and the promotion of Lionel Cranfield, later Earl of Middlesex. Cranfield, a skilled London merchant, took the royal accounts in hand and made the unpopular economies that kept government afloat.

Buckingham, whose power rested upon his relationship with the king, wholeheartedly supported James's desire to reestablish peace in Europe. For years James had angled to marry his son Charles to a Spanish princess. There were many obstacles to this plan, not the least of which was the insistence of the pope that the marriage lead to the reconversion of England to Roman Catholicism. When negotiations remained inconclusive, James, in 1621, called his third Parliament with the intention of asking for money to support the Protestant cause. By this means he hoped to bully Philip IV of Spain into concluding the marriage negotiations and into using his influence to put an end to the German war.

Parliament, believing that James intended to initiate a trade war with Spain, readily granted the king's request for subsidies. But some members mistakenly also believed that the king wished their advice on military matters and on the prince's marriage. When James learned that foreign policy was being debated in the lower House, he rebuked the members for their temerity in breaching the royal prerogative. Stunned, both because they thought that they were following the king's wishes and because they believed in their freedom to discuss such matters, members of the Commons prepared the Protestation of 1621, exculpating their conduct and setting forth a statement of the liberties of the House. James sent for the Commons journal and personally ripped the protestation from it. He reiterated his claim that royal marriages and foreign policy were beyond the ken of Parliament and dryly noted that less than a third of the elected members of the House had been present when the protestation was passed.

The Parliament of 1621 was a failure at all levels. No legislation other than the subsidy bill was passed; a simple misunderstanding among the members had led to a dramatic confrontation with the king; and judicial impeachments were revived, costing the king the services of Lord Chancellor Bacon. James, moreover, was unable to make any progress with the Spaniards, and supporting the European Protestants drained his revenue. By 1624 royal indebtedness had reached £1 million. The old king was clearly at the end of his power and influence. His health was visibly deteriorating, and his policies were openly derided in court and country. Prince Charles and Buckingham decided to take matters into their own hands. In 1623 they traveled incognito to Madrid.

Their gambit created as much consternation in England as it did in Spain. James wept inconsolably, believing that his son would be killed or imprisoned. The Spaniards saw the end of their purposely drawn-out negotiations. Every effort was made to keep Charles away from the infanta, and he only managed to catch two fleeting glimpses of the heavily veiled princess. Nevertheless, he confided in Buckingham that he was hopelessly in love. Buckingham and John Digby, Earl of Bristol, the ambassador to Spain, were almost powerless to prevent the most damaging concessions. Charles even confessed himself willing to be instructed in the Catholic faith. Yet the more the prince conceded, the more embarrassed the Spaniards became. Nothing short of an ultimate Catholic reestablishment in England would be satisfactory, and they began to raise obviously artificial barriers. Even the lovesick prince realized that he was being humiliated. Shame turned to rage as he and Buckingham journeyed home.

There they persuaded the bedridden king to call another Parliament for the purpose of declaring war on Spain. The Parliament of 1624 was given free rein. All manner of legislation was passed; subsidies for a trade war with Spain were voted; and issues of foreign policy were openly discussed. Firmly in control of political decision-making, Charles and Buckingham worked to stave off attacks upon James's fiscal policies, especially the granting of monopolies to royal favourites. The last Parliament of James's reign was his most successful. On March 27, 1625, the old king died.

CHARLES I (1625–49)

Father and son could hardly be more different than were James and Charles. Charles was shy and physically deformed. He had a speech defect that made his pronouncements painful for him and his audiences alike. Charles had not been raised to rule. His childhood had been spent in the shadow of his brother, Prince Henry, who had died in 1612, and Charles had little practical experience of government. He was introverted and clung tenaciously to a few intimates. His wife, Henrietta Maria, French, Roman Catholic, and hugely unpopular, received Charles's loyalty despite great political cost. So did Buckingham, who survived the change in monarchs and consolidated his grip on government.

The politics of war. Along with his kingdom, Charles I inherited a domestic economic crisis and the war with Spain. A series of bad grain harvests, continued dislocation of the cloth trade, and a virulent plague that killed tens of thousands all conspired against the new king. Under the pressure of economic crisis, members of the Parliament of 1625 were determined to reform the customs and to limit the crown's right to levy impositions. The traditional lifelong grant of tonnage and poundage was thus withheld from Charles so that reform could be considered. But reform was delayed and, despite the appearance of illegality, the king collected these levies to prevent bankruptcy.

The Spanish war progressed no better than the domestic economy. Buckingham organized an expedition to Cádiz, but its failure forced Charles to summon another Parliament. From the start the Parliament of 1626 was badly managed, and members of both Houses thirsted for Buckingham's blood. Where James had sacrificed his ministers to further policy, Charles would not. Parliament was dissolved without granting any subsidies.

The forced loan

On the advice of his council the king decided to replace the lost subsidies by a benevolence, or forced loan. Lists of wealthy subjects were compiled and privy councillors traveled the land, attempting to persuade them to give freely. Fiscally the loan was a great success, but politically it was a catastrophe. More than £260,000 was raised within a single year, a sum equal to five subsidies. But the demand for money that could not be gotten in a parliamentary way profoundly alienated the ruling elites. A number of prominent gentlemen refused to contribute to the loan, and they were imprisoned "by the special command of the king." In the Five Knights' Case (1627) the judges uneasily ruled that the prisoners could not be bailed as they were not accused of specific offenses. Lawyers, such as Sir Edward Coke, and country gentlemen, such as Sir John Eliot, now feared that the common law insufficiently protected their lives and liberties. This sentiment was compounded by the fact that soldiers were being billeted in citizens' homes, local militias were forced to raise, equip, and transport men to fight abroad, and provost marshals declared martial law in peaceful English communities.

Yet the extremity of these expedients was matched by the seriousness of the international situation. Incredibly, England was now at war with both France and Spain, and Buckingham was determined to restore his reputation. Instead, the campaign of 1627 was a disaster, and the duke's

landing at the Île de Ré a debacle. It was hard to see how Charles could protect him from his critics once the Parliament of 1628 assembled.

The defeats of 1627 made emergency taxation more necessary than ever, and the new Parliament, 27 of whose members had been imprisoned for refusing to contribute to the loan, assembled with a sense of profound disquiet. It was proposed to grant the king five subsidies for defense but to delay their passage until the Petition of Right (1628) could be prepared. The petition asserted four liberties: freedom from arbitrary arrest; freedom from nonparliamentary taxation; freedom from the billeting of troops; and freedom from martial law. Couched in the language of tradition, it was presented to the king as a restatement of ancient liberties. In this spirit he accepted it, more in hope of receiving his subsidies than in fear that the petition would restrain his actions.

Petition of Right

Between the two sessions of this Parliament the Duke of Buckingham was assassinated (August 23). While the king wept in his palace, people drank to the health of the assassin in the streets; Buckingham had become a symbol of all that was wrong in the nation. But with the favourite removed, there was a void in government. Buckingham had been in charge of military and domestic policy, and there was no one else who had the confidence of the king or the ability to direct the royal program. When Charles I, grief-stricken, attempted to manage the second session of Parliament by himself, all the tensions came to a head. In the Commons some members wanted to challenge violations of the Petition of Right, especially the continued collection of tonnage and poundage without parliamentary authority. Others were equally agitated about changes in religious policy caused by the emergence of Arminianism (see below). When the level of bitterness reached new heights, the king decided to end the session. But before he could do so, two hotheaded members physically restrained the speaker, while Three Resolutions (1629) condemning the collection of tonnage and poundage as well as the doctrine and practice of Arminianism were introduced. Parliament broke up in pandemonium, with both king and members shocked by the "carriage of diverse fiery spirits."

Peace and reform. The dissolution of the Parliament of 1628 in 1629 and the king's clear intention to govern for a period without this troublesome institution necessitated a reversal of policy. Over the next two years peace treaties ended England's fruitless involvement in continental warfare in which more than £2 million had been wasted and royal government brought into disrepute. The king was also able to pacify his subjects by launching a campaign of administrative and fiscal reform that finally allowed the crown to live within its own revenues. Customs increased to £500,000 as both European and North American trade expanded. Under capable ministers like Richard Weston, Earl of Portland, prerogative income also increased. Ancient precedents were carefully searched to ensure that the crown received its full and lawful dues. Fines were imposed on those who had not come forward to be knighted at the king's accession. These distraints of knighthood yielded more than £170,000. The boundaries of royal forests were resurveyed and encroachers fined. Fees in the court of wards were raised and procedures streamlined. With effort and application annual royal revenue reached £1 million.

The most important of Charles's fiscal schemes was not technically a design to squeeze monies into the royal coffers. While the king's own rights might underwrite the needs of government, they could do nothing toward maintaining the navy, England's sole military establishment. Thus Charles expanded the collection of ship money, an ancient levy by which revenue was raised for the outfitting of warships. Although ship money was normally only collected in the ports in times of emergency, Charles extended it to inland communities and declared pirates a national menace. At first there was little resistance to ship money, but, as it was levied year after year, questions about its legitimacy were raised. Hampden's case (1637) turned upon the king's emergency powers and divided the royal judges who narrowly decided for the crown. But legal opinion varied so significantly that revenue dropped and the stirring of a taxpayer revolt could be felt.

Ship money

Religious reform. Fears about the state of the church, which erupted at the end of the Parliament of 1628, had been building for several years. Charles had become drawn to a movement of church reform that aroused deep hostility among his Calvinist subjects. The doctrines of predestination and justification by faith alone formed the core of beliefs in the traditional English church. Yet slowly competing doctrines of free will and the importance of works along with faith, advocated by the Dutch theologian Jacobus Arminius, spread to the English church. Arminians were viewed as radical reformers despite the fact that their leaders were elevated to the highest positions in church government. In 1633 William Laud, one of the ablest of the Arminians, became archbishop of Canterbury. Laud stressed ceremony over preaching. He believed in the "beauty of holiness" and introduced measures to decorate churches and to separate the communion table from the congregation. Both of these practices were reminiscent of Roman Catholicism, and they came at a time when Protestants everywhere feared for the survival of their religion. Nor did it help that the queen openly attended mass along with some highly placed converted courtiers. Anti-popery was the single strain that had united the diverse elements of Protestant reform, and it was now a rallying cry against innovations at home rather than abominations abroad.

Arminianism

Laud's movement for church reform was soon exported to the king's Presbyterian subjects in Scotland. Charles ruled three kingdoms with different customs, laws, and religions. His desire to bring them into conformity was always strong, though his ability to do so was weak. In 1637 the attempt to introduce a new prayer book in Scotland was met with riot and ultimately with rebellion. Neither the king nor the Scots would brook an attack on what each saw as their own church. In 1638 a Scottish National Covenant bound the nation to resist innovations in religion, and the following year an army of defense was raised.

The Bishops' Wars (1639–40) brought an end to the tranquillity of the 1630s. Charles had to meet rebellion with force, and force required money from Parliament. He genuinely believed that he would be supported against the rebels, failing to comprehend the profound hostility that Laud's innovations had created in England. The Short Parliament (1640) lasted less than a month before the king dissolved it rather than permit an extended discussion of his inadequacies. He scraped some money together and placed his troops under the command of his able and ruthless deputy, Thomas Wentworth, Earl of Strafford. But English troops fighting for pay proved no match for Scottish troops fighting for religion. In 1640 the Scots invaded England and captured Newcastle, the vital source of London's coal. Charles was forced to accept a humiliating treaty whereby he paid for the upkeep of the Scottish army and agreed to call another Parliament.

The Long Parliament. With his circumstances more desperate than ever, Charles I summoned Parliament to meet in November 1640. The king faced a body profoundly mistrustful of his intentions. The reform movement in the Commons was led by John Pym, a Somerset yeoman who was prominent by his oratorical skills in debate and his political skills in committee. Pym was a moderate, and for the next three years he ably steered compromises between those who wanted too much and those who would settle for too little. In the Lords, Viscount Say and Sele and the Earls of Warwick and Manchester worked in tandem with Pym and his allies, leading or following as occasion required.

The Long Parliament (1640–53) opened with the imprisonment of Strafford and Laud, the architects of the Scottish fiasco. Strafford was put on trial and ultimately attainted for treason. The dubious legality of the charges against him forced the Commons to proceed by bill rather than impeachment, and thus both Lords and monarch had to approve the charge. The Lords were cowed by London mobs and Charles by the mistaken belief that Strafford's blood would placate his opponents. But Strafford's execution in May was just the beginning.

In fact, parliamentary reform took two different tacks. The first was to limit the king's constitutional authority in order to protect the existence of Parliament and the

liberties of the subject. The second was to reconstitute the church. In February the Triennial Act (1641) was passed, mandating the summoning of Parliament every three years. In May the king's power to dissolve the Long Parliament was removed. Charles was forced to accept both bills. Meanwhile, the Commons relentlessly investigated the legal basis of the king's fiscal expedients, amending the laws that Charles had so scrupulously followed. Ship money and distraints of knighthood were declared illegal, royal forests were defined, and the prerogative courts of High Commission and Star Chamber were abolished. Again the king acceded.

Triennial Act

Church reform proved more treacherous. Parliamentary leaders agreed that Charles and Laud had introduced intolerable innovations, but where some were satisfied by their removal, others wished that they be replaced by even greater novelties. In December 1640 an orchestrated petitioning campaign called upon Parliament to abolish episcopacy, root and branch. Pym and his supporters were as yet unwilling to propose such a sweeping change, fearing lest it divide the Commons and create a crisis with the Lords. Nevertheless, the equally radical proposal to remove the bishops from the upper House was passed in May, and when the Lords rejected it, the Commons responded with the Root and Branch Bill.

Pym's fear that the religious issue might break apart the parliamentary consensus was compounded by his fear of provoking the king to counterattack. Throughout the first six months of the session Charles had meekly followed Parliament's lead. But there were ominous signs that the worm would turn. His leading advisers, the queen among them, were searching for military options. The radical attack upon the church allowed the king to portray himself as the conservator of religious matters as "they were in the purest time of Queen Elizabeth's days." Week by week sympathy for the king was growing, and in August Charles determined to conclude a peace treaty with the Scots. This successful negotiation removed the crisis that had brought the Long Parliament into being. When Charles returned to London at the end of November, he was met by cheering crowds and a large body of members of the two Houses.

While the king resolved one crisis in Scotland, another emerged in Ireland. Catholics, stung by the harsh repression of Strafford's rule, rose against their Protestant overlords and slaughtered thousands in a bloody rebellion. Though the reality was grim enough, the exaggerated reports that reached London seemed to fulfill the worst fears of a popish plot. Urgently an army had to be raised, but only the king had military authority, and in the present circumstance he could not be trusted with a force that might be used in London rather than Londonderry. In despair over the situation in Ireland and deeply suspicious of the king's intentions, the leaders of the Long Parliament debated the Grand Remonstrance, a catalog of their grievances against the king.

The Grand Remonstrance (1641) divided the Commons as nothing else had. It passed by only 11 votes, and the move to have it printed failed. Many were appalled that the remonstrance was to be used as propaganda "to tell stories to the people." For the first time members of Commons began to coalesce into opposing factions of royalists and parliamentarians.

The Grand Remonstrance

The passage of the Grand Remonstrance was followed by Pym's attempt to create a militia. Bills were proposed to put the army under parliamentary control and to give Parliament the right to nominate officers. The political situation had reached a state of crisis. In Parliament rumours spread of a royal attack upon the Houses, and at court wild talk of an impeachment of the queen was reported. It was Charles who broke the deadlock. On Jan. 4, 1642, he rode to Westminster intending to impeach five members of the Commons and one of the Lords on charges of treason. But, because the king's plan was no secret, the members had already fled. Thus Charles's dramatic breach of parliamentary privilege badly backfired. He not only failed to obtain his objective but also lost the confidence of many moderates left in Parliament. After ensuring the safe departure of his wife and children out of the country, Charles abandoned his capital and headed north.

The initiative had returned to Pym and his allies, who now proceeded to pass much of their stalled legislation, including the exclusion of the bishops from the Lords and the Impressment Bill (1642), which allowed Parliament to raise the army for Ireland. In June a series of proposals for a treaty, the Nineteen Propositions (1642), were presented to the king. They called for parliamentary control over the militia, the choice of royal counselors, and religious reform. Charles rejected them outright, though in his answer he seemed to grant Parliament a coordinate power in government, making the king but one of the three estates. The king, however, had determined to settle the matter by main force. His principal advisers believed that the greatest lords and gentlemen would rally to their king and that Parliament would not have the stomach for rebellion. On Aug. 22, 1642, the king raised his standard bearing the device "Give Caesar His Due."

The Nineteen Propositions

Civil war and revolution. The civil war that began in 1642 was one that neither king, Parliament, nor nation wanted. It was a war that was as dangerous to win as to lose. The parliamentarians could only maintain the fiction that they were fighting to "preserve the safety of the king," as the commission of their commander, Robert Devereux, Earl of Essex, stated. The king's fiction was that he was opposing a rebellion. Most of the nation remained neutral, hoping that differences would be composed and fighting ended.

The first years of war were as halfhearted as these justifications. Parliament held the tactical advantages of controlling the navy and London. While the navy protected the coast from foreign invasion, London provided the funds and manpower for battle. The king held the strategic advantage of knowing that he had to recapture his capital. He relied upon the aristocracy for men and arms. In the first substantial engagement of the war, at Edgehill (1642), Charles's cavalry proved superior to Parliament's, and he followed this first encounter by marching to the capital. At Brentford (1642) on the outskirts of London, the City militia narrowly averted the king's triumph. For the next two years, however, the war was fought to a desultory standstill.

Almost from the beginning the members of Parliament were divided over their goals. A war group argued that Charles could not be trusted until he learned the lesson of military defeat. A peace group countered that the longer the war ground on, the less likely Charles would be to compromise. Both of these groups were loose coalitions, and neither of them dominated parliamentary politics. Until his death in 1643 Pym steered a course between them, supporting the Oxford Propositions (1643) for peace as well as creating the administrative machinery to raise and finance armies. The excise, modeled on impositions, and the monthly assessments, modeled on ship money, increased levels of taxation to new heights. The king burdened the communities his forces controlled just as heavily.

In 1643 the war widened. Charles negotiated a cease-fire with the Catholic rebels in Ireland that allowed him to bring Irish troops to England. Parliament negotiated the Solemn League and Covenant (1643) with the Scots, who brought an army to England in return for guarantees of a Presbyterian church establishment. Initially Parliament benefited most. A combination of English and Scottish troops defeated royalist forces at Marston Moor (1644) and took York. But ultimately, religious differences between Scottish Presbyterians and English Independents vitiated the alliance. As the parliamentary commanders bickered, their forces were defeated at Lostwithiel (1644) and at Newbury (1644). While another round of peace negotiations began, the unsuccessful Uxbridge Proposals (1645), Parliament recast its military establishment and formed the New Model Army.

There was little new about the New Model Army other than centralization. Remnants of three armies were combined to be directed by a parliamentary committee. This committee included the parliamentary generals who were displaced by the Self-Denying Ordinance (1645), an act that excluded members of Parliament from civil and military office. The New Model Army was commanded by

The New Model Army

Sir Thomas Fairfax, and eventually the cavalry was led by Lieutenant-General Oliver Cromwell.

The new parliamentary army was thought so weak that the king hoped to crush it in a single blow and thus end the war. Instead, the Battle of Naseby on June 14, 1645, delivered the decisive blow to the royalists. Even though the parliamentary forces only just managed to carry the day, despite their numerical superiority, their victory was decisive. It destroyed the king's main armies and left open a path to the west, where his other substantial forces were defeated at Langport (1645). The following year the king surrendered to the Scots, erroneously believing that they would strike a better bargain.

For four years the political divisions at Westminster had been held in check by the military emergency. But the king's defeat released all restraints. In Parliament coherent parties began to form around the religious poles provided by Presbyterians and Independents and around the political poles of peace and war. Denzil Holles, one of the five members of Parliament Charles had tried to arrest in 1642, came to head the most powerful group. He pushed through a Presbyterian church settlement, negotiated a large loan from the city of London, and used the money to ransom the king from the Scots. Holles' peace plan was to remove the main points of difference between king and Parliament by disbanding the army and settling the disputes about the church, the militia, and the rebellion in Ireland. His party was opposed by a group led by Sir Henry Vane the Younger and Oliver Cromwell, who desired toleration for Independents and were fearful of disbanding the army before an agreement was reached with Charles I.

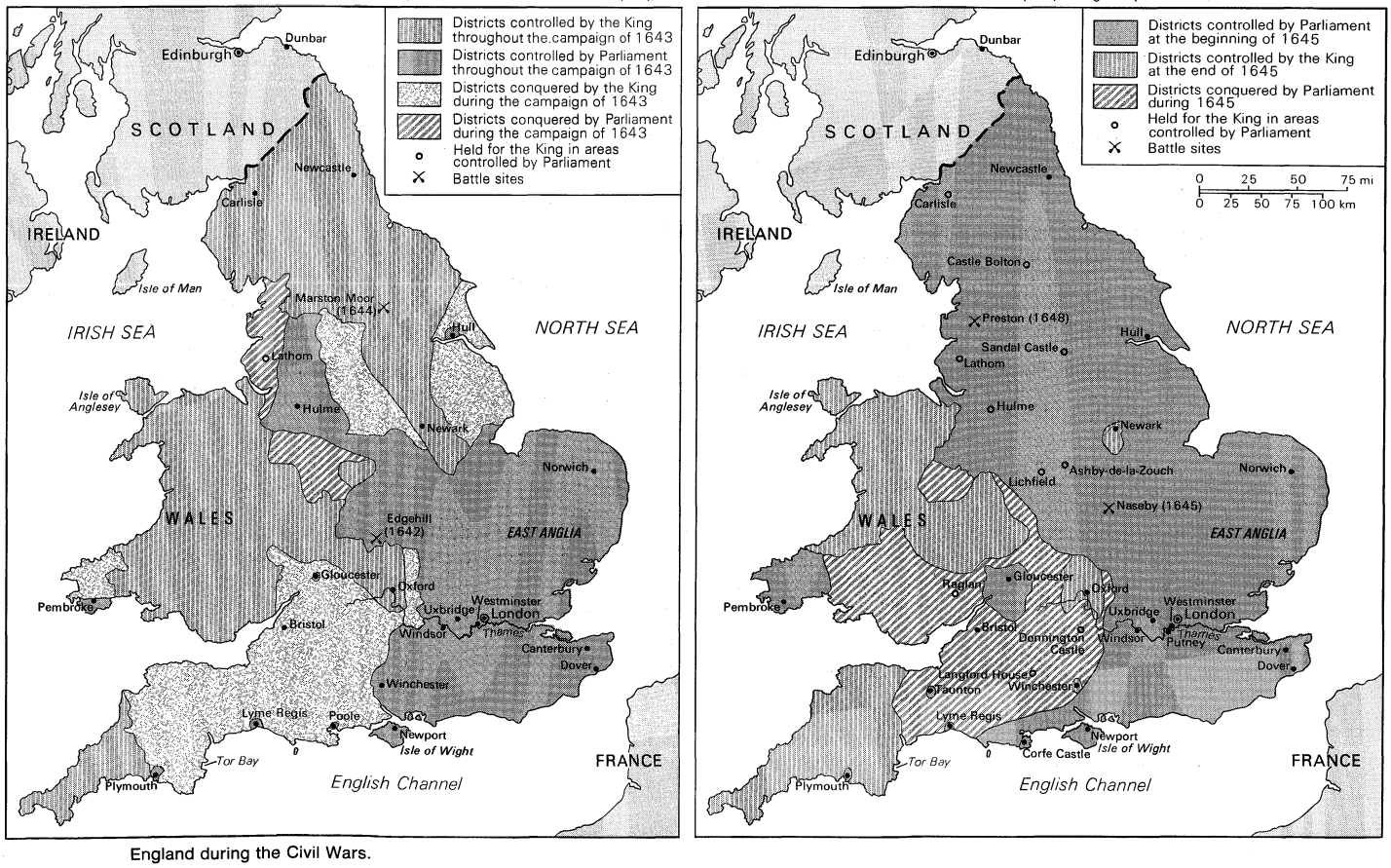
But war weariness in both Parliament and nation swept all before it. In January 1647 Charles was returned to English custody, and Holles moved forward with his plan to send a portion of the army to Ireland, assign a small force to English garrisons, and disband the rest. But in this he reckoned without the army. In the rank and file, concern about arrears of pay, indemnity, and liability for impressment stirred the soldiers to resist Irish service. A movement that began over material grievances soon turned political as representatives were chosen from the rank and file to present demands through their officers to Parliament. Holles attempted to brush this movement aside and to push through his disbandment scheme. At this the army rose up, driving out those of its officers who supported the disbandment, seizing Charles at Holmby House on June 3 and demanding the impeachment of Holles and his main supporters. At the beginning of August 1647 the army marched into London, and Holles with 10 of his allies fled the capital.

The army's intervention transformed civil war into revolution. Parliament, which in 1646 had argued that it was the fundamental authority in the nation, by 1647 was but a pawn in a new game of power politics. The perceived corruption of Parliament made it, like the king, a target of reform. Initiative was now in the hands of the king and the army, and Charles I tried to entice Cromwell and Henry Ireton, the army's leading strategist, to bargain his restoration for a tolerant church settlement. But the officers were only one part of a politicized army that was bombarded with plans for reorganizing the state. Among the most potent plans were those of the Levelers, led by John Lilburne, who desired that a new compact between ruler and ruled, the Agreement of the People (1647), be made. This was debated by the council of the army at Putney in October. The Levelers' proposals, which had much in common with the army's, called for the reform of Parliament through elections based upon a broad franchise and for a generally tolerant church settlement. Turmoil in the army led Fairfax and Cromwell to reassert military discipline, while the machinations of Charles led to the second Civil War (1648).

The Levelers

Charles had now managed to join his English supporters with discontented Scots who opposed the army's intervention in politics. Though the fighting was brief, it was bloody. Fairfax stormed Colchester (1648) and executed the ringleaders of the English rebellion, and Cromwell and several New Model regiments defeated the invading Scots at the Battle of Preston (1648).

Adapted from R. Treharne and H. Fullard (eds.), *Muir's Historical Atlas: Ancient, Medieval and Modern*, 9th ed. (1965); George Philip & Son Ltd., London



England during the Civil Wars.

The second Civil War hardened attitudes in the army. The king was directly blamed for the unnecessary loss of life, and for the first time alternatives to Charles Stuart, "that man of blood," were openly contemplated. Parliament, too, was appalled by the renewal of fighting. Moderate members believed that there was still a chance to bring the king to terms, despite the fact that he had rejected treaty after treaty. While the army made plans to put the king on trial, Parliament summoned its strength for one last negotiation, the abortive Treaty of Newport. Even now the king remained intransigent, especially over the issue of episcopacy. New negotiations infuriated the army because it believed that Parliament would sell out its sacrifices and compromise its ideals. On Dec. 6, 1648, army troops, under the direction of Colonel Thomas Pride, purged the House of Commons. Forty-five members were arrested and 186 were kept away. A rump of about 75 active members were left to do the army's bidding. They were to establish a High Court of Justice, prepare a charge of treason against the king, and place him on trial in the name of the people of England. Pride's Purge was a last-minute compromise made to prevent absolute military rule. With Cromwell deliberately absent in the north, Ireton was left to stave off the argument, made by the Levelers, that Parliament was hopelessly corrupt and should be dissolved. The decision to proceed by trial in the High Court of Parliament was a decision in favour of constitutional forms, however much a shadow they had become.

The Rump Parliament

The king's trial took place at the end of January. The Court of Justice was composed of members of Parliament, civilians, and army officers. There was little enthusiasm for the work that had to be done. No more senior judge than John Bradshaw could be found to preside, and he wore a hat ringed with iron in fear of assassination. The charges against the king, however politically correct, had little legal basis, and Charles deftly exposed their weakness. But like Strafford before him, Charles was to be sacrificed to the law of necessity if not the law of England. On Jan. 30, 1649, at the wall of his own palace, Charles I was beheaded. A witness recorded in his diary, "such a groan went up as I had never before heard."

Commonwealth and Protectorate. The execution of the king aroused hostility not only in England but also throughout Europe. Regicide was considered the worst of all crimes, and not even the brilliance of John Milton in *The Tenure of Kings and Magistrates* (1649) could persuade either Catholic or Protestant powers that the execution of Charles I was just. Open season was declared against English shipping, and Charles II was encouraged to reclaim his father's kingdom.

Despite opposition and continued external threats, the government of the Commonwealth was declared in May 1649 after acts had been passed to abolish the monarchy and the House of Lords. Political power resided in a Council of State, the Rump Parliament (which swelled from 75 to 213 members in the year following the king's execution), and the army. The military was now a permanent part of English government. Though the soldiers had assigned the complex tasks of reform to Parliament, they made sure of their ability to intervene in political affairs.

At first, however, the soldiers had other things to occupy them. For reasons of security and revenge, Ireland had to be pacified. In the autumn Oliver Cromwell, with a large detachment of troops, defeated royalist and Catholic forces at Drogheda and Wexford (1649), where he put civilian populations to the sword. Large parts of the northern Irish provinces were confiscated and used to pay off soldiers' arrears. While the suffering that Cromwell meted out horrified the Irish, it propitiated those who demanded revenge for the massacres of 1641. Cromwell's reputation soared, and he was the logical choice to command the forces that now had to face another grave challenge. In 1650 Charles II landed in Scotland, was declared king, and assembled a formidable army. Cromwell's military skills were never more in evidence than at Dunbar (1650), where his heavily outnumbered forces decisively defeated the Scots. But the war dragged on for another year. When England was invaded, Cromwell again defeated Scottish forces at Worcester (1651). Charles II barely escaped with his life.

Military victories in Ireland and Scotland

Victorious wars against the Irish, Scots, and Dutch (1652) made the Commonwealth a feared military power. But the struggle for survival defined the Rump's conservative policies. Little was done to reform the law. An attempt to abolish the court of chancery created chaos in the central courts. Little agreement could be reached on religious matters, especially on the vexing question of the compulsory payment of tithes. Most ominously, nothing at all had been done to set a limit for the sitting of the Rump and to provide for franchise reform and the election of a new Parliament. This had been the principal demand of the army, and the more the Rump protested the difficulty of the problem, the less patient the soldiers became. In April, when it was clear that the Rump would set a limit to its sitting but would nominate its own members to judge new elections, Cromwell marched to Westminster and dissolved Parliament. The Rump was replaced by an assembly nominated mostly by the army high command. The Nominated Parliament (1653) was no better able to overcome its internal divisions or untangle the threads of reform than the Rump. After five months it dissolved itself and returned power to Cromwell and the army.

The problems that beset both Rump and Nominated parliaments resulted from the diversity of groups that supported the revolution, ranging from pragmatic men of affairs, lawyers, officeholders, and local magistrates whose principal desire was to restore and maintain order to zealous visionaries who wished to establish heaven on earth. The republicans, like Sir Henry Vane the Younger, hoped to create a government based upon the model of ancient Rome and modern Venice. They were proud of the achievements of the Commonwealth and reviled Cromwell for dissolving the Rump. But most political reformers based their programs on dreams of the future rather than the past. They were millenarians, expecting the imminent Second Coming of Christ. Some were social reformers like Gerrard Winstanley whose followers, agrarian communists known as Diggers, believed that the common lands should be returned to the common people. Others were mystics, like those called Ranters, who believed that they were infused with a holy spirit that removed sin from even their most reprehensible acts. The most enduring of these groups were the Quakers, whose social radicalism was seen in their refusal to take oaths or doff their hats and whose religious radicalism was contained in their emphasis upon inner light. Ultimately, all of these groups were persecuted by successive revolutionary governments, which were continually being forced to establish conservative limits to individual and collective behaviour.

The failure of the Nominated Parliament led to the creation of the first British constitution, the Instrument of Government (1653). Drafted by John Lambert, the Instrument created a lord protector, a Council of State, and a reformed Parliament that was to be elected at least once every three years. Cromwell was named protector, and he chose a civilian-dominated Council to help him govern. The Protectorate tackled many of the central issues of reform head-on. Commissions were appointed to study law reform and the question of tithes. Social legislation against swearing, drunkenness, and stage plays was introduced. Steps were taken to provide for the training of a godly ministry, and even a new university at Durham was begun.

The Instrument of Government

But the protector was no better able to manage his parliaments than had been the king. The Parliament of 1654 immediately questioned the entire basis of the newly established government, with the republicans vigorously disputing the office of lord protector. The Parliament of 1656, despite the exclusion of many known opponents, was no more pliable. Both were a focus for the manifold discontents of supporters and opponents of the regime.

Among opponents, royalists were again active, though by now they were reduced to secret associations and conspiracies. Penruddock's rising (1655) in the west was effectively suppressed, but royalist opposition led to the imposition of military rule by means of the appointment of regional major-generals. They were so extremely unpopular, however, that despite their effectiveness the offices were abolished within a year.

By now it was apparent that the regime was held to-

gether by Cromwell alone. Within his personality resided the contradictions of the revolution. Like the gentry he desired a fixed and stable constitution, but like the zealous he was infused with a millenarian vision of a more glorious world to come. As a member of Parliament from 1640 he respected the fundamental authority that Parliament represented, but as a member of the army he understood power and the decisive demands of necessity. In the 1650s many wished him to become king, but he refused the crown, preferring the authority of the people to the authority of the sword. When he died in 1658, all hope of continued reform died with him.

For a time Richard Cromwell was elevated to his father's titles and dignity. But he was no match in power or skill. The republicans and army officers who had fought Oliver tooth and nail now hoped to use his son to dismantle the civil government that under the Humble Petition and Advice (1657) had come to resemble nothing so much as the old monarchy. An upper House of Lords had been created, and the court at Whitehall was every bit as ceremonious as that of the Stuarts. While some demanded that the Rump be restored to power, others clamoured for the selection of a new Parliament on the basis of the old franchise, and this took place in 1659. By then there was a vacuum of power at the centre; Richard Cromwell, incapable of governing, simply left office. A rebellion of junior officers led to the reestablishment of the Rump.

But all was confusion. The Rump was incapable of governing without financial support from the city and military support from the army. Just as in 1647, the city demanded military disbandment and the army demanded satisfaction of its material grievances. But the army was no longer a unified force. Contentions among the senior officers led to an attempt to arrest Lambert, and the widely scattered regiments had their own grievances to propound. The most powerful force was in Scotland, commanded by George Monck, once a royalist and now one of the ablest of the army's senior officers. When one group of officers determined to dissolve the Rump, Monck marched his forces south, determined to restore it. Arriving in London, Monck quickly realized that the Rump could never govern effectively and that only the restoration of Charles II could put an end to the political chaos that now gripped the state. In February 1660 Monck reversed Pride's Purge, inviting all of the secluded members of the Long Parliament to return to their seats under army protection. A month later the Long Parliament dissolved itself, paving the way for the return of the king.

George Monck

The later Stuarts

CHARLES II (1660–85)

The Restoration. Charles II arrived in London on the 30th birthday of what had already been a remarkably eventful life. He came of age in Europe, a child of diplomatic intrigues, broken promises, and unfulfilled hopes. By necessity he had developed a thick skin and a shrewd political realism. This was displayed in the Declaration of Breda (1660), in which Charles offered something to everyone in his terms for resuming government. A general pardon would be issued, a tolerant religious settlement would be sought, and security for private property would be assured. Never a man for details, Charles left the specifics to the Convention Parliament (1660), which was composed of members of the competing religious and political parties that contended for power amid the rubble of the Commonwealth.

The Convention declared the restoration of the king and lords, disbanded the army, established a fixed income for the king by maintaining the parliamentary innovation of the excise tax, and returned to the crown and the bishops their confiscated estates. But it made no headway on a religious settlement. Despite Charles's promise of a limited toleration and his desire to accept Presbyterians into the Anglican fold, as detailed in the Worcester House Declaration (1660), enthusiasts from both left and right wrecked every compromise.

It was left to the Cavalier Parliament (1661–79) to make the hard choices and to demonstrate that one of the changes that had survived the Revolution was the independence of Parliament. Despite Charles's desire to treat his father's adversaries leniently and to find a broad church settlement, the Cavalier Parliament established a rigid Anglican orthodoxy. It began the alliance between squire and parson that was to dominate English local society for centuries. The bishops were returned to Parliament, a new prayer book was authorized, and repressive acts were passed to compel conformity. The imposition of oaths of allegiance and nonresistance to the crown and an oath recognizing the king's supremacy in the church upon all members of local government in the Corporation Act (1661) and then upon the clergy in the Act of Uniformity (1662) led to a massive purge of officeholders. Town governors were put out of their places, and nearly a fifth of all clergymen were deprived of their livings. Authority in the localities was now firmly in the hands of the gentry. The Conventicle Act (1664) barred dissenters from holding separate church services, and the Five Mile Act (1665) prohibited dispossessed ministers from even visiting their former congregations.

The Cavalier Parliament

This program of repressive religious legislation was the first of many missed opportunities to remove the underlying causes of political discontent. Though religious dissenters were not a large percentage of the population, their treatment raised the spectre of permanently divided local communities and of potentially arbitrary government. This legislation (the Clarendon Code) is inappropriately associated with the name of Lord Chancellor Clarendon, for he as well as the king realized the dangers of religious repression and attempted to soften its effects. Indeed, in central government the king relied upon men of diverse political backgrounds and religious beliefs. Clarendon, who had lived with the king in exile, was his chief political adviser, and Charles's brother James, Duke of York, was his closest confidant and entrusted with the vital post of lord admiral. Monck, who had made the restoration possible, was raised to Duke of Albemarle and continued to hold military authority over the small standing army that, for the first time in English history, the king maintained.

War and government. Charles II could not undo the effects of the Revolution, but they were not all negative. The Commonwealth had had to fight for its survival, and in the process England had become a potent military power. Wars against France and Spain had expanded English colonial dominions. Dunkirk and Jamaica were seized, Barbados was colonized, and the North American colonies flourished. Colonial trade was an important source of royal revenue, and Charles II continued Cromwell's policy of restricting trade to English ships and imposing duties on imports and exports. The Navigation Acts (1660 and 1663) were directed against the Dutch, still the most powerful commercial force in Europe. The Cromwellian Navigation Act (1651) had resulted in the first Anglo-Dutch War (1652–54), and Charles's policy had the same effect. In military terms the Dutch Wars (1665–67; 1672–74) were a standoff, but in economic terms they were an English triumph. The American colonies were consolidated by the capture of New York, and the policy of the Navigation Acts was effectively established. Colonial trade and English shipping mushroomed.

The Dutch Wars

In the long run Charles's aggressive foreign policy solved the crown's perpetual fiscal crises. But in the short run it made matters worse. The London plague (1665) and fire (1666) were interpreted as divine judgments against a sinful nation. These catastrophes were compounded when the Dutch burned a large portion of the English fleet in 1667, which led to the dismissal and exile of Clarendon. The crown's debts led to the Stop of the Exchequer (1672), by which Charles suspended payment of his bills. The king now ruled through a group of ministers known as the Cabal, an anagram of the first letters of their names. None of the five was Anglican, and two were Roman Catholic.

Charles had wearied of repressive Anglicanism, underestimating its strength among rural gentry and clergy, and desired comprehension and toleration in his church. This fit with his foreign-policy objectives, for in the Treaty of Dover (1670) he allied himself with Catholic France against Protestant Holland. In exchange he received a

large subsidy from Louis XIV, and in the treaty's secret clauses, known only to the king's Catholic ministers, the promise of an even larger one if Charles undertook, at some unspecified moment, to declare himself a Catholic. That moment came for the king on his deathbed, by which time his brother and heir, the Duke of York, had already openly professed his conversion. In 1672 Charles promulgated the Declaration of Indulgence, which suspended the penal code against all religious nonconformists, Catholic and dissenter alike. But a declaration of toleration could not bring together these mortal enemies, and the king found himself faced by a unified Protestant front. Parliamentary Anglicans would not vote money for war until the Declaration was abrogated. The passage of the Test Act (1673), which the king reluctantly signed, effectively barred all but Anglicans from holding national office and forced the Duke of York to resign the admiralty.

The Popish Plot. Anti-Catholicism united the disparate elements of English Protestantism as did nothing else. Anglicans vigorously persecuted the Protestant sects, especially Quakers and Baptists, who were imprisoned by the thousands whenever the government claimed to have discovered a radical plot. John Bunyan's *Pilgrim's Progress* (1678), which became one of the most popular works in the English language, was composed in jail. Yet dissenters held out against persecution and continued to make their converts in towns and cities. They railed against the debauchery of court life, naming the Duke of York, whose shotgun wedding had scandalized even his own family, and the king himself, who acknowledged 17 bastard children but did not produce one legitimate heir. Most of all they feared a Catholic revival, which by the late 1670s was no paranoid delusion. The alliance with Catholic France and rumours of the secret treaty, the open conversion of the Duke of York, heir to the throne, and the king's efforts to suspend the laws against Catholic officeholders were potent signs.

Not even the policy of Charles's new chief minister, Thomas Osborne, Earl of Danby, could stem the tide of suspicion. An Anglican, Danby tried to move the crown back into alliance with the majority of country gentry, who wanted the enforcement of the penal code and the end of the pro-French foreign policy. He arranged the marriage of Mary, eldest daughter of the Duke of York, to William of Orange, the Dutch stadtholder. Yet, like the king, Danby admired Louis XIV and the French style of monarchy. He attempted to manage Parliament, centralize crown patronage, shore up royal finance, and maintain a standing army—in short, to build a base for royal absolutism. Catholicism and absolutism were so firmly linked in the popular mind the Danby was soon tarred by this broad brush. In 1678 a London dissenter named Titus Oates revealed evidence of a plot by the Jesuits to murder the king and to establish Roman Catholicism in England. Although both the evidence and the plot were a total fabrication, England was quickly swept up in anti-Catholic hysteria. The murder of the Protestant magistrate who had first heard Oates's revelations lent credence to a tissue of lies. Thirty-five alleged conspirators in the Popish Plot were executed, harsh laws against Catholics were revived, and Danby's political position was undermined when it was revealed that he had been in secret negotiation with the French. Parliament voted his impeachment and began to investigate the clauses of the Anglo-French treaties. A second Test Act (1678) was passed, barring all but Anglicans from Parliament, and an exception for the Duke of York to sit in the lords was carried by only two votes. After 18 years Charles II dissolved the Cavalier Parliament.

The exclusion crisis and the Tory reaction. The mass hysteria that resulted from the Popish Plot also had its effects on the nation's governors. When Parliament assembled in 1679, a bill was introduced to exclude the Duke of York from the throne. This plunged the state into its most serious political crisis since the Revolution. But, unlike his father, Charles II reacted calmly and decisively. First he co-opted the leading exclusionists, including the Earls of Shaftesbury, Halifax, and Essex, into his government, and then he offered a plan for safeguarding the church during his brother's reign. But when the Commons passed the Exclusion Bill, Charles dissolved Parliament and called new elections. These did not change the mood of the country, for in the second Exclusion Parliament (1679) the Commons also voted to bypass the Duke of York in favour of his daughter Mary and William of Orange, though this was rejected by the Lords. Again Parliament was dissolved, again the king appealed to the country, and again an unyielding Parliament met at Oxford (1681). By now the king had shown his determination and had frightened the local elites into believing that there was danger of another civil war. The Oxford Parliament was dissolved in a week, the "Whig" councillors, as they were now called, were dismissed from their places, and the king appealed directly to the nation for support.

The king also appealed to his cousin Louis XIV, who feared exclusion as much as Charles did, if for different reasons. Louis provided a large annual subsidy to increase Charles's already plentiful revenues, which had grown with English commerce. Louis also encouraged him to strike out against the Whigs. An attempt to impeach the Earl of Shaftesbury was foiled only because a Whig grand jury refused to return an indictment. But the earl was forced into exile in Holland, where he died in 1683. The king next attacked the government of London, calling in its charter and reorganizing its institutions so that "Tories," as his supporters were now called, held power. Quo warranto proceedings against the charters of many urban corporations followed, forcing surrenders and reincorporations that gave the crown the ability to replace disloyal local governors.

In 1683 government informants named the Earl of Essex, Lord William Russell, and Algernon Sidney as conspirators in the Rye House Plot, a plan to assassinate the king. Though the evidence was flimsy, Russell and Sidney were executed and Essex took his own life. There was hardly a murmur of protest when Charles II failed to summon a Parliament in 1684 as he was bound to do by the Triennial Act. He was now fully master of his state—financially independent of Parliament and politically secure, with loyal Tory servants predominating in local and national government. He died at the height of his power in 1685.

JAMES II (1685–88)

Church and king. Unlike his brother, James II did not dissimulate for the sake of policy. He dealt plainly with friend and foe alike. James did not desire to establish Catholicism or absolutism and offered ironclad guarantees for the preservation of the Anglican church. He did desire better treatment for his coreligionists and the repeal of the Test Acts. James came to the throne amid declarations of loyalty from the ruling elite. The Parliament of 1685 was decidedly royalist, granting the king customs revenues for life as well as emergency military aid to suppress Monmouth's Rebellion (1685). The Duke of Monmouth, an illegitimate son of Charles II, was Shaftesbury's personal choice for the throne, had Exclusion succeeded. He recruited tradesmen and farmers as he marched through the west country on the way to defeat at Sedgemoor. The rebellion was a fiasco as the local gentry refused to sanction civil war. Monmouth was executed and more than 600 of his supporters were either hanged or deported in the brutal aftermath of the rebellion, the Bloody Assizes (1685).

Monmouth's Rebellion

The king misinterpreted Monmouth's failure to mean that the nation would place the legitimate succession above all else. During the rebellion James had dispensed with the Test Act and appointed Catholics to military command. This led to a confrontation with Parliament, but the king's dispensing power was upheld in *Godden* v. *Hales* (1686). James made it clear that he intended to maintain his large military establishment, to promote Catholics to positions of leadership, and to dispense with the penal code.

These decisions could hardly have come at a worse moment. In France, Louis XIV revoked the Edict of Nantes, the legislation that had protected the rights of French Protestants for nearly a century. The repression of Huguenot congregations inflamed English public opinion. Thus the king's effort on behalf of Catholics was doomed from the start. He had vainly hoped the Parliament of 1685 would repeal the Test Acts. When his attempt to

open the universities to Catholics was met by rigid opposition, he forced a Catholic head upon Magdelan College, Oxford, but only after an open break with the fellows and unpleasant publicity. Moreover, his effort to forge an alliance with dissenters proved unsuccessful. When James showed favour to William Penn and the Quakers, his leading Anglican ministers, the Earls of Clarendon and Rochester, resigned.

By now the king was set upon a collision course with his natural supporters. The Tory interest was made up of solid support for church and king; it was James's mistake to believe that they would support one without the other. In 1687 he reissued the Declaration of Indulgence, which suspended the penal laws against Catholics and dissenters. This was a temporary measure, for James hoped that his next Parliament would repeal the penal code in its entirety. To that end he began a systematic investigation of the parliamentary boroughs. Agents were sent to question mayors, lieutenants, and justices of the peace about their loyalty to the regime and their willingness to vote for MPs who would repeal the Test Acts. Most gave temporizing answers, but those who stood out were purged from their places. For the first time in English history the crown was undertaking to pack Parliament.

The Revolution of 1688. The final crisis of James's reign resulted from two related events. The first was the refusal of seven bishops to read the Declaration of Indulgence in their churches. The king was so infuriated by this unexpected check to his plans that he had the bishops imprisoned, charged with seditious libel, and tried. Meanwhile, in June 1688 Queen Mary gave birth to a male heir, raising the possibility that there would be a Catholic successor to James. When the bishops were triumphantly acquitted by a London jury, leaders of all political groups within the state were persuaded that the time had come to take action. Seven leading Protestants drafted a carefully worded invitation for William of Orange to come to England and investigate the circumstances of the birth of the king's heir. In effect, the leaders of the political nation had invited a foreign prince to invade their land.

Invitation to William of Orange

This came as no surprise to William, who had been contemplating an invasion since the spring of 1688. William, who was organizing a Grand Alliance against Louis XIV, needed England as an ally rather than a rival. All Europe was readying for war in the summer of 1688, and James had powerful land and sea forces at his disposal to repel William's invasion. The crossing, begun on October 19, was a feat of military genius, however propitious the strong eastern "Protestant wind" that kept the English fleet at anchor while Dutch ships landed at Torbay (November 5). William took Exeter and issued a declaration calling for the election of a free Parliament. From the beginning the Anglican interest flocked to him. James could only watch as parts of his army melted away.

Yet there was no plan to depose the king. Many Tories hoped that William's presence would force James to change his policies; many Whigs believed that a free Parliament could fetter his excesses. When James marched out of London, there was even the prospect of battle. But the result was completely unforeseen. James lost his nerve, sent his family to France, and followed after them, tossing the Great Seal into the Thames. James's flight was a godsend, and when he was captured en route William allowed him to escape again. At the end of December William arrived in London, summoned the leading peers and bishops to help him keep order, and called Parliament into being.

The Convention Parliament (1689) met amid the confusion created by James's flight. For some Tories, James II was still the king; some were willing to contemplate a regency; others to allow Mary to rule with William as consort. But neither William nor the Whigs would accept such a solution. William was to be king in his own right, and in February the Convention agreed that James had "abdicated the government and that the throne has thereby become vacant." At the same time the leaders of the Convention prepared the Declaration of Rights to be presented to William and Mary. The declaration was a restatement of traditional rights, but the conflicts between Whigs and Tories caused it to be watered down considerably. Nevertheless, the Whigs did manage to declare the suspending power and the maintenance of a standing army in peacetime illegal. But many of the other clauses protecting free speech, free elections, and frequent parliaments were cast in anodyne formulas and the offer of the throne was not conditional upon the acceptance of the Declaration of Rights.

WILLIAM III (1689–1702) AND MARY II (1689–94)

The Revolution settlement. The Revolution of 1688 was a constitutional crisis, which was resolved through legislation. The Bill of Rights (1689), a more conservative document than even the declaration, was passed into law, but it established the principle that only a Protestant could wear the crown of England. A new coronation oath required the monarch to uphold Protestantism and the statutes, laws, and customs of the realm as well. A Triennial Act (1694) reestablished the principle of regular parliamentary sessions.

The Bill of Rights

Two other pieces of legislation tackled problems that had vexed the nation since 1640. The Mutiny Act (1689) restrained the monarch's control over military forces in England by restricting the use of martial law. It was passed for one year only, though when it lapsed between 1698 and 1701, the crown's military power was not appreciably affected. The Toleration Act (1689) was the most disappointing part of the whole settlement. It was originally intended to be part of a new comprehensive religious settlement in which most mainline dissenters would be admitted into the church. This failed for the same reasons that comprehension had failed for 30 years; the Anglican clergy would not give up its monopoly, and dissenters would not compromise their principles. The Toleration Act permitted most forms of Protestant worship; Unitarians were explicitly excluded, as were Catholics and Jews. But the Test Acts that theoretically prevented dissenters from holding government office or sitting in Parliament were continued in force.

The sinews of war. William III had come to England to further his continental designs, but English politics conspired against him. The first years of his reign were dominated by the constitutional issues of the revolution settlement, and he became increasingly frustrated with the political squabbling of Whigs and Tories. Moreover, holding the English throne was proving more difficult than taking it. In 1690, with French backing, James II invaded Ireland and routed Protestant forces. William personally led an army to the Battle of Boyne (1690) where James's forces were crushed. But the compromise settlement that he wished to impose on Ireland, which would have made it secure from French intrigue, was rejected by Parliament, which executed the full rigours of the penal code upon Irish Catholics.

The Irish wars impressed upon William's English subjects that, as long as the French backed James, they were now part of the great European struggle. Parliament granted William vast subsidies for the Nine Years' War (1688–97), more than £4.5 million in a two-year period alone, but also established a right to oversee the expenditure of public monies. This led to both economies and accountability, and it forged a new political alliance among "country" forces that were uneasy about foreign entanglements and suspicious of corruption at court. William's war was going badly on land and sea. The French fleet inflicted heavy losses on a combined Anglo-Dutch force and heavier losses on English merchant shipping. The land war was a desultory series of sieges and reliefs, which again tipped in favour of France.

Year by year the financial costs mounted. Between 1688 and 1702 England accumulated more than £14 million of debt, which was financed through the creation of the Bank of England (1694). The bank was a joint-stock company empowered to discount bills and issue notes. It lent to the government at a fixed rate of interest—initially 8 percent—and this interest was secured by a specific customs grant. Investors scrambled for the bank's notes, which were considered gilt-edged securities, and more than £1.2 million was raised on the initial offering. Not surprisingly,

The Bank of England

a growing funded debt created inflation and led to a financial crisis in 1696. But the underlying English economy was sound, and military expenditures fueled production.

The establishment of a funded national debt and the Bank of England was the work of the Whigs in alliance with the London mercantile establishment. The Tories and the country party were alternately suspicious and jealous of Whig success. In order to secure funds for his campaigns, William had been forced to allow the Whigs to dominate government, much against his inclination. An attempted assassination of the king in 1696 gave the Whigs an opportunity to impose an oath on the political nation that William was the "rightful and lawful king." This directly challenged Tory consciences, which had been tender since the death of Queen Mary in 1694. Many resigned office rather than affirm what they did not believe. The ascendancy of the so-called Junto Whigs might have been secured had not European events once again intruded into English affairs. In 1697 the Nine Years' War ended with the Treaty of Rijswijk, in which Louis XIV formally recognized William III as king of England.

Treaty of Rijswijk

A great revulsion and war weariness now took hold of the nation. Parliament voted to disband most of the military establishment, including William's own Dutch guards, and a vigorous public debate against the existence of a standing army ensued. Taxes were slashed, accounts were audited, and irregularities exposed. The Junto Whigs, who were associated with war and war profiteers, fell. A new coalition of country and Tory MPs, led by Sir Robert Harley, launched a vigorous campaign of retrenchment. It had not progressed very far by 1700 when the deaths of the Duke of Gloucester and Charles II of Spain redefined English and European priorities.

The Duke of Gloucester was the only surviving child of the queen's sister, Princess Anne, despite her 18 pregnancies. Because William and Mary were childless, the duke was the long-term Protestant heir to the throne. His death created a complicated problem that was resolved in the Act of Settlement (1701), which devolved the throne upon the grandchildren of James I, that is, on Sophia of Hanover and her son George. In clauses that read like a criticism of the policies of William III the act stipulated that the sovereign must be a member of the Anglican church and that his foreign policy was to be directed by Parliament and his domestic policy by the privy council. It also limited the right of the king to dismiss judges at pleasure. Although many of the more restrictive clauses of the act were repealed in 1706, the Act of Settlement asserted a greater degree of parliamentary control over the monarchy than had been obtained since 1649.

The consequences of the death of Charles II of Spain were no less momentous. Years of futile negotiations to divide the vast Spanish empire among several claimants came to an end when Louis XIV placed his grandson on the Spanish throne and began making preparations to unite the kingdoms into a grand Bourbon alliance. Louis's aggressive stance overcame even the torpor of British public opinion, especially when he renounced William's legitimacy and welcomed James III to his court as rightful king of England. William constructed another anti-French coalition and bequeathed to Queen Anne the War of the Spanish Succession (1701–14).

ANNE (1702–14)

Queen Anne, daughter of James II and the last of the Stuarts, inherited a nation that was bitterly divided politically. Her weak eyesight and indifferent health forced her to rely more upon her ministers than any of her predecessors, but she was no less effective for that. Anne had decided views about people and policies, and these did much to shape her reign. She detested the party divisions that now dominated central politics and did all she could to avoid being controlled by either Whigs or Tories. While she only briefly achieved her ideal of a nonpartisan ministry, Anne did much to disappoint the ambitions of nearly all party leaders.

Whigs and Tories. The most significant development in political life over the last quarter century had been the growth of clearly defined and opposing parties, which had taken the opprobrious titles Whigs (Scottish horse thieves) and Tories (Irish cattle rustlers). Parties had first formed during the exclusion crisis of 1679–81, but it was the Triennial Act (1694) that unintentionally gave life to party conflict. Nine general elections were held between 1695 and 1713, and these provided the structure whereby party issues and party leaders were pushed to the fore. Though party discipline was still in its infancy and ideology was a novel aspect of politics, clearly recognizable political parties had emerged by the end of the reign of William III. In general, the Tories stood for the Anglican church, the land, and the principle of passive obedience. They remained divided over the impending Hanoverian succession, wistfully dreaming that James III might convert to Protestantism so that the sanctity of the legitimate succession could be reaffirmed. From their country houses the Tories opposed an expensive land war and favoured the "blue sea" strategy of dominating the Atlantic and Mediterranean shipping lanes. Their leaders had a self-destructive streak. Only Robert Harley, Earl of Oxford, was a politician of the first rank, and he always shrank from being labeled a Tory. The Tories generally had a majority in the Commons and a friend on the throne, but they rarely attained power.

The Whigs were blessed with brilliant leadership and an inexhaustible supply of good luck. John Churchill, Duke of Marlborough, was the outstanding military figure of his day. His victories at Blenheim (1704) and Ramillies (1706) rank among the greatest in British history. During the first part of the reign his wife, Sarah, was the queen's confidante, and together they were able to push Anne to support an aggressive and expensive foreign policy. Continental warfare was costing £4 million a year, paid for by a tax on land, and after the early years successes were few and far between. Sidney Godolphin kept Marlborough supplied and financed and ably managed the Whig interest by disciplining government officeholders to vote for Whig policies in Parliament. Among these policies were support for dissenters who, to avoid the rigours of the Test Acts, would take Anglican communion. Both the queen and the Tories were opposed to these occasional conformists, and three bills to outlaw the practice were passed through the Commons but defeated in the Lords. When the Tories attempted to attach one of these to the military appropriations bill, even the queen condemned the maneuver.

John Churchill, Duke of Marlborough

For the first half of Anne's reign Whig policies were dominant. Marlborough's victories set off a wave of nationalistic pride and forced even Tories to concede the wisdom of a land war. Unfortunately, military success built overconfidence, prompting the Whigs to adopt the fruitless policy of "no peace without Spain," which committed them to an increasingly unattainable conquest of Iberia. Yet the capture of both Gibraltar (1704) and Minorca (1708) made England the dominant sea power in the western Mediterranean and paid handsome commercial dividends. So too did the unexpected union with Scotland (1707). Here again, Godolphin was the dominant figure, calling the Scottish bluff when they announced they would not accept the Hanoverian succession. Godolphin passed an Aliens Act (1705) that would have prohibited all trade between England and Scotland—no mere scare tactic in light of the commercial policy that was crippling the Irish economy. Rather than risk economic strangulation, Scottish leaders negotiated for a permanent union, a compact the English monarchy had sought for more than a century. The union was a well-balanced bargain: free trade was established, Scottish Presbyterianism was protected, and provisions were made to include 45 Scottish members in the English House of Commons and 18 members in the House of Lords. England gained security on its northern border, and the Whigs gained the promise of a peaceful Hanoverian succession.

Union with Scotland

Tories and Jacobites. Whig successes were not welcomed by the queen, who had a personal aversion to most of their leaders, especially after her estrangement from Sarah Churchill. As in the reign of William, war weariness and tax resistance combined to bring them down. Robert Harley, Earl of Oxford, and Henry St. John, Viscount Bolingbroke, vied for leadership of a reinvigorated Tory

party that rallied support with the cry "church in danger." In 1710 a Whig prosecution of a bigoted Anglican minister, Henry Sacheverell, badly backfired. Orchestrated mob violence was directed against dissenting churches, and Sacheverell was impeached by only a narrow margin and given a light punishment. When the Tories gained power, they were able to pass legislation directed against dissenters, including the Occasional Conformity Act (1711), which forbade dissenters to circumvent the Test Acts by occasionally taking Anglican communion. The Tories also concluded the War of the Spanish Succession. By the Treaty of Utrecht (1713) England expanded its colonial empire in Canada and the Caribbean and maintained possession of Gibraltar and Minorca in the Mediterranean.

Occasional Conformity Act

But the Tories had their own Achilles' heel. They were deeply divided over who should succeed Anne, divisions that became public during the queen's serious illness in 1713. Though there were far more Hanoverian Tories than Jacobite Tories (supporters of James II and his son, James III), the prospect of the succession of a German, Lutheran prince with continental possessions to defend did not warm the hearts of isolationist, Anglican country gentlemen. Both Harley and Bolingbroke were in correspondence with James III, but Harley made it plain that he would only support a Protestant succession. Bolingbroke's position was more complicated. A brilliant politician, he realized that the Tories would have little to hope for from the Hanoverians and that they could only survive by creating huge majorities in Parliament and an unshakable alliance with the church. Conflict between Tory leaders and divisions within the rank and file combined to defeat Bolingbroke's plans. After Anne died in August 1714, George I acceded to the British throne and Bolingbroke, having tainted the Tory party with Jacobitism for the next half century, fled to France. (M.A.K.)

18th-century Britain, 1714–1815

THE STATE OF BRITAIN IN 1714

When Georg Ludwig, elector of Hanover, became king of Great Britain on Aug. 1, 1714, the country was in some respects bitterly divided. Fundamentally, however, it was prosperous, cohesive, and already a leading European and imperial power. Abroad, Britain's involvement in the War of the Spanish Succession had been brought to a satisfactory conclusion by the Treaty of Utrecht (1713). It had acquired new colonies in Gibraltar, Minorca, Nova Scotia, Newfoundland, and Hudson's Bay, as well as trading concessions in the Spanish New World. By contrast, Britain's rivals, France, Spain, and the Dutch Republic, were left weakened or war-weary by the conflict. It took France a decade to recover, and Spain and Holland were unable to reverse their military and economic decline. As a result Britain was able to remain aloof from war on the Continent for a quarter of a century after the Hanoverian succession, and this protracted peace was to be crucial to the new dynasty's survival and success.

War had also strengthened the British state at home. The need to raise men and money had increased the size and scope of the executive as well as the power and prestige of the House of Commons. Taxation had accounted for 70 percent of Britain's wartime expenditure (£93,644,560 between 1702 and 1713), so the Commons' control over taxation became a powerful guarantee of its continuing importance.

Britain's ability to pay for war on this scale demonstrated the extent of its wealth. Agriculture was still the bedrock of the economy, but trade was increasing, and more men and women were employed in industry in Britain than in any other European nation. Wealth, however, was unequally distributed, with almost a third of the national income belonging to only 5 percent of the population. But British society was not polarized simply between the rich and the poor; according to writer Daniel Defoe there were seven different and more subtle categories:

1. The great, who live profusely.
2. The rich, who live plentifully.
3. The middle sort, who live well.
4. The working trades, who labour hard, but feel no want.
5. The country people, farmers etc., who fare indifferently.
6. The poor, who fare hard.
7. The miserable, that really pinch and suffer want.

From 1700 to the 1740s Britain's population remained stable at about seven million, and agricultural production increased. So, although men and women from Defoe's 6th and 7th categories could still die of hunger and hunger-related diseases, in most regions of Britain there was usually enough basic food to go around. This was crucial to social stability and to popular acquiescence in the new Hanoverian regime.

But early 18th-century Britain also had its weaknesses. Its Celtic fringe—Wales, Ireland, and Scotland—had been barely assimilated. The vast majority of Welsh men and women could neither speak nor understand the English language. Most Irish men and women spoke Gaelic and belonged to the Roman Catholic church, in contrast with the population of the British mainland, which was staunchly Protestant. Scotland, which had only been united to England and Wales in 1707, still retained its traditional educational, religious, legal, and cultural practices. These internal divisions were made more dangerous by the existence of rival claimants to the British throne. James II, who had been expelled in the Glorious Revolution of 1688, died 13 years later, but his son, James Francis Edward Stuart, the Old Pretender, pressed his family's claims from his exile in France. His Catholicism and Scottish ancestry ensured him wide support in Ireland and the Scottish Highlands; his cause also commanded sympathy among sections of the Welsh and English gentry and, arguably, among the masses.

Controversy over the succession sharpened partisan infighting between the Whig and Tory parties. About 50 Tory MPs (less than a seventh of the total number) may have been covert Jacobites in 1714. More generally, Tories differed from Whigs over religious issues and foreign policy. They were more anxious to preserve the privileges of the Anglican church and more hostile to military involvement in continental Europe than Whig politicians were inclined to be. These attitudes made the Tories vulnerable in 1714. The new king was a Lutheran by upbringing and wanted to establish wider religious toleration in his new kingdom. As a German he was deeply interested in European affairs. Consequently he regarded the Tory party as insular in its outlook as well as suspect in its allegiance.

BRITAIN FROM 1715 TO 1742

The supremacy of the Whigs. Even before he arrived in Britain, George I had decided to exclude the two leading Tory ministers, Robert Harley, Earl of Oxford, and Henry St. John, Viscount Bolingbroke. In their place he appointed two Whig politicians, Charles, Viscount Townshend, and James, Viscount Stanhope, as secretaries of state. Townshend's brother-in-law, Robert Walpole, became paymaster general. Walpole, who came from a minor Norfolk gentry family, was an extremely able politician, shrewd, greedy, and undeviatingly Whig. He encouraged the new king's partisan bias, turning it unremittingly to his advantage. A general election was held in February 1715, and, due in part to royal influence, the Whigs won 341 seats to the Tories' 217. In December the Old Pretender landed in Scotland, provoking an armed rebellion that was quickly suppressed. The proved involvement of a small number of Tory landowners led to Tories being purged not only from state office but also from the higher ranks of the army and navy, the diplomatic service, and the judicial system. To make their capture of the state even more secure, the Whigs passed the Septennial Act in 1716. It allowed general elections to occur at seven-year intervals instead of every three years, as mandated by the Triennial Act of 1694. The intention was to tame the electorate, which during Anne's reign had shown itself to be volatile and far more inclined to vote Tory than Whig.

Robert Walpole

Having defeated their Tory opponents, the Whig leaders began to quarrel among themselves. In 1717 Walpole and Townshend left office and went into open opposition. Stanhope stayed on, with Charles Spencer, Earl of Sunderland, now serving as secretary of state. At the same time the heir apparent to the throne, George, Prince of Wales,

quarreled with his father and began to flirt with Opposition groups in Parliament. These events set the pattern for future political conflicts. From then on until the 1750s the Opposition in Parliament would be a hybrid group of Whig and Tory sympathizers. And from then on until the early 19th century Oppositions in Parliament would enjoy sporadic support from successive princes of Wales. In 1717 the rebel Whigs were a serious threat in large part because Walpole was such a skillful House-of-Commons politician. As peers, Sunderland and Stanhope were confined to the House of Lords and lacked spokesmen in the Commons who could match Walpole's ruthlessness and talent. He showed his power by mobilizing a majority of MPs against the Peerage Bill in 1719. Had this legislation passed, it would have limited the king's prerogative to create new peers, thereby cementing the Whig administration's majority in the House of Lords. To prevent further blows of this kind, the Whig elite ended its schism in April 1720. The royal family temporarily buried its differences at the same time.

South Sea Bubble

The restoration of unity was just as well, as 1720 saw the bursting of what became known as the South Sea Bubble. The South Sea Company had been founded in 1711 as a trading and finance company. In 1719 its directors offered to take over a large portion of the national debt previously managed by the Bank of England. The Whig administration supported this takeover, and in return the company made gifts (in effect, bribes) of its new stock to influential Whig politicians, including Stanhope and Sunderland, and to the king's mistress, the Duchess of Kendal. In 1720 investing in the South Sea Company became a mania among those who could afford it and some who could not; South Sea stock was at 120 in January and rose to 1,000 by August. But in September the inevitable crash came. Many landed and mercantile families were ruined, and there was a nationwide shortage of specie. Parliament demanded an inquiry, thus raising the possibility that members of the government and the royal family would be openly implicated in financial scandal. This disaster proved to be Walpole's opportunity, and he did not waste it. He used his influence in the Commons to blunt the parliamentary inquiry and managed gradually to restore financial confidence. The strain of the investigation killed Stanhope, and Sunderland too died in 1722. Walpole duly became first lord of the treasury and chancellor of exchequer, while Townshend returned to his post as secretary of state.

The Atterbury plot

Walpole's position as the king's favourite minister was finally assured when he exposed the Atterbury plot. Francis Atterbury was bishop of Rochester. Always a Tory and High Churchman, he drifted after the Hanoverian succession into Jacobite intrigue. In 1721–22 he and a small group of conspirators plotted an armed invasion of Britain on behalf of the Old Pretender. The plot was uncovered by the secret service, which was more efficient in this period than it was until World War II. Atterbury was tried for treason by Parliament and sent into exile. This coup, one politician aptly wrote at the time, was the "most fortunate and greatest circumstance of Mr Walpole's life. It fixed him with the King, and united for a time the whole body of Whigs to him, and gave him the universal credit of an able and vigilant Minister."

Robert Walpole. Walpole has often been referred to as Britain's first prime minister, but historically this is incorrect. The title had in fact been applied to certain ministers in Anne's reign and was commonly used as a slur or simply as a synonym for first minister. During Walpole's period of dominance it was certainly used more frequently, but it did not become an official title until the early 20th century. Some historians have also claimed that Walpole was the architect of political stability in Britain, but this interpretation needs to be qualified. There is no doubt that from 1722 to his resignation in 1742 Walpole stabilized political power in himself and a section of the Whig party. Nor can there be any doubt that his foreign and economic policies helped the Hanoverian dynasty to become securely entrenched in Britain. But it should not be forgotten that Walpole inherited a nation that was already wealthy and at peace. He built on foundations that were already very strong. And, although he was to dominate political life for 20 years, he never succeeded in stamping out political, religious, and cultural opposition entirely, nor did he expect to do so.

Opposition to Walpole in Parliament began to develop as early as 1725. When William Pulteney, an ambitious and talented politician, was dismissed from state office, he and 17 other Whig MPs aligned themselves with the 150 Tory MPs remaining in the House of Commons. These dissidents (who called themselves Patriot Whigs) grew in number until, by the mid-1730s, more than 100 Whig MPs were collaborating with the Tories against Walpole's nominally Whig administration. Some were motivated primarily by disappointed ambition. But many Whigs and Tories genuinely believed that Walpole had arrogated too much power to himself and that he was corrupt and an enemy to liberty. These accusations were expressed not just among politicians in London but also in the growing number of newspapers and periodicals in Britain at large. In 1726 Pulteney and the one-time Tory minister Lord Bolingbroke founded their own journal, *The Craftsman* (the implication of the title being that Walpole governed by craft alone). It was widely read among the political classes, not least because many of the most gifted writers working in London had been drawn into the Opposition camp. Jonathan Swift, Alexander Pope, and, for a time, Henry Fielding all wrote against Walpole. So did John Gay, whose triumphantly successful *The Beggar's Opera* (1728) was a satire on ministerial corruption.

The Craftsman

But, despite its flamboyance and innovative tactics, the Opposition for a long time lacked high-level support. Frequent disagreements occurred between its Patriot Whig and Tory sectors. These weaknesses helped Walpole to keep the Opposition at bay until 1742. But there were other reasons for his prolonged stay in power: he retained the support of the crown, resisted military involvement in Europe, pursued a moderate religious policy, and adopted a skillful economic policy. Moreover, in the general elections of 1727 and 1734 he was able to manipulate the electoral system to maintain himself in power.

George II and Walpole. George I died in June 1727 and was buried in Hanover. He was succeeded by his eldest son, who became George II. Initially the new king planned to dismiss Walpole and appoint his personal favourite, Spencer Compton, in his place. Closer familiarity with Walpole's gifts, however, dissuaded him from taking this step, as did his formidable wife, Queen Caroline, who remained an important ally of the minister until her death in 1737. Walpole cemented his advantage by securing the king a Civil List (money allowance) from Parliament of £800,000, a considerably larger sum than previous monarchs had been able to enjoy. Royal favour, in turn, shored up Walpole's parliamentary majority. Because the monarch appointed and promoted peers, he had massive influence in the House of Lords. In addition, he appointed the 26 bishops of the Church of England, who also possessed seats in the House of Lords. He alone could promote men to high office in the army, navy, diplomatic service, and bureaucracy. Consequently, MPs who held such offices (the so-called placemen), and those who wanted to hold them in the future, were likely to support Walpole as the king's minister out of self-interest, if for no other reason. Walpole, however, could never take royal support for granted. George II was an irritable but by no means an insignificant figure who retained great influence in terms of patronage, military affairs, and foreign policy. He demanded respect from his minister and had to be carefully managed.

Foreign policy. Once the Hanoverian succession had taken place, Whig ministers became as eager to remain at peace with France as the Tories had been. Walpole certainly adhered to this view, and for good reasons. Although Britain now possessed the world's most powerful navy, it could not match France in land forces. War with France, moreover, was likely to lead to an invasion of Hanover, which was naturally unwelcome to George I and his successor. It would also give the Old Pretender the prospect of French military aid to launch an invasion against Britain itself. In 1717 Stanhope negotiated a Triple Alliance with

the French and the Dutch. This treaty was maintained by Walpole and Townshend throughout the 1720s. By 1730, however, it was attracting considerable criticism from the Opposition, and in the Second Treaty of Vienna, signed in March 1731, Walpole jettisoned the Anglo-French alliance in favour of an alliance with Austria. But whether forming an alliance with the French or the Austrians, Walpole always considered it his primary aim to keep Britain out of war in continental Europe. In 1733 Austria, Saxony, and Russia went to war against France, Spain, and Sardinia in the War of the Polish Succession (1733–38). The Austrians asked for British aid under the terms of the Treaty of Vienna, but Walpole refused to give it. By keeping out of European entanglements for so long, Walpole appeased some of the traditionally insular Tory MPs. He also kept direct taxation low, which pleased many landed families. The land tax was cut to two shillings in the pound (10 percent) in 1730 and to one shilling in the pound two years later.

Religious policy. Walpole's religious policy was also designed to foster social and political quiescence. Traditionally the Whig party had supported wider concessions to the Protestant dissenters (Protestants who believed in the doctrine of the Trinity but who refused to join in the worship of the state church, the Church of England). They had been given freedom of worship under the Toleration Act of 1689 but were barred from full civil rights and access to university education in England. In 1719 the Whigs had repealed two pieces of Tory legislation aimed against dissent, the Schism and the Occasional Conformity acts. These concessions ensured that Protestant dissenters would be able to establish their own educational academies and hold public office in the localities, if not in the state.

Concessions to dissenters

There was always a danger, however, that too many concessions to Protestant dissent would alienate the Church of England, which enjoyed wide support in England and Wales. There were 5,000 parishes in these two countries, each containing at least one church served by a vicar (minister) or a curate (his deputy). For much of the 18th century these Anglican churches provided the only large, covered meeting places available outside of towns. They served as sources of spiritual comfort and also as centres for village social life. At religious services vicars would not only preach the word of God but also explain to congregations important national developments: wars, victories, and royal deaths and births. Thus churches often supplied the poor, the illiterate, and particularly women with the only political information available to them. Weakening the Church of England therefore struck Walpole as unwise, for at least two reasons. Its ministers provided a vital service to the state by communicating political instruction to the people. The church, moreover, commanded massive popular loyalty, and assaults on its position would arouse nationwide discontent. Walpole therefore determined to reach an accommodation with the church, and in 1723 he came to an agreement with Edmund Gibson, Bishop of London. Gibson was to ensure that only clergymen sympathetic to the Whig administration were appointed to influential positions in the Church of England. In return, Walpole undertook that no further extensive concessions would be made to Protestant dissenters. This arrangement continued until 1736.

Economic policies. Finally, Walpole's long tenure of power was assisted by national prosperity. The gross national product rose from £57.5 million in 1720 to £64.1 million in 1740, an increase of 11.5 percent. Walpole encouraged trade by abolishing some customs duties, but his main economic concerns were to reduce interest payments on the national debt and to foster agriculture by switching taxation from land to consumption. He succeeded in reducing interest payments on the debt by 26 percent during his time in office, but his efforts to reduce the land tax in favour of more excises almost led to political disaster. In 1732 he revived a duty on salt, which enabled him to cut the land tax to one shilling in the pound. In 1733 he proposed to levy excise taxes on the sale of wine and tobacco, but the Opposition in Parliament launched a ferocious and successful campaign against these proposals. It claimed that excises weighed unfairly on the poor, whereas the land tax was mainly paid by the prosperous. It claimed, too, that excise collectors, and there were more than 6,000 of them employed by the state by this time, intruded into citizens' private affairs and were a danger to British liberties. This crisis led to nationwide riots and demonstrations, and Walpole's House-of-Commons majority seemed in jeopardy. In April 1733 he decided to retreat. He continued, however, until 1740 to keep the land tax at a low rate, thereby winning important support from the nation's dominant landed class.

Land tax and excise tax

The electoral system. The fiasco over the excise might have toppled Walpole, since a general election was scheduled for 1734. In fact, however, his administration retained a comfortable majority in the House of Commons. One reason for this was that Britain's electoral system at this time did not adequately reflect the state of public opinion. Until the Reform Act of 1832 England returned 489 MPs. Eighty of these were elected by the 40 county constituencies; 196 smaller constituencies called boroughs returned two MPs each, and two other boroughs, including London, the capital city, returned four MPs each. Oxford and Cambridge universities were also allowed four representatives in Parliament. Wales returned only 24 members of Parliament and Scotland 45. Their limited representation indicated the extent to which these countries were subordinated to England in the British political system at this time.

The system was not even remotely democratic. Power in this society was intimately and inextricably connected with the possession of property, particularly landed property. To be eligible for election as an MP, a man had to possess land worth £600 per annum if he was representing a county constituency and worth £300 per annum in the case of a borough constituency. To vote, adult males had to possess some kind of residential property or, in certain borough constituencies, be registered as freemen. Women were not given the vote until 1918.

In all, some 350,000 Britons may have been able to vote in the 1720s, which was roughly one in four of the adult male population. There was no secret ballot, and voting took place in public. Consequently, many voters were liable to be influenced or coerced by their landlords or employers or bribed by the candidates themselves. Bribery was particularly widespread and effective in the smaller boroughs where there were often fewer than 100 voters and sometimes fewer than 50. These constituencies were called rotten or pocket boroughs. In the borough of Malmesbury, for example, in the English county of Wiltshire, there were only 13 voters, few of whom voted strictly in accordance with their own conscience or opinions: "It was no odds to them who they voted for," one inhabitant declared, "it was as master pleased." Large electorates could be found, however, in some areas. The northern English county constituency of Yorkshire had 15,000 voters in 1741. In Bristol, a major port on the western coast of England, 5,000 men had the vote—approximately one-third of the city's adult male population. In these larger constituencies public opinion could make itself felt at election time. The problem for the Opposition in 1734 was that there were few such populous, open constituencies but very many rotten borough seats such as Malmesbury. Since government candidates usually had more to bribe voters with in the way of money and favours, Walpole was able to win the majority of these boroughs and therefore retain his majority in the House of Commons despite his unpopularity after the excise crisis.

Rotten boroughs

Walpole's loss of power. Walpole's luck and political grasp only began to fail in 1737. In that year Queen Caroline, one of his most important allies, died. At this time, too, Frederick Louis, Prince of Wales, George II's eldest son and heir apparent, followed Hanoverian family tradition; he quarreled with his father and aligned himself with the Opposition. This damaged Walpole's position in two ways. The king, born in 1683, was now in his 50s, which was elderly by the standards of the time. Many young ambitious MPs, such as William Pitt, were inclined to join Prince Frederick, because they saw in him the political future. Moreover, as Prince of Wales, Frederick owned a large part of the county of Cornwall and conse-

quently controlled numerous rotten boroughs. In the 1734 election the Cornish constituencies had returned 32 pro-government MPs to Parliament; but at the next general election in 1741, when Prince Frederick used his electoral influence against Walpole, only 17 pro-government candidates were returned by this county. Walpole lost another important ally to the opposition, John, Duke of Argyll. Argyll was a member of the Cabinet, the most important Whig landowner in Scotland, and head of Clan Campbell. In the 1734 election his influence in Scotland helped to ensure that 34 of the country's 45 elected MPs were pro-government. But by the 1741 election he had defected to the Opposition, and the electoral repercussions were serious. On this occasion Scottish constituencies only elected 17 pro-government MPs.

But Walpole's main enemies were time and war. By 1737 he was in his 60s and had dominated politics for 15 years. Some ambitious Whigs resented his prolonged monopoly on power; others anticipated his retirement or death and judged it prudent to distance themselves from his administration. And some of Walpole's policies were now widely viewed as dubious, even anachronistic. Whereas he wanted to keep Britain out of war, many government and Opposition MPs, and even some members of Walpole's own Cabinet, favoured going to war with Spain to gain colonial and commercial objectives. Such a war policy was strongly backed by commercial opinion in London and in the nation's main trading cities.

The War of Jenkins' Ear

It was a sign of Walpole's declining powers that he was unable to prevent the drift into war in 1739. The War of Jenkins' Ear (so called after an alleged Spanish atrocity against a British merchant navy officer, Captain Robert Jenkins) was initially successful. Admiral Edward Vernon became a popular and Opposition hero when he captured the Spanish settlement of Portobelo (in what is now Panama) in November 1739. But his victory was followed by several defeats, and Britain soon became embroiled in a wider European conflict, the War of the Austrian Succession. Walpole survived the general election of 1741, but with a greatly reduced majority. His political doom was sealed in the fall of that year when the Tory and Whig sectors of the Opposition managed finally to agree on a strategy to defeat him. Walpole eventually resigned from his offices in early 1742. He still retained the king's favour, and, although sections of the Opposition wanted to impeach him for corruption, he was given a peerage, entered the House of Lords as Earl of Orford, and died in his bed in 1745. Nonetheless, the fact that he had to resign despite George II's continuing support indicated an important development in the British political system. Although monarchs retained the rights to choose their own ministers, they could no longer retain a chief minister who was unable to command a majority of votes in the House of Commons. If they wanted to remain in office, chief ministers now needed to possess parliamentary as well as royal support.

BRITAIN FROM 1742 TO 1754

Political events after Walpole's resignation demonstrated once again the artificiality and inner tensions of the Opposition. Its Tory sector (some 140 MPs strong) had expected that a new administration would be formed in which some of their leaders would be given state office. They hoped that the proscription of their party, implemented after 1714, would be reversed and that various changes in domestic and foreign policy would be made. But now that Walpole was out of the picture many of their Patriot Whig allies wanted nothing more to do with Tories or Tory measures. The leading Patriot Whig, William Pulteney, accepted a peerage and became Earl of Bath. Six other Patriot Whigs accepted government office, including John, Baron Carteret (later Earl of Granville), who became the new secretary of state. Spencer Compton, now Earl of Wilmington, became the new first lord of the treasury and nominal head of the government. Fourteen former members of Walpole's administration retained their posts, including Henry Pelham and his older brother, Thomas Pelham-Holles, Duke of Newcastle. The Tories, as well as many people outside Parliament, had expected the fall of Walpole to result in a revolution in government and society, but this did not occur. Instead, all that had happened was a reshuffling of state employment among patrician Whigs, which caused widespread disillusionment and anger. It was with the Patriot Whigs in mind that Samuel Johnson, a staunch Tory, was later to describe patriotism in his *Dictionary* as the last resort of the scoundrel.

Carteret

When Wilmington died in 1743, Carteret took over as head of the administration. He was a clever and subtle man, able to speak many European languages, and fascinated by foreign affairs. These qualities naturally endeared him to the king. His status as a royal favourite was confirmed when he accompanied George on a military expedition to Germany in defense of the electorate of Hanover. In June George commanded his British and Hanoverian troops at the Battle of Dettingen (the last battle in which a British monarch commanded), defeating the opposing French forces. But the victory was not followed up and aroused little patriotic enthusiasm in Britain. Instead, accusations that the king and Carteret were sacrificing British interests to Hanoverian priorities were openly expressed in Parliament and in the press. The Pelham brothers took advantage of this discontent (and Carteret's absence) to undermine his political position. In November 1744 he was forced to resign, though during the next 18 months George II continued to consult with him privately on political business. These intrigues infuriated Henry Pelham, who was now first lord of the treasury and chancellor of the exchequer, and his brother Newcastle, who was secretary of state.

The Jacobite rebellion. Britain's involvement in the War of the Austrian Succession, Tory and popular anger at the political deals that followed Walpole's resignation, and the infighting among the Whig elite were the background to the Jacobite rebellion of 1745–46 (the Forty-five). Since Britain was now at odds with France, the latter power was willing to sponsor an invasion on behalf of the Stuart dynasty. It hoped that such an invasion would win support from the masses and from the Tory sector of the landed class. Although a handful of Tory conspirators encouraged these hopes, the degree of their commitment is open to question. A large-scale French naval invasion of Britain in early 1744 failed in part because these men would not commit themselves to action. In July 1745 the Old Pretender's eldest son, Charles Edward Stuart (the Young Pretender), landed in Scotland without substantial French aid. In September he and some 2,500 Scottish supporters defeated a British force of the same size at the Battle of Prestonpans. In December, with an army of 5,000 men, he marched into England and got as far south as the town of Derby, some 150 miles from London.

Landing of the Young Pretender

Charles's initial success owed much to the ineptitude, the unconcern even, of Britain's rulers. One problem was that the standing army was too small, consisting of some 62,000 men. Because of Britain's involvement in the War of the Austrian Succession, the bulk of this force was in Flanders and Germany. Only 4,000 men had been left to defend Scotland, and most of them were raw recruits. Moreover, hampered by internal divisions, the administration was slow to respond. When the Young Pretender landed, the Pelhams were anxious but Carteret, now Earl of Granville, was not. Nor, at the beginning, was George II, who was actually in Hanover when his rival for the throne landed. As a result of these squabbles and misunderstandings, Parliament did not assemble until Oct. 17, 1745. Because by law only Parliament could authorize money to pay the militia (Britain's civil defense force), this delay seriously impeded early resistance to the Jacobite force. The city of Carlisle in the north of England surrendered to the rebels in November largely because its militia had received no pay from the government or from anyone else for two months.

Some historians have argued that the mass of Britain's population cared little which dynasty ruled them at this time and that the Young Pretender would have regained the kingdom for the Stuarts if only he had pressed on to London. Clearly, this thesis can never be proved one way or the other. The Jacobites, however, did not try to march on to London but retreated to Scotland. Nonethe-

less, it is probably significant that the Young Pretender attracted scarcely any English supporters on his march to Derby. Only in Manchester, which had a large Catholic population, did he gain recruits—some 200 men, mostly unemployed weavers. No Tory landowner or politician joined him, nor did any men of influence or wealth come out in his favour. By contrast, once the seriousness of the invasion was recognized, many individuals joined home-defense units or subscribed money against it. Between September and December 57 civilian loyal associations are known to have been founded in 38 different counties. Merchants and traders in the prosperous towns—Liverpool, Norwich, Exeter, Bristol, and most of all London—were particularly prominent in loyalist activity.

Although many Britons had become disillusioned by events after Walpole's fall, probably few were seriously tempted by the prospect of a Jacobite restoration. The Young Pretender, a Roman Catholic, was viewed as the pawn of France, Britain's enemy and prime commercial and imperial competitor. Traditionally the Catholic religion and French politics were associated with absolutist government, religious persecution, and assaults on liberty. These prejudices worked against the Young Pretender's appeal, as did prejudices against the Scottish Highlanders, the bulk of his armed supporters, who were regarded as terrifying barbarians by many of the English. The lack of mass English support for the Stuarts in 1745 dissuaded the French government from sending substantial military aid to the rebels. On April 16, 1746, the Duke of Cumberland (George II's second son) defeated the Jacobite army at Culloden in northern Scotland. This was the last major land battle to occur in Great Britain. The Young Pretender escaped to France and finally died in 1788, sodden with drink and disillusionment.

Firmer integration of Scotland

The main result of the Forty-five was the British government's decision to integrate Scotland, and particularly the Scottish Highlands, more fully into the rest of the kingdom. Despite the Act of Union of 1707, clan chieftains had retained considerable judicial and military powers over their followers. But these powers were destroyed by the Abolition of Heritable Jurisdictions (Scotland) Act of 1747. Other statutes required oaths of allegiance to the Hanoverian dynasty from the Episcopalian clergy, banned the wearing of kilts and tartans in an attempt to erode distinctive Highlands practices, and confiscated arms. The administration also confiscated the estates of Highlands chieftains who had rebelled and used the proceeds to encourage trade and agriculture in Scotland. Indeed, the gradual pacification of Scotland and its partial integration into a united Britain probably owed more to growing prosperity than to legal changes. By the mid-1750s Scotland's population was estimated at 1,265,380, and it continued to grow at a rapid rate until the 1830s. Linen production doubled between 1750 and 1775, and coal mining, iron smelting, and agricultural productivity also began to expand. Economic and demographic growth was particularly dramatic in towns such as Glasgow, Edinburgh, Aberdeen, and Dundee. The Act of Union had made Britain the largest free-trade area in Europe, and, as more Scots came to profit from trading and manufacturing links with England, more had a vested interest in maintaining the status quo.

The rule of the Pelhams. Defeating the rebellion also strengthened the position of the Pelhams. In February 1746, George II attempted to replace them with Granville but failed. Thereafter Henry Pelham and Newcastle insisted upon and received the king's full confidence. The attempted invasion widened once again the gulf between the Whig and Tory parties. The Whigs became for a time more united, and the Tories did badly in the general election of 1747, winning only 110 seats. The only serious opposition Pelham faced after that date came from the heir to the throne, Frederick, Prince of Wales. Although Frederick had abandoned the Opposition in 1742, his impatience to succeed to the throne soon prompted him to drift back into political intrigue against his father and his father's ministers. He claimed to be motivated by some of Lord Bolingbroke's political ideas. In 1738, during Frederick's earlier phase of opposition, Bolingbroke had written *The Idea of a Patriot King,* arguing that a future ideal monarch could unify and purify the nation by seizing the initiative to abolish faction and ruling over an administration based on virtue rather than on party. Frederick's avowed commitment to a nonparty government attracted Tory as well as a few Whig MPs to his support in the late 1740s. But their schemes and hopes were dashed when Frederick died in 1751. His eldest son, George (the future George III), became heir to the throne, and serious opposition to Pelham effectively ceased. Debate in Parliament became so muted, one politician wrote, that a bird might have built its nest in the Speaker's wig and never be disturbed.

Both Pelham and Newcastle were overshadowed by their more famous predecessor Robert Walpole and by their charismatic successor, William Pitt the Elder. Like Walpole, both brothers regarded themselves as staunchly Whig though their ideology was by no means clear-cut. Like Walpole, they had little enthusiasm for British involvement in European wars. They helped to negotiate the Treaty of Aix-la-Chapelle (1748), which ended the War of the Austrian Succession. Like Walpole, too, the Pelhams sought to reduce the national debt and to keep taxation on land low. But unlike Walpole, they avoided corruption; both lost rather than made money during their political careers. And Henry Pelham was more interested in domestic reform than Walpole had been.

Domestic reforms. The Gin Act of 1751 was designed to reduce consumption of raw spirits, regarded by contemporaries as one of the main causes of crime in London. In 1752 Britain's calendar was brought into conformity with that used in continental Europe, previously 11 days ahead. It was once believed that protests against this change—"give us back our 11 days," crowds are supposed to have chanted—represented nothing more than parochial ignorance. In fact the adoption of the new calendar, though it ultimately benefited commerce and international relations, initially played havoc with monthly rental payments and wages in the short term. In 1753 the Marriage Act was passed to prevent secret marriages by unqualified clergymen. From then on, every bride and groom had to sign a marriage register or, if they were illiterate, make their mark upon it. This innovation has been of enormous value to historians, enabling them to establish how many Britons were able to write at this time and, by inference, how many could read.

The new calendar

BRITISH SOCIETY BY THE MID-18TH CENTURY

Joseph Massie's categories. From the Hanoverian succession to the mid-18th century the texture and quality of life in Britain changed considerably but by no means evenly. Change was far more pronounced in the towns than in the countryside and among the prosperous than among the poor. The latter category was still very large; in the late 1750s an economist named Joseph Massie estimated that the bottom 40 percent of the population had to survive on less than 14 percent of the nation's income. The rest of his calculations can be summarized as follows:

Estimate of Social Structure in 1759

annual income	number of families	social types
£5,000+	310	peers, great landowners
£1,000+	1,000	greater gentry
£600+	3,400	great merchants, squires
£100+	104,900	small landowners, clergy, traders, professionals
£50+	160,000	small traders, lesser clergy, farmers
Below £50	1,093,000	the rest

Massie's calculations were not exact since no official census was implemented in Britain until 1800. But his figures were probably broadly correct and are the best available for this period. It is noticeable that his top three categories had close connections with the land, still the bedrock of wealth, status, and power. The greatest landowners (Massie's 310 families) owned estates ranging from 10,000 to 20,000 acres. Many of them belonged to the peerage, that is, they were dukes, marquesses, earls,

viscounts, or barons. Such hereditary titles, which could only be granted by the crown, carried with them the right to sit in the House of Lords. In the reigns of George I and George II there were some 170 of these peers. Almost all of them possessed fine houses in London as well as one or more mansions in the counties where their land lay. The Dukes of Marlborough (Winston Churchill's ancestors), for example, dominated large parts of Oxfordshire from their stately home of Blenheim (built 1705–30). The Earls of Carlisle in Cumberland built Castle Howard in the same period, spending £35,000 on the house and a further £24,000 on the gardens. Together with the greater gentry and the squires, listed in Massie's second and third categories, great landowners such as these owned considerably more than half of the cultivatable land in Britain.

Not all wealthy men were landowners. The foundation of the Bank of England in 1694 and other finance companies made it possible to make fortunes on the stock market, and the expansion of trade and industry forged powerful mercantile dynasties such as the Whitbreads (brewing), Smiths (banking), and Strutts (textiles). Some of these self-made families purchased landed estates to advertise their new wealth; others made do with smart town houses or country villas. But although it was possible to be rich and influential in this society without owning broad acres, it was the landed elite that set the cultural tone and dominated positions of power in both central and local government. Every peer in the House of Lords and a majority of MPs in the House of Commons owned land. Landowners also monopolized the office of lord lieutenant. Lords lieutenant were the crown's leading representatives in each of the English and Welsh counties. (Only in the 1790s was this office extended to Scotland.) Appointed by the king, they were responsible for maintaining law and order in their counties and for organizing civil defense measures during time of war. To assist them in these tasks, they appointed deputy lieutenants and justices of the peace—offices usually held by the squires and lesser gentry in the countryside and by merchants and landed gentlemen in the towns. None of these offices carried salaries—a clear indication that they were confined to the prosperous. But they brought with them considerable local influence and status and were often much sought after.

Mercantile dynasties

Less is known about Massie's 4th, 5th, and 6th social categories than about the landowning classes. And much less is known about small merchants, tradesmen, professionals, artisans, and labourers in Wales and Scotland than about their English equivalents. Most historians believe that the middle-income groups were increasing in number in the mid-18th century. Professional opportunities in law, medicine, schoolteaching, banking, and government service certainly expanded at this time. In the town of Preston in Lancashire, for example, there was only one attorney in 1702; by 1728 there were 17. Growing prosperity also increased job opportunities in the leisure and luxury industries. Urban directories show that there were more musicians and music teachers, more dancing masters, booksellers, caterers, and landscape gardeners than in the 17th century. And there were more shops. Shops had expanded even into rural areas by the 1680s, but in the 18th century they proliferated at a much faster rate. By 1770 the new town of Birmingham in Warwickshire had 129 shops dealing in buttons and 56 selling toys, as well as 35 jewelers. Not for nothing would Napoleon Bonaparte later describe Britain as a nation of shopkeepers.

Proliferation of shops

Urban development. The centre of this commercial culture was the city of London. As the only real national metropolis London was unique in its size and multiplicity of functions. By 1750 it contained more than 650,000 citizens—just under one in 10 of Britain's population. By contrast, only one in 40 Frenchmen lived in Paris in this period. The Hague held only one in 50 of the inhabitants of the Netherlands, and Madrid was the home of just one in 80 Spaniards. Some of these great European capitals had no resident sovereign. Many others, such as Vienna and St. Petersburg (now Leningrad), were grand ceremonial and cultural centres but effectively isolated from the economic life of their national hinterland. London was different. It was not only the location of the Court and of Parliament but also the nation's chief port, its financial centre, the home of its printing industry, and the hub of its communications network. Britain's rulers were brought into constant proximity with powerful economic lobbies from all parts of the nation and with a large and constantly fluctuating portion of their subjects. Britons seem to have been more mobile than their fellow Europeans in this period, and then as now many traveled to the capital to find work and excitement. Perhaps as many as one in six Britons spent part of their working life in London in the 18th century.

London easily dwarfed the other British towns. In 1750 its nearest rival, Norwich, had fewer than 50,000 people. Nonetheless, the provincial towns, although functioning on quite a different scale from that of the metropolis, were also growing in size and importance at this time. In 1700 only 10 of them contained more than 10,000 people. By 1750 there were 17 towns with populations of that size, and by 1800 there were more than 50. As towns grew, they became better organized and safer, more pleasant places to live in. Because more stone was used in buildings, the risk of destruction by fire began to lessen. Towns acquired insurance companies and fire engines to protect their citizens. Supplies of clean water improved. Urban planning and architecture became more sophisticated and splendid, and the results can still be seen today in towns like Stamford in Lincolnshire or Bath in Somerset (now Avon). These provincial centres developed cultural lives of their own, with new theatres, assembly rooms, libraries, Freemason lodges, and coffeehouses. By mid-century there were at least nine coffeehouses in Bristol, six in both Liverpool and Chester, two in Northampton, and at least one in most substantial market towns. Such establishments supplied their male customers with newspapers and a place to gossip as well as with liquid refreshments. They also often served as a base for clubs, debating societies, and spontaneous political activity. Schools grew in number, in both the towns and the surrounding countryside. In just one English county, Northamptonshire, the local newspaper press advertised the establishment of more than 100 new schools between 1720 and 1760.

Provincial towns

Change and continuity. Historians have differed sharply over the impact these commercial and cultural innovations had on British society as a whole. Some have argued that only a minority of men and women were touched by them and that the countryside, which contained the majority of the population, continued on in its traditional ways and values. This is certainly true of parts of Britain. The Scottish Highlands, the mountainous central regions of Wales, and some English regions such as East Anglia remained predominately rural and agricultural. Old beliefs and superstitions lingered on there and elsewhere, often into the late 19th century. Although Parliament repealed the laws against witchcraft in the 1730s, for example, many men and women, and not just the illiterate, continued to believe in its power. (John Wesley, the founder of Methodism, was convinced that witches and the Devil had a real corporeal existence on earth.) It is true, too, that many of the new consumer goods that improved the quality of life for the prosperous—porcelain china, armchairs, fine mirrors, newspapers, and manufactured toys—were beyond the economic reach of the poor. And although new styles of interior decoration transformed the dwellings of the landed and mercantile classes—the sale of wallpaper, for example, had risen from 197,000 yards in 1713 to more than two million yards in 1785, a 10-fold increase—they rarely reached the impoverished. Some agricultural labourers and miners had only one set of clothes and lived in mud-lined cottages, caves, or cellars. Beggars, vagrants, and the unemployed might not possess even these basic commodities.

Yet it would be wrong to postulate too stark a contrast in life-styles between the town and countryside, between the wealthy and the lower orders. Points of contact between the various layers of British society were in fact increasing at this time. More and more country landowners, their womenfolk, and their servants succumbed (without, one suspects, too much trouble) to the temptation of spending some months every year sampling the pleasures of their

neighbourhood provincial town, consulting its lawyers and financial agents, and patronizing its shops. Many urban merchants, taking advantage of better roads and coach services, went to live in the countryside while maintaining their businesses in town. Lower down the social scale, hawkers and peddlers (itinerant traders) carried town-produced goods into the country areas and sold them there. Conversely, the growing demand for food in urban areas sucked in men and goods from the countryside. English drovers braved the old Roman roads and faltering bridle paths, the only routes available in Welsh counties such as Caernarvon and Anglesey, in order to purchase meat cattle for London and other towns. Every year tens of thousands of black cattle from the Scottish Highlands were driven southward until they reached the Smithfield meat market in London. Demand for manufactured goods fostered the spread of inland trade, as did increasing industrial specialization in the different British regions. Daniel Defoe illustrated this point by describing the multiple provenance of an affluent man's suit of clothes:

> A coat of woollen cloth from Yorkshire, a waistcoat of cullamancoe from Norwich, breeches of strong drugget from Devizes and Wiltshire, stockings of yarn from Westmoreland, a hat of felt from Leicestershire, gloves of leather from Somerset, shoes from Northampton, buttons from Macclesfield, or, if metal, from Birmingham, garters from Manchester, and a shirt of handmade linen from Lancashire or Scotland.

In short, Britain was not a static society, and the towns and the countryside were not entirely separate spheres. Men and women moved about to seek pleasure, to do business, to sell goods, to marry, or to find work; and their ideas and impressions shifted over time.

The revolution in communications. Increased mobility was made possible by a revolution in communications. In the earlier 18th century long-distance travel was rare and the idea of long-distance travel for pleasure was a contradiction in terms. The speediest coach journey between London and Cambridge (just 60 miles) took at least a day. Traveling from the capital to the town of Shrewsbury by coach took more than three days, and the journey to Edinburgh could last as long as 10 days. Some travelers made their wills before starting, as coaches easily overturned on bad roads or in swollen rivers. By 1750, however, privately financed turnpike roads had spread from London and its environs to major English provincial centres like Bristol, Manchester, Newcastle, Leeds, and Birmingham. In the 1760s and '70s they spread further into Wales and Scotland. The postal service also improved in this period, though again much more slowly in the Celtic fringe than in England. In 1765 only 30 Scottish towns enjoyed a daily postal service.

Long-distance travel

But the most dramatic advance in inland communication came in the form of the printed word. London's first daily newspaper appeared in 1702. By 1760 it had four dailies and six tri-weekly evening papers that circulated in the country at large as well as in the capital. But the provinces also generated their own newspapers, their own books, dictionaries, magazines, printed advertisements, and primers. In 1695 Parliament passed legislation allowing printing presses to be established freely outside London. Between 1700 and 1750 presses were founded in 57 English provincial towns, and they proliferated at an even faster rate in the last third of the 18th century:

Number of English Towns Acquiring Printing Presses, 1700–1800

years	number of towns
1700–1710	9
1710–1720	14
1720–1730	8
1730–1740	14
1740–1750	12
1750–1760	11
1760–1770	12
1770–1780	33
1780–1790	39
1790–1800	56

By 1725 no fewer than 22 provincial newspapers had emerged. By 1760 there were 37 such papers and by 1780, 50. In Scotland seven newspapers and periodicals were in existence by 1750, including the monthly *Scots Magazine,* which was printed in Edinburgh but could also be purchased from booksellers at Aberdeen, Glasgow, Dundee, Perth, and Stirling. Wales had no English-language newspaper until 1804, but many English papers found their way there.

Newspapers

By 1760 more than nine million newspapers were sold in Britain every year. Because they were expensive by the standards of the time (three or four pennies), one copy of a paper may have been shared and read by as many as 20 different people. There is little doubt that this explosion of newsprint helped to integrate the nation. All provincial newspapers and periodicals were parasitic on the London press. They borrowed large extracts from the more popular and controversial London papers and pamphlets. Increasingly, too, they broke the law and reprinted London journalists' accounts of debates in the House of Commons and House of Lords (printing parliamentary debates was illegal until 1770). Consequently, by the time of the Seven Years' War (1756–63), larger numbers of Britons than ever before had some access to political information. They were more aware of their country's military victories and defeats and more conscious of political scandals and protest. Politics was no longer just the preserve of the politicians at court, in Parliament, and in the country houses.

BRITAIN FROM 1754 TO 1783

Henry Pelham died in 1754 and was replaced as head of the administration by his brother, the Duke of Newcastle. Newcastle was shrewd, intelligent, and hard-working and possessed massive political experience. But he lacked self-confidence and a certain breadth of vision, and he was hampered by being in the House of Lords. In 1755 Henry Fox was appointed secretary of state and acted as the administration's spokesman in the Commons. Fox's promotion alienated a man who was far more interesting and remarkable than either of these ministers, William Pitt the Elder. Pitt had entered Parliament as an Opposition MP in the 1730s. In 1746 he had been appointed paymaster general, a highly lucrative state office. But Pitt, whose ambition was for fame and recognition rather than money, remained unsatisfied. The king, however, disliked him and successfully obstructed his career. In 1755 he dismissed Pitt, who began to attack Newcastle on imperial and foreign policy issues.

Conflict abroad. Although Britain and France had technically been at peace since 1748, both powers continued to harass each other in their colonial settlements in North America, the West Indies, and India. When the French attacked the British colony of Minorca in May 1756, war broke out; Britain allied itself with Prussia and France with Austria. Like every 18th-century war, this one began badly for Britain; it lost Oswego in North America as well as Minorca. There was an outcry in the press, and Newcastle and Fox resigned. In November Pitt was appointed secretary of state with William Cavendish, Duke of Devonshire, serving as nominal head of the new administration. But Pitt, still lacking royal approval or an adequate majority in the Commons, was dismissed by the king in April 1757. He returned to power in June, forming what was to be a highly effective wartime coalition with Newcastle. Pitt captured the attention and imagination of Parliament and of the people by his rhetoric and charisma; Newcastle employed his experience and industry to raise more than £160 million during the course of the war. But what cemented the coalition was Britain's naval and military successes. General Robert Clive defeated the French at the Battle of Plassey in 1757 and subsequently managed to establish British dominance in Bengal and the Carnatic, the two most profitable regions of India for European traders. Two years later large sections of the French fleet were destroyed at the naval battle of Quiberon Bay. When Quebec fell to General James Wolfe in 1759, British control of Canada was effectively secured. The island of Guadeloupe was captured in the same dramatic year, as were French trading bases on the west coast of Africa.

Wartime coalition

Most of these gains were confirmed by the Treaty of Paris (1763), though Britain restored Guadeloupe to the French in return for control of Canada. In the short term these victories resulted in a mood of patriotic exultation, especially among merchants. They looked to the new colonies to provide both fresh stocks of raw materials and eager markets for British manufactured goods: "Trade," Edmund Burke gloated, "had been made to flourish by war." This global victory, however, had been purchased at a high price. The conquest of Canada freed the American colonists from the fear of a French invasion from the north. Anxiety on this score had helped to foster American attachment to Britain. Now these fears had been relieved, and as early as 1760 some Britons and Americans anticipated that this would lead to difficulties. Furthermore, the enormous cost of the conflict led to drastic and sometimes damaging postwar economies, not least the deterioration of the Royal Navy, which would be an important factor in Britain's defeat in the American Revolution (1775–83). Postwar economies also forced British governments to explore new fiscal expedients, which aroused discontent at home and in the American colonies. Finally, the apparent unity and strength of Britain's elite during the Seven Years' War was deceptive: Newcastle and many of his allies were elderly men, Pitt was difficult and unstable, and old Whig and Tory alignments had ceased to have much meaning. All these factors helped to make the early reign of George III a period of conflict and instability.

Political instability in Britain. George II died in October 1760 and was succeeded by his grandson, who became George III. The new king became one of the most controversial British monarchs. In the first 10 years of his reign administrations changed no fewer than seven times. In October 1761 Pitt resigned and Newcastle was made to share power with the royal favourite, John Stuart, Earl of Bute. In May 1762 Newcastle too resigned, and Bute alone led the government until his resignation in April 1763. Bute was replaced by George Grenville, who was in turn dismissed in July 1765. For the next year Charles Watson-Wentworth, Marquess of Rockingham, served as first lord of the treasury. But in July 1766 Rockingham was sacked and replaced by Pitt, now elevated to the House of Lords as Earl of Chatham. Chatham soon lapsed into manic depression, and from 1768 to 1770 Augustus Henry Fitzroy, Duke of Grafton, led the government. Only in 1770 did the king find a minister whom he felt he could trust and deal with: Frederick, Lord North. Such high political instability undoubtedly hampered British efforts to resolve the problem of its American colonies.

But division and instability were not just confined to the court and parliament. The 1760s were a period of bad harvests, rising food prices, and sporadic unemployment. These economic and social problems helped to fuel the public agitation over John Wilkes, a Protestant dissenter and the son of a London malt distiller. In 1757 he bribed a rotten borough to elect him as its member of Parliament. An interesting, irresponsible, and cheerfully immoral man, Wilkes became well known in London society but failed to obtain a government post. His disappointment, as well as a bent toward iconoclasm, pushed him into opposition journalism. In April 1763 issue number 45 of his paper, the *North Briton* (a reference to the then chief minister Lord Bute, who was Scottish), was judged seditious. The government reacted by issuing a general warrant under which Wilkes and 48 additional persons were arrested. But Sir Charles Pratt, chief justice of the court of commons pleas, determined that this was a breach of Wilkes's parliamentary privilege, and he acquitted him. Soon after Wilkes fled to France to avoid another trial, this time for obscenity. In 1764 he was expelled from the Commons and tried in absentia for sedition, libel, and obscenity. But, as he did not return, he was declared an outlaw for impeding royal justice. In 1768, deeply in debt, he returned and was elected MP for the county of Middlesex, the most populous county constituency in England.

John Wilkes

Since Wilkes was still an outlaw, Parliament declared him ineligible for election, and for a time he was imprisoned in the Tower of London. Due in large part to Wilkes's organizational and propaganda skills, this precipitated a nationwide agitation; Wilkes was seen not only in England but also in the American colonies as a martyr for liberty. His plight raised the question of whether the will of the people or the decision of a Parliament elected by only a fraction of the people was supreme. In 1769 the Society for the Supporters of the Bill of Rights was founded to aid Wilkes and to press for parliamentary reform. Its members demanded parliamentary representation for important new towns such as Birmingham, Leeds, and Manchester, the abolition of rotten boroughs, and general admission to the franchise for men of movable property (*i.e.*, traders, merchants, and professionals). The English, as well as the American colonists, were becoming more interested in the connection between parliamentary representation (or the lack of it) and the obligation to pay taxes.

Society for the Supporters of the Bill of Rights

The American Revolution. The American issue was the final and most volatile element in the instability of the 1760s. Tension mounted, as far as British governments were concerned, primarily for two reasons. First, from this decade onward imperial organization received increased attention, and attempts were made to tighten British rule in Ireland and India as well as in the American colonies, a development that caused friction. Fiscal need was the second and more pressing problem. In 1763 the national debt stood at £114 million, and it continued to grow. Since the burden of taxation was already heavy for Britons, the government naturally looked to other sources of revenue. This was the background to George Grenville's decision, in 1765, to pass the Stamp Act, a measure designed to raise revenue in the American colonies by putting a tax on all legal and commercial papers. But it stirred up intense resentment in the colonies and, indirectly, in Britain, when the Americans boycotted British goods. In 1766 Rockingham repealed the Stamp Act while maintaining Parliament's right to legislate for the colonies. In 1767 Charles Townshend, then chancellor of exchequer, levied duties on certain imports into the colonies, including a duty on tea, and linked this proposal with plans to remodel colonial government. These measures exacerbated American discontent, though Parliament was not made to realize how much until 1774.

Historians have long disagreed over the question of how far George III himself was responsible for these tumultuous events. The Declaration of Independence (1776) unambiguously condemned the king as a tyrant. The so-called 19th-century British Whig historians also criticized the king in very harsh terms, maintaining, at their most extreme, that as a young prince he was indoctrinated with archaic and inflated ideas of royal power. When he came to the throne, he supposedly ousted his Whig ministers, replacing them with Tories, who were more sympathetic to royal ambitions. His arbitrary aims and policies, it was claimed, provoked the Wilkite agitation in Britain and drove the American colonists to rebel. George was consequently held directly responsible for the break-up of the British Empire. Finally, he was charged with employing bribery and corruption to persuade Parliament to do his bidding.

Assessment of George III's responsibility

Twentieth-century historians, in particular the Polish-born scholar Lewis Namier, have revised many of these extreme judgments. It has now been established that the king was neither educated in arbitrary ideas, nor did he preside over a Tory revival. Ministers such as Bute, Grenville, Townshend, and North regarded themselves as Whigs. But by the 1760s and '70s "Whig" and "Tory" were terms that had lost precise ideological significance, and the breakdown of these old partisan divisions undoubtedly contributed to ministerial instability at this time. There is little evidence that the king used corrupt influence to make Parliament accept his American policy. Indeed, it is unlikely that he initially even possessed an American policy; royal correspondence shows that he was rarely closely interested in American affairs before 1774. The colonists' drift toward opposition and independence was probably caused as much by their distance from London and their increasing prosperity as it was by British fiscal measures.

But George III cannot be entirely exonerated. When he succeeded, he was only 22, immature, idealistic, and not well-educated. His appointment of his decorative

favourite, Lord Bute, was a breach of the convention that monarchs should choose chief ministers possessed of political experience and proven abilities. In his dealings with other politicians George showed himself throughout his reign to be intransigent and obstinate, and he often confused his own personal feelings with the public welfare. He can scarcely be blamed for wanting to retain such an important part of his empire as the American colonies, but he can legitimately be criticized for insisting that the American war be continued after 1780, by which time it had become clear to his chief minister, Lord North, that Britain had lost.

Domestic responses to the American Revolution. Even at its outbreak in 1775 British attitudes to the American war were mixed. Many Protestant dissenters regarded the Americans as their brethren, for political and religious reasons. The City of London, and other commercial centres such as Glasgow, Norwich, and Newcastle, objected to the war because it disrupted highly profitable Anglo-American trade. Many British newspapers and cartoons adopted a pacifist and sometimes even a pro-American line. Other Britons believed, with George III, that rebellion against a monarch was sinful and that Parliament's authority must be preserved. Conventional patriotism became stronger after 1778, when France, Spain, and belatedly the Dutch, allied themselves with the Americans against Britain.

The next two years proved profoundly difficult. Fears that the French would invade Ireland as a prelude to invading the British mainland led ministers to encourage the creation of an Irish volunteer force some 40,000 strong. The Irish Protestant elite, led by Henry Grattan, used this force and the French threat to extract concessions from London. In 1783 Ireland was granted legislative independence, though it remained subject to George III. Declining British fortunes abroad also revived the issue of parliamentary reform. By 1779 three different reform groups had emerged, all of whom favoured peace with America. The Marquess of Rockingham and his parliamentary supporters (including his secretary, Edmund Burke) wanted to reduce official corruption and George III's influence in government. Another group, led by Christopher Wyvill, a one-time Anglican clergyman, wanted a moderate reform of the representative system. Wyvill and some of his supporters played with the idea of a national association, an assembly of reformers from each county in Britain, that would exist parallel to Parliament and be superior to it in constitutional zeal. A third small group, led by Charles James Fox, a Whig MP, and by former Wilkite activists, wanted more extensive political reform, including the secret ballot and annual general elections. In 1780 they founded the Society for Constitutional Information, which was designed to build public support for political change through the systematic production and distribution of libertarian propaganda.

The Gordon Riots

It was unlikely that any of these reforms would be implemented. But the Gordon Riots of June 1780 made it certain that they would not be. In 1778 Parliament had made minor concessions to British Roman Catholics, who were excluded from civil rights. Anti-Catholic prejudice, however, had been a powerful emotion in Britain since the Reformation in the 16th century, and Roman Catholicism tended to be associated by many with political absolutism and persecution. A movement to repeal the Catholic Relief Act of 1778, the Protestant Association, started in Scotland under the leadership of an unstable individual called Lord George Gordon. The movement reached London and exploded there in riots that lasted for eight days. More than 300 people were killed, and more damage was done to property than would be done in Paris during the French Revolution. For a time these riots gave reform and popular agitation a bad name. To many, the very name of Wyvill's National Association was dangerously suggestive of the Protestant Association, and the parliamentary reform movement lapsed until the 1790s.

Disasters at home were followed by further disasters abroad. Late in 1781 Britain learned of General Charles Cornwallis' surrender in America at the Battle of Yorktown. Parliamentary pressure to end the war now became irresistible. When in March 1782 Lord North's majority in the Commons fell to nine votes, he resigned, against the wishes of George III. A new administration, formed under Lord Rockingham, was committed to peace with America and moderate constitutional reform at home. When Rockingham died in July 1782, William Petty, Earl of Shelburne, became first lord of the treasury. In November of that year it was he who had the thankless task of concluding peace with the Americans and formally acknowledging their independence and British defeat in the Treaty of Paris.

BRITAIN FROM 1783 TO 1815

Defeat abroad and division at home led many Britons to believe that their country was in irreversible decline. The war had cost more than £236.4 million and had apparently brought only humiliation and the loss of one of the most profitable regions of the British Empire. Yet recovery was rapid, and by the time Britain again went to war—in 1793, against revolutionary France—it was wealthier and more powerful than it had been at the beginning of George III's reign.

In February 1783 Britain made a far from disadvantageous peace with its European enemies. Minorca and Florida were ceded to the Spanish, but Gibraltar was retained. France was given settlements in Senegal and Tobago, but Britain recovered other West Indian islands lost during the war. Holland gave Britain freedom of navigation in its spice islands and an important trading base in India. Nonetheless, this peace damaged Shelburne's reputation, and he resigned. A coalition administration was formed, led by Lord North and Charles James Fox. The king disliked it and ruthlessly sabotaged it. The Fox–North coalition planned to cement its authority by passing a bill to reform the government of British settlements in India, previously administered by the East India Company alone. The India Bill passed the Commons but, like every other piece of legislation not directly concerned with taxation, it had to be approved by a majority in the House of Lords. In advance of the vote the king let it be known that he would regard any peer who supported the bill with disfavour. The Lords duly threw the bill out in December 1783, providing the king with an excuse to dismiss Fox and North and replace them with William Pitt the Younger, the second son of the late Earl of Chatham. The general election of 1784 supplied Pitt with a parliamentary majority.

The Fox–North coalition

William Pitt the Younger. Pitt lived and died a bachelor, totally obsessed with political office. He was clever, single-minded, confident of his own abilities, and a natural politician. But perhaps his greatest asset in the early 1780s was his youth. He had entered Parliament in 1780 and was just 24 when he became first minister in 1783. Consequently, he was not associated in the public mind with the American debacle but seemed instead to promise a new era. Moreover, although he and George III never developed a close relationship, he did enjoy the king's support. Knowing that the alternative to Pitt was Fox (whom he hated), the king dealt with Pitt in a responsible manner. In 1788–89 the king suffered a major bout of insanity (or, according to some scholars, porphyria, a hereditary blood disease). Although he recovered, he thereafter interfered in politics far less than in his early reign. Pitt in turn treated the king tactfully. He dropped his early enthusiasm for parliamentary reform, and in 1801 he resigned over the issue of Roman Catholic emancipation (the extension of civil rights to Catholics) rather than force the king to accept it.

Royal support aided Pitt's control of his cabinet and political patronage. But what sustained him most in the 1780s and early 1790s was the quality and success of his measures. He reduced the national debt by £10 million between 1784 and 1793, in part by increasing tax revenue. He fostered legitimate trade and reduced smuggling by cutting import duties on certain commodities such as tea. In 1786 he signed an important commercial agreement, the Eden Treaty, with France. It was in keeping with the argument made by the economist Adam Smith in his *The Wealth of Nations* (1776) that Britain should be less economically dependent on trade with America and become

more adventurous in exploring trading opportunities in continental Europe. At home, Pitt strove for cheaper and more efficient administration; for example, he set up a stationery department to supply government offices with the necessary paper at a more economical rate. Abroad, he restored Britain's links with continental Europe and implemented imperial reorganization. In 1788 he signed the Triple Alliance between Britain, Prussia, and Holland, thereby ensuring that in a future war his country would not be bereft of allies as it had been during the American Revolution. In 1790 he demonstrated Britain's renewed power and prestige by negotiating a peace between Austria and Turkey. In 1784 he passed his own India Act, creating a board of control regulating Indian affairs and the East India Company. The board's members were nominated by the king from among the privy councillors. Finally, in 1791 the Canada Constitutional Act was passed. London became responsible for the government of both Lower and Upper Canada, but both provinces were given representative assemblies.

The Triple Alliance

Economic growth and prosperity. Many of Pitt's reforms and policies, such as his India Act, had been devised by previous ministers. But even though he did not originate all of his schemes, Pitt nonetheless deserves credit for actually implementing them. For all his priggish ruthlessness and occasional dishonesties (perhaps because of them), Pitt undoubtedly contributed to the restoration of national confidence; indeed, for many people, he became its very personification. But British recovery had wider and more complex causes than just one man's measures. At bottom, it was rooted in accelerating economic growth and unprecedented national prosperity:

Rate of British Economic Growth, 1770–1800

year	exports	industrial production
1700	100	100
1770	256	199
1780	246	197
1790	383	285
1800	544	387

These figures illustrate two striking points. First, in the 1770s British export performance and industrial productivity were perceptibly damaged by the American war. But, second, Britain's economic recovery after the war was rapid and dramatic. Particularly noticeable is the fact that the wars with revolutionary and Napoleonic France (1793–1802 and 1803–15) did not slow Britain's buoyant prosperity. Although Napoleon tried to blockade Britain in 1808 and again in 1811–12, he never succeeded in cutting the lifeline of its trade. In the period 1794–96 British exports averaged £21.7 million per annum. In the period 1804–06 the equivalent figure was £37.5 million, and during 1814–16, £44.4 million. These figures demonstrate how quickly Britain regained its American markets after 1783 and how extensive its other colonial markets were. But they are also one of many signs that the nation was experiencing the first Industrial Revolution.

The Industrial Revolution. Some historians have questioned whether the term Industrial Revolution can really be applied to the economic transformation of late 18th- and early 19th-century Britain. They point out that in terms of employment the industrial sector may not have overtaken the agricultural sector until the 1850s and that even then the average unit of production employed only 10 people. Large, anonymous factories did not become common until the late 19th century. Other scholars have argued, rightly, that industry did not suddenly take off in the 1780s and that even in 1700 Britain was a more industrialized state than its European competitors. But, despite all these qualifications, the available evidence suggests that by 1800 Britain was by far the most industrialized state in the world and that, because of this, its rate of economic growth must have accelerated in the last third of the 18th century.

Perhaps the most powerful evidence one can cite for these statements (which are inevitably controversial, given the ferocity and rapid fluctuations of the debate on the Industrial Revolution) is Britain's ability to sustain an unprecedented growth in its population from 1780 onward without suffering from major famines or acute unemployment. In 1770 the population was about 8.3 million. By 1790 it had reached 9.7 million; by 1811, 12.1 million; and by 1821, 14.2 million. By the latter date, it is estimated that 60 percent of Britain's population was 25 years of age or below. By comparison, while a similar rate of demographic growth occurred in Ireland, there was no Irish Industrial Revolution. Partly as a result of this, Ireland suffered the great famine in the 1840s, whereas there was no similar famine in Britain.

To say this is not to deny the dark side of early industrialization. The conditions of work were often brutal, particularly for the young. Industrial safety was minimal, and environmental pollution and unguarded machines led to horrific injuries. Mechanization ruined the livelihoods of some skilled craftsmen, most notably the handloom weavers. Nonetheless, it is probable that without industrialization the social costs of rapid population growth in Britain would have been far greater.

Causes of the Industrial Revolution

Although it is not easy to account for Britain's early industrialization, some facts stand out. Britain, unlike its prime European rival, France, was a small, compact island. Except in northern Scotland, it had no major forests or mountains to disrupt or impede its internal communications. The country possessed a range of natural ports facing the Atlantic, plenty of coastal shipping, and a good system of internal waterways. By the 1760s there were already 1,000 miles of inland canals in Britain; over the next 70 years 3,000 more miles of canals were constructed. Britain was also richly endowed with coal and iron ore, and these minerals were often located close together in counties such as Staffordshire, Northumberland, Lancashire, and Yorkshire.

Most importantly perhaps, Britain could draw on an ample supply of customers for its goods, both at home and overseas. Its colonies fed it with raw materials while also serving as captive customers. And its expanding population meant buoyant demand at home even in wartime when foreign trade was disrupted. The best illustration of these advantages is the cotton industry. Its Indian settlements supplied Britain with ever-increasing amounts of raw cotton, and annual cloth production soared from 50,000 pieces of cloth in 1770 to 400,000 pieces in 1800. Much of this output in textiles was consumed by the home market. Some scholars have argued that the increased wearing of cotton (which could be easily washed) as distinct from woolen clothes (which could not) improved health conditions, thus contributing to Britain's population expansion.

British opinion on the French Revolution

Britain during the French Revolution. The outbreak of the French Revolution in July 1789 initially heightened British national confidence. Some Britons welcomed it in the belief that civil commotion would weaken their prime European competitor. Many others, William Wordsworth, Samuel Taylor Coleridge, William Godwin, and Mary Wollstonecraft among them, felt confident that revolutionary France would become a new and enlightened state and that this process would in turn accelerate political, religious, and social change in Britain. By contrast, Edmund Burke's fierce denunciation in *Reflections on the Revolution in France* (1790) met with little immediate support, even among the political elite. Only when the new French regime guillotined Louis XVI and threatened to invade Holland did mainstream opinion in Britain begin to change and harden. In February 1793 Britain and France went to war.

There has been much debate over the degree to which British opinion on the war was united. Some historians have argued that Thomas Paine's best-seller, *The Rights of Man* (1791–92), fostered mass enthusiasm for democratic reform and mass alienation from Britain's ruling class. Paine attacked the monarchy, aristocracy, and all forms of privilege, and he demanded not only manhood suffrage and peace but also public education, old-age pensions, maternity benefits, and full employment. While he did not directly advocate a redistribution of property to fund these reforms, some contemporary radicals certainly did. A Newcastle schoolmaster, Thomas Spence, for ex-

ample, issued a penny periodical, *Pig's Meat* (a reference to Burke's savage description of the British masses as "the swinish multitude"), calling for the forcible nationalization of land.

These developments in radical ideology were made more significant by simultaneous developments in radical organization. In January 1792 a small coterie of London artisans led by a shoemaker, Thomas Hardy, formed a society to press for manhood suffrage. It cost only a shilling to join, and the weekly subscription was set at a penny so as to attract as many members as possible. These plebeian reformers, making use of Britain's growing communications network, corresponded with similar societies that had sprung up in response to the Revolution in the English provinces and in Scotland. In October 1793 Scottish radicals held what they styled a British Convention in Edinburgh, and a few of the English corresponding societies managed to send delegates there. They issued a manifesto demanding universal manhood suffrage and annual elections and affirming their faith in the principles of the French Revolution.

Corresponding societies

In terms of the number of men involved, these initiatives were always limited. Corresponding societies were far more widespread in London and the industrial north than in predominantly rural areas such as central Wales. Only a small proportion of rural and industrial labourers, as distinct from artisans, seems to have joined them. Even in the radical bastion of Sheffield (population 31,000) the local corresponding society attracted only 2,000 members, and most of these did not attend its meetings regularly. A minority of these activists were overtly Francophile and some may have wanted a French invasion of Britain and the establishment of a republican regime. Most corresponding-society members, however, seem to have been deeply attached to the British constitution and to have wanted only to reform it. But if these societies were not extensive or proto-revolutionary, they were still important and recognized as such. Contemporaries realized that for the first time in the 18th century working men throughout the nation were beginning to organize to achieve political change.

Pitt's ministry acted ruthlessly to suppress them. Leading Scottish radicals were arrested and given harsh sentences. In England habeas corpus was temporarily suspended, laws were passed prohibiting public meetings and demonstrations, and Thomas Hardy was tried for treason but acquitted. By 1795 the corresponding societies had formally ceased to meet. A minority of radicals, however, continued to agitate for reform in secret, some of them engaging in sedition. Particularly prominent in this respect were Irish dissidents. By now large numbers of Irish immigrants lived and worked in British towns. Some of them sympathized with the Irish Rising of 1798 and formed secret societies to overturn the government. Several Irish agitators were involved in the Spithead and Nore naval mutinies of 1797 that for a time immobilized the Royal Navy. In 1803 an Irishman and former shipmate of Horatio Nelson, Edward Despard, was executed in London for plotting a coup d'état. Just how dangerous and well-supported these various incidents were is uncertain. But there can be no doubt that successive British wartime administrations felt obliged to devote extensive resources to maintaining order at home. even though they were also fighting an unprecedentedly massive war abroad.

Irish dissidents

The Napoleonic Wars. The Napoleonic Wars were massive in their geographic scope, ranging, as far as Britain was concerned, over all of the five continents. They were massive, too, in terms of expense. From 1793 to the Battle of Waterloo in June 1815 the wars cost Britain more than £1,650,000,000. Only 25 percent of this sum was raised by government loans, the rest coming largely from taxation, not least from the income tax that was introduced in 1798. But the wars were massive most of all in terms of manpower. Between 1789 and 1815 the British army had to expand more than sixfold, to about a quarter of a million men. The Royal Navy, bedrock of British defense, aggression, trade, and empire, grew further and faster still. Before the wars it had employed 16,000 men; by the end of them, it employed more than 140,000. Because there was an acute danger between 1797 and 1805 that France would invade Britain, the civil defense force also had to be expanded. The militia was increased, and by 1803 more than 380,000 men were acting as volunteers in home-based cavalry and infantry regiments. In all, one in four adult males in Britain may have been in uniform by the early 19th century.

Despite these financial and military exertions, British governments found it extremely difficult to defeat France. In part this was because Pitt the Younger's abilities were more suited to peace than to war. But the main reason the conflict was so protracted was France's overwhelming military superiority on land. The historian Paul Kennedy has written of British and French power in this period:

> Like the whale and the elephant, each was by far the largest creature in its own domain. But British control of the sea routes could not by itself destroy the French hegemony in Europe, nor could Napoleon's military mastery reduce the islanders to surrender.

The first coalition of anti-French states, consisting of Britain, Russia, Prussia, Spain, Holland, and Austria, disintegrated by 1796. A British expeditionary force to aid Flanders and Holland was defeated, and Holland was occupied by the French. By 1797 the cost of maintaining its own forces and subsidizing those of its European allies had brought Britain to the verge of bankruptcy. For a time the Bank of England suspended payments in cash.

The British response to these developments was to concentrate on home defense and to consolidate its imperial and naval assets. Britain won a string of important naval victories in 1797, and in 1798 at the Battle of the Nile, Nelson defeated the French fleet anchored off Egypt, thereby safeguarding British possessions in India. Pitt also tried to solve the problem of Ireland. In 1801 the Act of Union took effect amalgamating Ireland with Great Britain and creating the United Kingdom. The Dublin Parliament ceased to exist, and Ireland's Protestant voters were allowed to return 100 MPs to Westminster. Pitt had hoped to sweeten the union by accompanying it with Roman Catholic emancipation, that is, by allowing Irish Catholics to vote and hold state office if they possessed the necessary property qualifications. George III opposed this concession, however, and Catholics were not admitted to full British citizenship until 1829. Pitt resigned and was succeeded as first minister by Henry Addington, the deeply conservative son of a successful doctor. It was his administration that signed the short-lived Treaty of Amiens with France in 1802.

Naval victories

War broke out again in May 1803. Once again, Britain demonstrated its power at sea but, until 1809, was unable to win substantial victories on land. Its fleet captured St. Lucia, Tobago, Dutch Guiana, the Cape of Good Hope, French Guiana, Java, Martinique, and other West Indian and African territories. Most importantly, in October 1805 Nelson defeated the French and Spanish fleets at Trafalgar, thereby preventing an invasion of Britain. Napoleon, however, inflicted serious military defeats on the Austrians, Prussians, and Russians and invaded Spain. At one stage Britain's only remaining European allies were Sweden, Portugal, Sicily, and Sardinia; in short, the country was without any significant allies at all. Political leadership was uneven and sometimes weak, and the long duration of the war and its damaging effects on trade aroused increasing criticism at home. Pitt had resumed his post as chancellor of the Exchequer and first lord of the Treasury in May 1804, but he died worn out by work and drink in January 1806. None of the three men who succeeded him as premier, William Wyndham Grenville, Baron Grenville (1806–07), William Henry Cavendish Bentinck, Duke of Portland (1807–09), and Spencer Perceval (1809–12), was able to establish himself in power for very long or to capture the public imagination.

Yet the war began to turn in Britain's favour in 1809, in large part because of Napoleon's strategic mistakes. When the Spanish rebelled against French rule, substantial British armed forces were dispatched to assist them under the command of Arthur Wellesley, later Duke of Wellington. Spain's new anti-French posture meant that Spain was once again open to British manufactured goods,

as were its colonies in Latin America. For a time this helped to reduce the commercial community's criticism of the conduct of the war. But demands for peace revived during the slump of 1811–12 and intensified when British relations with the United States, a vitally important market, began to deteriorate. One of the main irritants was the so-called Orders in Council, prohibiting neutral powers (like the United States) from trading with France. In 1812 commercial lobbies in Liverpool, Sheffield, Leeds, and Birmingham succeeded in getting the orders repealed, an indication of the growing political weight exercised by the manufacturing interest in Britain. Although this failed to prevent the Anglo-American War of 1812, neither Britain's trade nor its war efforts in Europe was seriously damaged by that conflict. Russia's break with Napoleon in 1812 opened up large markets for British goods in the Baltic and in northern Europe.

From 1812 onward Napoleon's defeat was merely a matter of time. In June 1813 Wellington defeated the French army in Spain at Victoria. The forces of Austria, Sweden, Prussia, and Russia expelled the French from Germany in the Battle of Leipzig (October 1813). This victory allowed Wellington, who had already crossed the Pyrenees, to advance upon Bayonne and Toulouse. Robert Stewart, Viscount Castlereagh, the secretary of state for foreign affairs, played the leading part in negotiating the Treaty of Chaumont in March 1814, which clarified allied war aims (including the expulsion of Napoleon), tightened allied unity, and made provision for a durable European settlement. The subsequent squabbles over the spoils of war were interrupted for a time when Napoleon escaped from his genteel exile on Elba and fought his last campaign from March to June 1815. Although his final defeat at Waterloo was accomplished by the allied armies, Britain secured prime credit. This textbook victory was to help Britain dominate Europe and much of the world for the next 100 years.

Waterloo

Imperial expansion. Britain's ultimate success against Napoleon, like its importance in this period as a whole, owed much to its wealth—its capacity to raise loans through its financial machinery and revenue through the prosperity of its inhabitants and the extent of its trade. But British success also owed much to the power of its navy and to the energy and aggressiveness of its ruling class, which was particularly apparent in the imperial expansion of this period. Britain sought to extend its control by legislation, by war, and by individual enterprise. The Acts of Union with Scotland in 1707 and with Ireland in 1801 tightened London's rule over its Celtic periphery, as did the laws passed to erode the autonomy of the Scottish Highlands after the rebellion of 1745. In the 1760s Britain sought not only to increase the revenue it gained from its North American colonies but also to shore up its military and administrative influence there. These measures failed, but Britain had more success with its Indian possessions. Between 1768 and 1774, in fact, the House of Commons devoted far more time to Indian affairs than to those of North America. Its discussions culminated in the passing of the India Act in 1784, which indicatively increased the government's authority over the East India Company and therefore over Britain's possessions in India.

Every major war Britain engaged in during this period increased its colonial power. The Seven Years' War was particularly successful in this respect, and so were the Napoleonic Wars. Between 1793 and 1815 Britain gained 20 colonies, including Tobago, Mauritius, Malta, St. Lucia, the Cape, and the United Provinces of Āgra and Oudh in India. By 1820 the total population of the territories it governed was 200 million, 26 percent of the world's total population. Not all of these acquisitions were formally directed by London. Captain James Cook's explorations of Australia and New Zealand after 1770 were in part an exercise in private enterprise and scientific inquiry. Nonetheless, British settlement of Australia at New South Wales began in 1787, in part because the mother country needed another repository for transported convicts previously sent to the North American colonies. The East India Company also retained considerable initiative in its military strategies. In 1819 Sir Thomas Stamford Raffles seized Singapore for the company and not on London's instructions. But, however acquired, all these acquisitions added to Britain's power and reputation. It was no accident, perhaps, that its two national anthems, "God Save the King" and "Rule Britannia," were composed in this period. For the privileged and the rich, this was preeminently an era of confidence and arrogance. (L.J.Co.)

Great Britain, 1815–1914

BRITAIN AFTER THE NAPOLEONIC WARS

The end of the long wars against Napoleon did not usher in a period of peace and contentment. Although both agricultural and industrial production had greatly, if unevenly, increased during the wars, the total national debt had nearly quadrupled since 1793. Of the total annual public revenue after 1815, more than half had to be employed to pay interest on this debt. Furthermore, the abolition of Pitt's income tax in 1816 meant that the debt burden fell on consumers—many of them with low incomes—and on industrialists. The archaic and regressive nature of the national taxation system, along with a mounting scale of locally levied poor-law rates, which fell heavily on middle-income groups, provoked widespread anxiety and criticism.

The postwar economy and society. The postwar period was marked by open social conflicts, most of them exacerbated by an economic slump. As the long-run process of industrialization continued, with a rising population and a cyclic pattern of relative prosperity and depression, many social conflicts centred on questions of what contemporaries called "corn and currency," agriculture and credit. Others were directly related to the growth of factories and towns and to the parallel development of middle-class and working-class consciousness.

The agriculturalists, who were predominant in Parliament, attempted to safeguard their wartime economic position by securing, in 1815, a new Corn Law designed to keep up grain prices and rents by taxing imported grain. Their political power enabled them to maintain economic protection. Nonetheless, many of them suffered, particularly after 1819, when there was a return to the gold standard, from a serious fall of agricultural prices. Debts contracted during the wars became more onerous as prices fell. There were many complaints of agricultural distress during the early 1820s.

The Corn Law of 1815

Many of the industrialists, an increasingly vociferous group outside Parliament, resented the passing of the Corn Law because it favoured the landed interests. Others objected to the return of gold in 1819, which was put into effect in 1821. Whatever their outlook, industrialists were beginning to demand a voice in Parliament. The term middle classes began to be used more frequently in social and political debate.

Town and village labourers were also unrepresented in Parliament, and it was they who bore the main brunt of the postwar difficulties. Bad harvests and high food prices left them hungry and discontented, and in the worst years, whenever bad harvests and industrial unemployment coincided, discontent assumed a political shape. Moreover, the development of a steam-driven factory system with new rhythms of work and new controls led to a breakdown in traditional family relationships and the growth of towns with structures of communication that were quite different from those of villages or preindustrial urban communities. These changes fostered the emergence, though it was not always shared, of the sense of a working class. There were radical riots in 1816, 1817, and particularly in 1819, the year of the Peterloo Massacre, when there was a clash in Manchester between workers and troops of the yeomanry, or local citizenry.

Local magistrates, without adequate police forces at their disposal, were often unsure how to deal either with secret "conspiracy" or with open challenges to authority, while the government of Robert Banks Jenkinson, Earl of Liverpool, with only rudimentary administrative machinery at its disposal, tended at first to follow a policy of repression. The Six Acts of 1819, associated with Henry Addington, Viscount Sidmouth, the home secretary, were designed to

reduce disturbances and to check the extension of radical propaganda and organization. They provoked sharp criticism from Whigs as well as from radicals, and they did not dispel the fear and suspicion that seemed to be threatening the stability of the whole social order.

Revival of confidence

There was a revival of confidence after 1821, as economic conditions improved and the government itself embarked on a program of economic reform. Sidmouth retired, to be succeeded by Sir Robert Peel; and Robert Stewart, Viscount Castlereagh, the foreign secretary, committed suicide. Even the king, George IV (1820–30), who had been drawn into the heart of politics when his estranged queen, Caroline, returned to England in 1820 and for a time became a radical heroine, ceased to be the target for continual radical abuse. Liverpool was a sufficiently able and sensible prime minister to work with new men and to move in new directions. Between 1821 and 1825 duties on raw material imports were reduced and tariff schedules were simplified; and in 1828, one year after Liverpool resigned, the fixed Corn Law of 1815 was replaced with a law providing for a sliding scale. During this same period Peel was reforming the criminal law. Even after the collapse of the economic boom of 1824–25 no attempt was made to return to policies of repression.

Foreign policy. There was a change of tone if not of principle in foreign policy, as in home affairs, after Castlereagh's suicide. Castlereagh himself, who had represented Britain at the Congress of Vienna in 1815, had refused to follow up the peace settlement he had signed, which entailed provisions for converting the Quadruple Alliance of the victorious wartime allies into an instrument of police action to suppress liberalism and nationalism anywhere in Europe. His policy was one of nonintervention. His successor at the Foreign Office, George Canning, propounded British objectives in colourful language and with a strong appeal to British public opinion and emphasized differences between British viewpoints and interests and those of the European great powers more than their common interests. In 1824 he recognized the independence of Spain's American colonies, declaring in a famous phrase that he was calling "the New World into existence to redress the balance of the Old." In 1826 he used British force to defend constitutional government in Portugal, while, in the tension-ridden area of the eastern Mediterranean, he supported the cause of Greek independence. Although he died in 1827, before the new Greek state came into existence, after having served for a few months as Lord Liverpool's successor as prime minister, his policies and styles were reasserted by Henry John Temple, Viscount Palmerston, who became foreign minister in 1830.

The beginning of political reform. Between the death of Canning and Palmerston's acceptance of office in a government presided over by the aristocratic Whig leader Charles Grey, 2nd Earl Grey, there had been a major shift in British politics. Canning's successor as prime minister, Frederick John Robinson, Viscount Goderich, who had been a successful chancellor of the exchequer under Liverpool, was unable to deal adequately with the tangle of Tory and Whig factions, and he was soon replaced by Wellington, the military hero of the Napoleonic Wars. It was the Wellington ministry that introduced the new Corn Law of 1828 and presented Peel, the prime minister's chief henchman, with the renewed opportunity of reforming the law and, in 1829, with the chance of creating a new paid and uniformed police force for London. Yet Wellington, more soldier than politician, had to tackle two very difficult tasks—coping with Irish disorders and holding together in the same government Tories who had supported and opposed Canning.

The issue of Catholic emancipation

Irish disorders centred, as they had since 1801, on the issue of Catholic emancipation, a favourite cause of the Whigs, who had been out of power since 1807. During the 18th century Catholics in England had achieved a measure of unofficial toleration, but in Ireland restrictions against Catholics holding office were still rigorously enforced. In 1823 Daniel O'Connell, a Dublin Roman Catholic lawyer, had founded the Catholic Association, the object of which was to give Roman Catholics in Ireland the same political and civil freedoms as Protestants. Employing impressive techniques of organization, he galvanized opinion in Ireland while at the same time mobilizing radical allies in England. In 1828 he won an election in County Clare, Ireland, so convincingly that Wellington, who—like the king—had always opposed Catholic emancipation, came to the conclusion that the government would have to push a measure for emancipation, which Canning had supported, through a Tory-dominated British Parliament. With difficulty he persuaded Peel, who had been tempted to resign, and the king, who had to be bullied, that an emancipation act was necessary and inevitable. Yet 128 "ultra-Tories" voted against the 1829 measure. Tory divisions left an opening for the Whigs, who, however, were divided on tactics. Some of them had joined the Canning ministry, but others had stayed aloof, biding their time. They had considerable support from financial interests and from religious dissenters, whose civil rights were recognized in 1828.

The death of George IV, in June 1830, speeded up events. After the accession of William IV (1830–37) and an inconclusive general election, Wellington, beset by many enemies, was defeated in November on a relatively unimportant motion on royal expenditure. In a year of renewed economic distress and of revolution in France, when the political reform issue was being raised again at public meetings in different parts of the country, Wellington had not made matters easier for himself by expressing complete confidence in the constitution as it stood. He decided, therefore, to resign, and the king sent for Earl Grey, who had been persona non grata with George IV, to ask him to form a new ministry. The government Grey assembled was predominantly aristocratic—and it included Canningites as well as Whigs—but the new prime minister, like most of his colleagues, was committed to introducing a measure of parliamentary reform. For this reason, 1830 marked a real parting of the ways. At last there was a break in the continuity of regime that led back to Pitt's victory over Fox in the 1780s and that had only temporarily been interrupted in 1806–07. Moreover, the new government, aristocratic or not, was the parent of most of the Whig–Liberal administrations of the next 35 years.

The year 1830 was also one of economic and social grievances, with religious issues still being thrown into the melee. In many parts of the southern countryside village labourers, backed, if not instigated, by the popular radical leader William Cobbett, were engaged in acts of violence against landlords and property (the "Captain Swing" disturbances), while in the midland and northern towns and cities, well-organized political reform movements were winning widespread support. The Whigs were as afraid of rural riots as the Tories and almost as suspicious of new urban radical leaders in cities like Birmingham. Corn laws, currency laws, poor laws, and game laws were all being attacked, while in the industrial north the demand was growing for new laws to protect factory labour. It was in such an atmosphere that the new Whig-led government prepared its promised reform bill.

THE POLITICS OF REFORM

Three approaches to reform

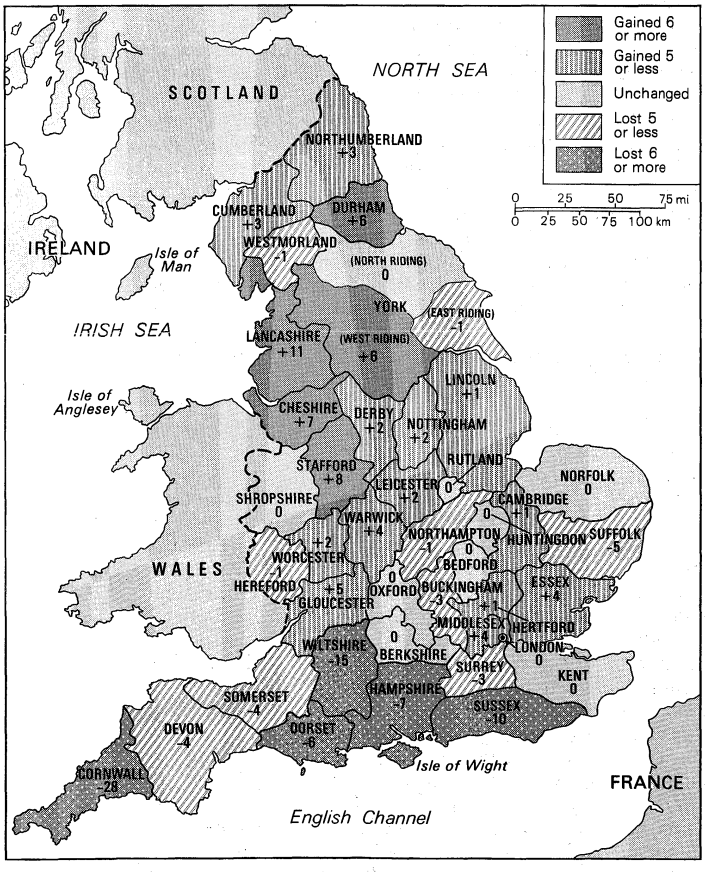
Whig interest in parliamentary reform went back to the 18th century, and Grey himself provided a link between two separate periods of public agitation. Yet, in the country as a whole, there were at least three approaches to the reform question. Middle-class "reformers" were anxious to secure representation for commercial and industrial interests and for towns and cities, like Birmingham and Manchester, that had no direct voice in Parliament. "Popular radicals," middle-class or working-class, were concerned with asserting rights as well as with relieving distress. "Philosophic radicals," the followers of the utilitarian philosophy of Jeremy Bentham, were strong ideological protagonists of parliamentary reform but deeply hostile both to the arguments and the tactics of the popular radicals, except when confident that they were in a position to deploy or control them. It was agitation in the country that kept the reform question on the boil between 1830 and 1832, while Grey, aloof from all forms of agitation, faced unprecedented constitutional difficulties with both the king and Parliament.

The Reform Act of 1832. A Whig reform bill, more sweeping than had been expected, was introduced in March 1831 and, on its first reading in the Commons, it was carried by one vote. In mid-April, however, after an opposition amendment had been successfully pressed, Grey induced a reluctant king to dissolve Parliament. At the ensuing general election the government won a clear majority on the single cry "the bill, the whole bill and nothing but the bill." A second reform bill passed the Commons with no difficulty but was defeated in October in the Lords. Immediately there was a public outcry, with mass meetings of "political unions" and, in some cities, riots. A third bill was then passed by the Commons, only to be thrown out again—on an amendment—in the Lords in May 1832. William refused Grey's request to create a number of new peers who would enable the bill to pass through the Lords, and, in consequence, Grey resigned and Wellington was called in. Such was the public mood, however, that he could not form a ministry, and Grey was reappointed, this time with a royal pledge that peers would be created if necessary. The threat was sufficient, and the bill passed, receiving the royal assent on June 7.

Achievement of the Reform Act

The Reform Act was in no sense a democratic measure. It was concerned with giving the middle classes a stake in government rather than with changing the basis of government. Yet it entailed a substantial redistribution of constituencies and a change in the franchise. The total electorate was increased by 57 percent to 217,000, but the artisans, the working classes, and some sections of the lower middle classes still remained outside "the pale of the constitution." No radical demands were met, even though the manner of passing the bill had demonstrated the force of organized opinion in the country, particularly in the large cities, which were now given representation. Those Tories who had prophesied that the act would mean revolution were wrong. The composition of the new House of Commons differed little from that of the old. It continued to reflect property rather than population, and landed interests remained by far the largest interests represented.

Further Whig reforms. Returned with a huge majority at the general election of December 1832, the Whigs carried out a number of other important reforms. A statute in 1833 ended slavery in the British colonies; in the same year the East India Company lost its monopoly of the China trade and became a purely governing body with no commercial functions. In 1834 a new Poor Law, recommended by a royal commission appointed in 1832, was passed; this law grouped parishes into unions and placed the unions under the control of elected boards of guardians, with a national Poor Law Board in London. In 1835 the Municipal Corporations Act was passed, which swept away old oligarchies in local government. Elected councils were to appoint town clerks and treasurers, and many unincorporated industrial communities were to be granted their first governmental powers. In some towns religious dissenters became the new governing class.



Membership change, by county, in the House of Commons as a result of the Reform Act of 1832 (England only).

The end of slavery was the final act in a long campaign in which a number of Tories had always played an important part, and the other reforming Whig measures owed much to the ideas of the philosophic radicals. The new Poor Law turned out to be an unpopular measure in many parts of the country, however, and led to outbreaks of disturbance. Its basic principle—that outdoor poor relief should cease and that conditions in workhouses should be "less eligible" than the worst conditions in the labour market outside—was as bitterly attacked by writers like Thomas Carlyle as by the workingmen themselves. In fact, procedures of inquiry and inspection, associated with this and many later reform measures, marked a change in the conduct of government that was at least as significant in the long run as the political reform of 1832. Public servants like Sir Edwin Chadwick, a disciple of Bentham, played an increasingly important, if controversial, part in administration, turning naturally from questions of poor relief to public health, education, and social reform. Inspired less by philanthropy than by belief in efficiency and economy, they extended the preoccupations of government at the very time when businessmen were seeking to demolish an inherited apparatus of economic control. Much administrative change in the 19th century was to have a momentum of its own, but its origins lay in new forms of social awareness, "public opinion," in an increasingly industrialized society.

The new Poor Law

The Whigs were not at ease in this changing context. Nor were they united in dealing with the problems either of England or of Ireland. Indeed, Grey's successor, William Lamb, Viscount Melbourne, home secretary from 1830 to 1832, was for a time pushed out of office in 1834, to return in 1835. He was adept in his dealings with the young Queen Victoria, who came to the throne in 1837, but incapable of finding effective answers to any of the pressing financial, economic, and social questions of the day. All of these questions multiplied after 1836, when a financial crisis ushered in a period of economic depression accompanied by a series of bad harvests. Social conflicts, never far from the surface, became more open and dramatic. Early Victorian England was turbulent and excited, and, if it had not been for Robert Peel, who succeeded Melbourne as prime minister after the general election of 1841 and had returned to power the "Conservatives" (as Peel liked to think of his party), there might well have been even greater disorder. The achievements of his Conservative ministry must be considered both within the context of the immediate social and political disturbance and within the perspectives of 19th-century British history as a whole.

SOCIAL CLEAVAGE AND SOCIAL CONTROL IN THE EARLY VICTORIAN YEARS

Chartism and the Anti-Corn Law League. As the economic skies darkened after 1836 and prophets like Carlyle anticipated cataclysmic upheaval, the two most disgruntled groups in society were the industrial workers and their employers. Each group developed new forms of organization and each turned from local to national extraparliamentary action. The two most important organizations were the Chartists and the Anti-Corn Law League. Chartism drew on a multiplicity of working-class grievances, extending working-class consciousness as it grew; the Anti-Corn Law

League, founded as a national organization in Manchester in 1839, was the spearhead of middle-class energies, and it enjoyed the advantage not only of lavish funds but also of a single-point program—the repeal of the restrictive Corn Laws.

Chartism, which aimed at parliamentary reform, took its name from the People's Charter published in London in May 1838. The charter contained six points, all of them political and all with a radical pedigree—annual parliaments; universal male suffrage; the ballot; no property qualifications for members of Parliament; payment of members; and equal electoral districts. These were old demands that would have been supported by 18th-century radicals. Behind the political demands at this time, however, was a fierce social discontent. In the provinces a main source of grievance was the new Poor Law of 1834, and in Lancashire and Yorkshire discontent focused on long working hours in the factories. Earlier localized agitations centring on such grievances were subsumed in Chartism because of its commitment to national political action. In addition, the failure to create effective trade unions during the early 1830s also diverted working-class efforts from economic to political action. The first trade unions failed because they could not maintain their strength at a time when economic conditions deteriorated and, more fundamentally, because employers then were far stronger than workers.

The Chartists also failed to secure any of the six points when they presented petitions to Parliament in 1839, 1842, and 1848. Problems of organization, local differences, disagreements about tactics (including the use of force), arguments about leadership (particularly about the leadership of Feargus O'Connor), and an improvement in economic conditions—first between 1844 and 1846 and then after 1848—dictated the details of the story. One of the few violent incidents was the small "rising" at Newport on the Welsh border in November 1839; it began like many of the other demonstrations, but it got out of hand before it failed.

Disagreement among Chartists

In Scotland "moral force" Chartism was particularly strong. In discontented Ireland, which might have provided support for Chartism, it was only after O'Connell's death in 1847 that the Chartists found allies.

The middle-class Anti-Corn Law League, led by Richard Cobden and John Bright, attempted to secure the repeal of the duties on imported grain, which were believed to raise the price of food for the workingman and benefit only the landowning classes. The league also had its difficulties, particularly at the outset. But it employed every device of propaganda, including the use of new media of communication, such as the penny post, which was introduced in 1840. The formula of the league was a simple one, designed to secure working-class as well as middle-class support. Repeal of the Corn Laws, it was argued, would settle the two great issues that faced Britain in the "hungry forties"—securing the prosperity of industry and guaranteeing the livelihood of the poor. The only barrier to salvation was the landlord. Most Chartists were unconvinced by this logic, but, in a landed Parliament, a few Anti-Corn Law Leaguers, led by Cobden, told Peel firmly that he would be "a criminal and a poltroon" if he did not repeal what they regarded as an immoral as well as an economically restrictive piece of legislation.

Peel and the Peelite heritage. Much depended on the nature of Peel's response to the problems of the time. Between 1832 and 1841 he had built up a disciplined party, most members of which accepted 1832 as a fait accompli. He himself, though brought up as a Tory, was a child of the Lancashire cotton industry and accepted industrialization as beneficial as well as inevitable. Afraid of violence, he sought, with a strong sense of public duty, to discover practical solutions to the complex issues of an industrializing society. He was the presiding genius of a powerful administration, strictly supervising the business of each separate branch of government, some of which were managed by very able lieutenants. From the start Peel attached top priority to financial reform. Beginning with his budget of 1842, he set about simplifying and reducing tariff restrictions on trade, and in the same year he reintroduced an income tax. In 1844 his Bank Charter Act laid the foundations of a sound national banking and credit system centred on the Bank of England. Finally, in 1846 he repealed the Corn Laws.

In this sequence of changes Peel alienated many of his followers, and repeal brought all the conflicts within his party to a head. A substantial section of the squirearchy rebelled, roused by the brilliant speeches of a young politician, Benjamin Disraeli, who in his writings had already approached the "condition of England question" in a totally different style from that of Peel. During the crisis Peel put his sense of duty to his sovereign, to posterity, and to his own conscience first and his obligations to his party second.

The results of repeal were important politically as well as economically. As a result of the split, party boundaries remained blurred until 1859, with the "Peelites" retaining a sense of identity even after Peel's premature death following a riding accident in 1850. Some of them, particularly William Gladstone, eventually became leaders of the late 19th-century Liberal Party, which emerged from the mid-century confusion. The protectionists, most of whom abandoned protection after 1852, formed the nucleus (around Edward Stanley, Earl of Derby, and Disraeli) of the later Conservative Party, but they were unable to secure a majority at any election until 1874. The minority governments they formed in 1852, 1858, and 1866 lacked any secure sense of authority. The Whigs, themselves divided into factions, returned to office in 1847 and held it for most of the mid-century years, but they were often dependent on radical and Irish support. Leadership in these years rested with strong or persuasive personalities, of whom Viscount Palmerston was the most prominent.

There was no time between 1846 and 1866, however, when extraparliamentary agitation assumed the dimensions it had done between 1838 and 1846. Only a fierce outburst of Chartism, this time with an injection of socialism and of Irish nationalism, disturbed the year. Ireland, positively or negatively, usually played an important part in the politics of crisis, for it was the failure of the Irish potato crop in 1845–46 and the threat of famine in Ireland that helped sway Peel to repeal the Corn Laws. Unfortunately, disaster was not averted in Ireland itself, and in the course of a few terrible famine years, about 500,000 Irish died and one million Irish emigrated. Peel's Whig successor, Lord John Russell, had no command over this situation, which challenged all the economic, administrative, and political assumptions of the time.

Irish famine

Social legislation and social control. In Britain, by contrast, steps were taken before and after Peel to assume a measure of public control over an increasingly urbanized and industrialized society. The question of public health, raised in the late 18th century and given a high degree of urgency during the 1830s, was the subject of several widely discussed reports before the passing of the first national Public Health Act in 1848. Chadwick was the leading spokesman of "the sanitary idea," which was canvassed vigorously by the novelists Charles Dickens and Charles Kingsley and which was to inspire George Eliot. Industrial questions also figured prominently in the social novels of the 1840s, including those of Disraeli; and statisticians, treated warily by most novelists, provided a different form of ammunition in the social debate. A report on conditions in the mines was followed by legislation in 1842 forbidding the employment underground of children under 10 years of age and of all women. Meanwhile, the reports of factory inspectors, appointed under an act of 1833, were attracting widespread interest; and in 1847, after earlier attempts had proved abortive, a Factory Act was passed limiting the hours of work for children in textile factories to 10: it was the first of many acts amending abuses and extending the principle of intervention.

In education there were significant new departures, limited in scope by the rivalry between churchmen and dissenters. In 1833 the first government grants had been made to the two main voluntary organizations sponsoring primary education, and in 1839 the first school inspectors had been appointed. J.P. Kay (later Sir James Kay-Shuttleworth), secretary of the Committee of Council on

Developments in education

Education, did everything he could to make the most of an inadequate system of provision, which, however, was extended somewhat with the first grants for teacher training and salaries in 1846. Kay's main object was to encourage the use of education as a means of introducing a measure of order and discipline into the working-class population, when older and more traditional methods of wielding authority through subordination had broken down. Yet just as the operation of the deterrent Poor Law directed attention in the long run to the need for more complex welfare policies, so the extension of educational provision, limited and belated though it was, involved the interplay of more varied motives and purposes. That there was no revolution in Britain in 1848, as there was in most countries in Europe, was largely due to the character of Britain's social structure, but the contemporary prophets of revolution, whether for or against, tended to overlook as other factors the effects of effort, local as well as national, to influence conditions of life through conscious policies.

The evolution of such policies was difficult for a number of reasons: first, there were entrenched vested interests, particularly obvious in questions of public health; second, there was a shortage of professional expertise, particularly in engineering and medicine; third, most radicals were highly suspicious of all forms of state intervention, particularly if they involved spending large sums of money or interfering with local freedoms; fourth, the orthodox economic policies of the day stressed the need for private rather than for public initiative. Yet there were aspects of the social structure that encouraged social action. Industrialists were sometimes especially sensitive to the problems of the workers in the countryside, particularly during the heyday of the Anti-Corn Law League, when the Leaguers were attacking "feudalism"; and some landlords, notably the evangelical philanthropist Lord Ashley, were especially sensitive throughout the 1830s and '40s to the problems of town workers. In such circumstances it was difficult to rest content with things as they were. In the words of one of Disraeli's characters, this was a "high-pressure" age. For the statistician G.R. Porter, whose *The Progress of the Nation* (1836–43) went through many editions, all "the elements of improvement" in the country were "working with incessant and increasing energy." He added that in his own lifetime he had seen "the greatest advances in civilisation that can be found recorded in the annals of mankind."

The pace of economic change. Not all of Porter's contemporaries would have agreed about the phrase "the greatest advances in civilisation," for some of them criticized the quality of life in the new society, and others pointed to the social contrasts reflected, for example, in the unequal distribution of incomes. There were radically contrasting approaches both to the political economy of the market and to the new coal, iron, and steam technology.

"Optimists" and "pessimists"

There was a debate between "optimists" and "pessimists," the latter suggesting that the condition of the working classes was actually deteriorating as national wealth increased. Most people were concerned, too, about the rise in population. At the first (defective) census of 1801, the population of England and Wales was about nine million and that of Scotland about 1.5 million. By 1851 the comparable figures were 18 million and three million. At its peak between 1811 and 1821, the growth rate for Britain as a whole was 17 percent for the decade. It took time to realize that the fears expressed so eloquently by Thomas Malthus that population would outrun subsistence were exaggerated and that, as population grew, national production would grow also.

Indeed, national income at constant prices increased nearly threefold between 1801 and 1851, substantially more than the increase in population; and the share of manufacturing, mining, and building in the national accounts of wealth increased sharply, as compared with the share of agriculture. In 1801 agriculture accounted for 34 percent and manufacturing, mining, and building for 28 percent. The comparable figures for 1851 were 21 percent and 40 percent. Cotton textiles remained the dominant new industry, with the cotton factory being thought of by one of its contemporary admirers, Edward Baines, as "the most striking example of the dominion obtained by human science over the powers of nature of which modern times can boast." There were 1,800 cotton factories in 1851. Raw cotton imports had increased unevenly from 101 million pounds in 1815 to 757 million pounds in 1851 and exports of manufactured cotton piece goods from 253 million yards in 1815 to 1,543,000,000 yards in 1851. Manchester was the centre of the cotton industry. Meanwhile, however, similar steam-driven technology accounted for the expansion of the woolen textiles industry over the same period, with Australia, which had provided no raw wool for Britain in 1815, providing 30 million pounds in 1851, and Bradford and Leeds were the centres of the woolen textile industry. It made no difference to the economics of textiles whether materials came from colonies, like Australia, or foreign countries, like the United States, or whether exports went to India or Brazil.

It was the textiles industries more than any others that illustrated Britain's dependence on international trade, a trade that it commanded not only through the size of the import bill or of manufacturing output but through the strength of its banking and other financial institutions and the extent of its shipping industry. Despite the advance of steam (registered steamship tonnage in 1815 was 1,000, and in 1851, 187,000), the tonnage of sailing ships also increased. There was no hardship comparable to that of the displaced handloom weavers, who were the victims of technological progress.

The new technology reached its peak in the age of the railway and the steamship. Coal production, about 13 million tons in 1815, increased five times during the next 50 years, and by 1850 Britain was producing more than 2 million tons of pig iron, half the world's output. Both coal and iron exports increased dramatically, with coal exports amounting to 3.3 million tons in 1851, as against less than 250,000 tons at the end of the French wars. Coal mining was scattered in the coal-producing districts; there were few large towns and miners lived a distinctive life, having their own patterns of work and leisure. Iron production was associated with larger plants and considerable urbanization. In South Wales, for example, one of the areas of industrial expansion, the Dowlais works were employing 6,000 people and turning out 20,000 tons of pig iron each year during the 1840s. Birmingham, Britain's second largest city, was the centre of a broad range of metallurgical industries. Organized mainly in small workshops, they differed sharply in character from the huge textile mills of Lancashire and Yorkshire.

Impact of the railroads

Industrialization preceded the coming of the railway, but the railway did much to lower transport costs, to consume raw materials, to stimulate investment through an extended capital market, and to influence the location of industry. The railway age may be said to have begun in 1830, when the line from Manchester to Liverpool, the country's most vigorously expanding port, was opened, and to have gone through its most hectic phases during the 1840s, when contemporaries talked of a "railway mania." By 1851, 6,800 miles of railway were open, some of them involving engineering feats of great complexity. There was as much argument among contemporaries about the impact of railways as there was about the impact of steam engines in factories, but there was general agreement about the fact that the coming of the railway marked a great divide in British social history. The novelist William Thackeray put it succinctly when he compared the last years of the stagecoach with the railway age: "Your railroad starts the new era, and we of a certain age belong to the new time and the old one." It was not until the 1870s and '80s that steamships brought this "new time" to its full realization, and by then British engineers and workmen had been responsible for building railways in all parts of the world. By 1890 Britain had more registered shipping tonnage than the rest of the world put together.

Some of the ships carried British emigrants overseas, both to distant parts of the empire and to the United States. Most of the emigrants owed little to government assistance, relying instead on the emigration industry that developed between 1815 and the late 1860s with its offices, agents, and publicity. Although individually emigrants

were propelled by a variety of motives, emigration as a whole, particularly the massive emigration from Ireland after the famine, served as a safety valve for the home country; it also made possible the opening up of large areas of the world.

The Great Exhibition of 1851. By 1890 it had become apparent that the British Industrial Revolution, far from being unique, was merely the first in a sequence of industrial revolutions and that Britain's early lead was becoming something of a handicap. In 1851, however, Britain was the workshop of the world and the main influence on the industrialization of other nations. The Great Exhibition of 1851 in London symbolized this economic supremacy. The exhibition was housed in a huge glass and iron building called by a journalist—with a touch of romance—"the Crystal Palace." There people from all parts of the world could examine machines of every kind, "England's arms of conquest . . . the trophies of her bloodless war." Part of the success of the exhibition was political as much as economic. The objects on display came from all parts of the world, including India and the countries with recent white settlements, such as Australia and New Zealand, that constituted the new empire. Many of the visitors who flocked to London came from European cities.

Victoria and Albert

The exhibition was a triumph not only for the economy but also for Victoria and her German husband, Albert, whom she had married in 1840. "In England," wrote a continental observer, "loyalty is a passion." Despite outbursts of opposition to Albert by the press, particularly in the patriotic mid-1850s, the family life of the Victorian court began to be considered increasingly as a model for the whole country. The fact that Albert had appreciated the significance of Peel's achievement and that he put his trust in the advancement of industry and science was as important as the fact that Victoria herself established monarchy on respectable foundations of family life. It was during the mid-1850s that the word "Victorian" began to be employed to express a new self-consciousness, both in relation to the nation and to the period through which it was passing. The death of Albert in 1861 and the subsequent withdrawal of the queen from public life led to a decline in the popularity of the court, but in time the queen's subjects, both at home and overseas, came once again to consider both her virtues and her obvious limitations as the very essence of "Victorianism" itself.

The colony of Australian settlers in the extreme southeast of the continent, with Melbourne as its largest city, received the name Victoria when it separated from New South Wales in 1851. It was important for its future that this event coincided with the discovery of gold, and the ensuing gold rush not only ensured the colony's future but also brought with it a new note of confidence in Britain itself.

MID-VICTORIAN SOCIETY AND CULTURE

After the excitements of the 1830s and '40s, mid-Victorian England was relatively quiet, with the family being regarded by most mid-Victorians as the central institution in society. Nationally a kind of balance was struck between the busy industrial north and midlands and the sleepy countryside described in the novels of Anthony Trollope. A kind of balance was also struck between the traditional ideal of "the gentleman" and the new ideal of the self-made man, allowing a place both for deference and dependence, on the one hand, and for individual advancement and acquisitiveness, on the other. There was far more talk during this period of self-help than there was of class conflict; indeed, the most comfortable social theory of the period rested on the assumption that class dividing lines could and should stay, provided that individuals in each class could move. Social discipline was strong, counting for more, perhaps, than the extension of the local police forces by an act of 1856.

Effects of economic prosperity

Victorian attitudes. Various kinds of balance rested, however, on economic as well as on psychological and sociological factors. From the early 1850s to the early 1870s, with occasional years of high unemployment and business failure (bad harvests counted for less as food imports increased), almost all sections of the population seemed to be benefiting from relative prosperity. Profits rose, as did wages and incomes from land. Indeed, those supporters of protection who had argued in the 1840s that free trade would ruin British agriculture were mocked by the mid-Victorian prosperity of agriculture in a golden age of high arable farming. There seemed to be little need in such a society for strong government, and it was only during the Crimean War (1854–56), when even many radicals looked for enemies abroad rather than at home and when the war itself was managed with obvious inefficiency, that either society or government seemed to be under great strain.

It was during these years, when great individual creative power was tapped, that Victorianism, perhaps the only "ism" in history attached to the name of a sovereign, came to represent a cluster of restraining moral attributes—"character," "duty," "will," earnestness, hard work, respectable comportment and behaviour, and thrift. These virtues were not only embraced by the striving bourgeoisie, but all of them also made an appeal to other class sections of the population, aristocratic or trade-unionist. Samuel Smiles, author of the best-seller *Self-Help* (1859), extolled and related them to one another in his many books. Later in the century, however, these values were taken apart and criticized, even lampooned, one by one, in the course of a late-Victorian revolt. More recently they have again received praise as essential "Victorian values."

Self-criticism and doubt

Yet despite their widespread appeal, all of these Victorian virtues were subjected to contemporary criticism. In the same year as Smiles's *Self-Help,* there appeared John Stuart Mill's essay *On Liberty,* a powerful attack on conformity and a classic liberal statement of freedom of discussion. Dickens frequently made fun of the Victorian smugness and unwillingness to face unpleasant facts, as represented by Mr. Podsnap in *Our Mutual Friend;* the critics John Ruskin and Matthew Arnold, each arguing from a different point of departure, questioned many of the accepted beliefs and prejudices of the age. Minority communicators in the widely read reviews were often extremely critical, and even Smiles himself, writing for a bigger public, thought that none of the virtues he was extolling came to men naturally nor could any of them be taken for granted in the middle of the 19th century. There was always a Victorian underworld. Belief in the family was accompanied by a high incidence of prostitution, and in every large city there were districts where every Victorian value was ignored or flouted. Mill's London was also the London of Henry Mayhew, whose sensational *London Labour and the London Poor* was published in book form in 1862. Many Victorians were as eager to read about crime as to read the Bible.

Religion. The critical sense of many of the great Victorians—at least as far as opinions were concerned—inevitably involved questions of religion as much as of society or politics, and Victorian doubt about inherited biblical religion was as much an acknowledged theme of the period as Victorian belief. Discoveries in geology and biology continued to challenge all accepted views of religious chronology handed down from the past. Perhaps the most profound challenge to religion came with Charles Darwin's *On the Origin of Species,* another of the great books of the remarkable year 1859. Yet the challenge was neither unprecedented nor unique. A year later *Essays and Reviews* was published; a lively appraisal of fundamental religious questions by a number of liberal-minded religious thinkers, it provoked the sharpest religious controversy of the century.

Behind such controversies, there were many signs of a confident belief on all sides that inquiry itself, if freely and honestly pursued, would do nothing to dissolve shared ideals of conduct. Even writers who were "agnostic" talked of the "religion of humanity" or tried to be good "for good's sake, not God's." Standards were felt to count in institutional as well as in private life. These were years, therefore, when the extension of the civil service involved the development of a remarkable code of institutional morality. Following a report by Sir Charles Trevelyan and Sir Stafford Northcote in 1853–54, a civil service commission was set up. Recruitment and promotion in most parts of the service were to depend on competitive examination.

An order-in-council of 1870 made this system mandatory, except for the Foreign Office. The extended civil service that took shape owed little to political patronage and was almost completely free from corruption.

Emphasis on conduct was, of course, related to religion. The English religious spectrum was of many colours. The Church of England was flanked on one side by Rome and on the other by religious dissent. Both were active forces to be reckoned with. The Roman Catholic church was growing in importance not only in the Irish sections of the industrial cities but also among university students and teachers. Dissent had a grip on the whole culture of large sections of the middle classes, dismissed too abruptly by Matthew Arnold as classes of "mutilated and incomplete men." Sometimes the local battle between Church of England and Dissent was bitterly contested, with Nonconformists opposing church rates (taxes), challenging closed foundations, and preaching total abstinence and educational reform. A whole network of local voluntary bodies, led either by Anglicans or Dissenters, usually in rivalry, came into existence, representing a tribute to the energies of the age and to its fear of state intervention.

The Church of England itself was a divided family, with different groups contending for positions of influence. The

The Oxford Movement

High Church movement (which emphasized the "Catholic" side of Anglicanism) was given a distinctive character, first by the Oxford Movement, or Tractarianism, which had grown up in the 1830s as a reaction against the new liberal theology, and then by the often provocative and always controversial ritualist agitation of the 1850s and '60s. The fact that prominent members of the Church of England flirted with "Romanism" and even crossed the Rubicon often raised the popular Protestant cry of the "church in danger." Peel's conversion to free trade in 1846 scarcely created any more excitement than John Henry Newman's conversion to Rome the previous year, while in 1850 Lord John Russell, Peel's successor as prime minister, tried to capitalize politically on violent antipapal feelings stimulated by the pope's decision to create Roman Catholic dioceses in England.

The Evangelicals, in many ways the most influential as well as the most distinctively English religious group, were suspicious both of ritual and of appeals to any authority other than that of the Bible. Their concern with individual conduct was a force making for social conformity during the middle years of the century rather than for that depth of individual religious experience that the first advocates of "vital religion" had preached in the 18th century. Yet leaders, such as Lord Ashley, were prepared to probe some Evangelical social issues, like housing, and to stir men's consciences; and, even if their preoccupation was with saving souls, their missionary zeal influenced developments overseas as well as domestic legislation. There were other members of the church who urged the cause of "Christian Socialism." Their intellectual guide was the outstanding Anglican theologian Frederick Denison Maurice. The Evangelicals, in particular, were drawn into substantial missionary activity in the empire and other parts of the world, frequently clashing with settlers and administrators, and sometimes with soldiers. They regarded it as their sacred duty to spread the gospel from "Greenland's icy mountains" to "India's coral strand."

Beyond the influence of both church and chapel there were thousands of people in mid-Victorian England who were ignorant of, or indifferent toward, the message of Christianity, a fact demonstrated by England's one religious census in 1851. Although movements like the Salvation Army, founded by William Booth in 1865, attempted to rally the poor of the great cities, there were many signs of apathy or even hostility. There was also a small but active secularist agitation; particularly in London forces making for what came to be described as "secularism" (more goods, more leisure, more travel) could undermine spiritual concerns. The great religious controversies of mid-Victorian England were not so much to be settled as shelved.

The Church of Scotland

In Scotland, where the Church of Scotland had been fashioned by the people against the crown, there was a revival of Presbyterianism in the 1820s and '30s. A complex and protracted controversy, centring on the right of congregations to exclude candidates for the ministry whom they thought unsuitable, ended in schism. In 1843, 474 ministers left the Church of Scotland and established a free church. Within four years they had raised more than £1,250,000 and built 654 churches. This was a remarkable effort, even in a great age of church and chapel building. It left Scotland with a religious pattern even more different from that of England than it had been in 1815. Yet many of the most influential voices in mid-Victorian Britain, including Carlyle and Smiles, were Scottish voices, and the conception of the gospel of work, in particular, owed much in content and tone, even if often indirectly, to Scottish Calvinism. In Wales there was a particularly vigorous upsurge of conformity, and the Welsh chapel was to influence late 19th-century and 20th-century British politics.

MID-VICTORIAN POLITICS

Religious questions helped divide the limited mid-Victorian electorate, with the dissenters encouraging, from their local bases, the development of liberalism, and churchmen often—but by no means universally—supporting the Conservative Party. Nonelectors' associations tried with varying degrees of success to keep radical issues alive, but party divisions remained based on customary allegiance as much as on careful scrutiny of issues, and there was still considerable scope for bribery at election times. The civil service might be pure, but the electors often were not. The Corrupt Practices Act of 1854 provided a more exact definition of bribery than there had been before, but it was not until a further act of 1883 that election expenses were rigorously controlled. During the mid-Victorian years the way to Parliament often led through the pigsty.

The Corrupt Practices Act

The prestige of the individual member of Parliament was high, and the fragmentation of parties after 1846 allowed him considerable independence. Groups of members supporting particular economic interests, especially the railways, could often determine parliamentary strategies. Contemporaries feared such interests less than they feared what was often called the most dangerous of all interests, executive government. Powerful government and large-scale "organic" reform were considered dangerous, and even those radicals who supported organic reform, like Cobden and Bright, were suspicious of powerful government. For most politicians, politics was identified not with theories or even programs but with pragmatic leadership.

In his interesting analysis of the English constitution (1867), the journalist Walter Bagehot, as knowledgeable about economics as about politics, considered the cabinet as "a board of control chosen by the legislature, out of persons whom it trusts and knows, to rule the nation." Its primary task was to administer, not to legislate. There was little legislation dealing with public health or trade unions or Irish agrarian problems until the late 1860s, although the Company Act of 1862 consolidated limited liability legislation passed during the 1850s. A judge told a number of men who were convicted of illegal activities during a strike of 1867,

> Everybody knows that the total aggregate happiness of mankind is increased by every man being left to the unbiased, unfettered determination of his own will and judgement as to how he will employ his industry and other means of getting on in the world.

Palmerston. Palmerston, an aristocrat born in 1784, stood out as the dominant political personality in mid-Victorian Britain, precisely because he was opposed to dramatic change and because he knew through long experience how to maneuver politics within the half-reformed constitution. In a period when it was difficult to collect parliamentary majorities he often forced decisions, as at the general election of 1857, on the simple question, "Are you for or against me?" He was skillful also in using the growing power of the press in order to reinforce his influence. At a time of party confusion, when the queen might well have played a key part in politics, Palmerston found the answer to royal opposition in popular prestige, carefully stage-managed. His chief preoccupation was with foreign affairs, and his approach was diametrically opposed to that of the court on several occasions.

There was no contradiction between his views on domestic and foreign policy. He preferred the English system of constitutional government, resting on secure social foundations, to continental absolutism, but, like Canning before him, he was anxious above all else to advance the interests of England as he saw them. The supremacy of British sea power, British economic ascendancy, and political divisions inside each of the main countries of Europe before and after the revolutions of 1848 gave him his opportunity. He liked to appear energetic. In 1850 he sent a blockading squadron to Greece to enforce payments of debts due to Don Pacifico, a Gibraltar-born British subject, and restated the doctrine of *civis Romanus sum* ("I am a Roman citizen," by which an ancient Roman could proclaim his rights throughout the empire) in a Victorian setting. In 1852, when he helped overturn Russell's shaky government, he had his revenge on Russell, who had dismissed him from his post as foreign secretary in December 1851 for welcoming Napoleon III's coup d'état in France. In January 1855, after George Gordon, Earl of Aberdeen, a Peelite, had shown his inadequacies as a war leader during the Crimean War, Palmerston was made prime minister for the first time. His interventions were not confined to Europe. In 1840–41 he had forced the China ports open to foreign trade and, by the Treaty of Nanking (1842), acquired Hong Kong for Britain. In 1857 he went to war in China again and, when defeated in Parliament, appealed triumphantly to the country. Although his government was defeated in 1858, he was back again as prime minister, for the last time, a year later.

The Don Pacifico affair

During the remarkable ministry of 1859–65, which included Russell as foreign secretary and the Peelite Gladstone as chancellor of the exchequer, it was impossible for Britain to dominate the international scene as effectively as in previous periods of Palmerstonian power. With efficient military power at his disposal, the Prussian prime minister, Otto von Bismarck, proved more than a match for Palmerston. The union of modern Italy, which Palmerston supported, the American Civil War, in which his sympathies were with the South, and the rise of Bismarck's Germany, which he did not understand, were developments that reshaped the world in which he had been able to achieve so much by forceful opportunism. When he died, in October 1865, it was clear that in foreign relations as well as in home politics there would have to be what Gladstone described as "a new commencement."

In the large urban constituencies the demand for a new and active liberalism had already been gaining ground, and at Westminster itself Gladstone was beginning to identify himself not only with the continued advance of free trade but also with the demand for parliamentary reform. In 1864 he forecast new directions in politics when he stated that the burden of proof concerning the case for reform rested not with the reformers but with their opponents. A year later he lost his seat at Oxford University and was returned "unmuzzled" as representative for a populous Lancashire constituency. The timing was right, because after the death of Palmerston the question of parliamentary reform was reopened and the Second Reform Bill was passed in 1867.

The Reform Bill of 1867. Yet it was Disraeli and not Gladstone who claimed the credit for the act of 1867. On Palmerston's death, Russell and Gladstone had introduced a modest and colourless bill that was severely mauled by both Conservatives and reform Liberals. The government resigned, and Derby and Disraeli took office. It was difficult to shelve the demand for reform, and the government decided to "dish the Whigs" and "take a leap in the dark." Agitation in the country was more vociferous on the issue than it had been since the days of the Chartists, and organizations, notably the Reform League, were engaged in stirring the public, alongside prominent individuals, notably John Bright. In a Parliament where Disraeli was in a minority, his only chance lay in accepting amendments, however radical, to the bill that he had introduced and in claiming them as his own. All reserves and safeguards were dropped, and, although he lost some of his own supporters, he eventually carried a bill very different from the one he had introduced. It added 938,000 new names to the register, amounting almost to a doubling of the electorate, and gave the vote to many workingmen in the towns and cities. The county franchise was not substantially changed, but 45 new seats were created by taking one member from existing borough constituencies with a population of less than 10,000. Disraeli hoped that, in return for passing this measure, urban workingmen would vote for him—he believed rightly that many of them were Conservatives already by instinct and allegiance—but, at the first general election under the new system, it was Gladstone who was returned as prime minister.

The "leap in the dark"

GLADSTONE AND DISRAELI

The choice between Gladstone and Disraeli was the first of many similar choices offered to an extended electorate—the choice not only between two programs but also between two men who were completely different in temperament and political outlook. Gladstone had made his mark first and lived on far longer (1898) than did his rival (1881). He saw politics in terms of moral principles and, in his ministry of 1868–74, introduced some of the most important Liberal legislation of the 19th century. Disraeli, who combined opportunism and political imagination, carried through an impressive program of social reform and embarked upon an active foreign policy both in Europe and overseas. Yet much of the Liberal legislation was the product of compromise rather than of principle, and much of the social legislation of the Conservatives was not considered important enough by them at the time to stress it in the election of 1880. In both parties new forces were stirring at the local level, and energetic efforts were underway to organize the electorate and the political parties along new lines. With the development of central party machinery and local organization, the role of the crown was reduced during this period to that of merely ratifying the result of elections. Although the queen greatly preferred Disraeli to Gladstone, she could not keep Gladstone out. Her obvious partisanship made some of her acts look unconstitutional, but they would not have been deemed unconstitutional in any previous period of history. The public during this period was more interested in the political leaders than in the queen, who lived in retirement and was sharply criticized in sections of the press.

Gladstone's first administration. The achievements of Gladstone's first administration were several: the disestablishment and partial disendowment of the Irish church, accomplished in 1869 in face of the opposition of the House of Lords; the Irish Land Act of 1870, providing some safeguards to Irish tenant farmers; William Edward Forster's Education Act of the same year, the first national act dealing with primary education; the Trade-Union Act of 1871, legalizing unions and giving them the protection of the courts; and the Ballot Act of 1872, introducing secret voting. There were many other important reforms, most of which were designed to broaden the span of individual opportunity or to reform cumbersome administrative machinery. In 1871, for example, the universities of Oxford and Cambridge were opened to Dissenters, while between 1868 and 1873 the cumbrous military machine was renovated by Gladstone's secretary for war, Edward Cardwell. The system of dual responsibility of commander in chief and secretary for war was abolished, and the subordination of the former to the latter was asserted. In 1873 the Judicature Act, amended in 1876, simplified the tangle of legal institutions and procedures.

Administrative reforms

Many of these reforms did not satisfy affected interests. The Irish Church Disestablishment Act failed to placate the Irish and alarmed many English churchmen, while the Education Act was passed only in the face of bitter Nonconformist opposition. The Dissenters objected that Forster's system did not break the power of the church over primary education, and, although the act was extended in 1880 when primary education was made compulsory and in 1891 when it became free, there were often noisy struggles between churchmen and Dissenters in the new school boards set up locally under the Forster Act. If the Education Act alienated many Dissenters, the Licensing Bills of 1871 and 1872 alienated their enemies,

the brewers. At the general election of 1874, therefore, months after Disraeli had described the Liberal leaders in one of his many memorable phrases as a "range of exhausted volcanoes," the brewers threw all their influence behind the Conservatives. "We have been borne down in a torrent of gin and beer," Gladstone complained.

Disraeli's second administration. Disraeli's ministry embarked upon a sizable program of social legislation. Gladstone, throughout his life, preferred cheap and free government to expensive and socially committed government. He was anxious, indeed, in 1873 to abolish income tax, on which the public finances of the future were to depend. Disraeli had always been interested in "the condition of England question" and, with the assistance of men like Richard Cross, the home secretary, justified at last his reputation as a social reformer. By the Employers and Workmen Act of 1875, "masters" and "men" were put on an equal footing as regards breaches of contract, while by a Trade-Union Act of 1875 that went much further than the Liberal act of 1871 trade unionists were allowed to engage in peaceful picketing and to do whatever would not be criminal if done by an individual. The Public Health Act of 1875 created a public health authority in every area; the Artizans' and Labourers' Dwellings Improvement Act of the same year enabled local authorities to embark upon schemes of slum clearance; a factory act of 1878 fixed a 56-hour week; while further legislation dealt with Friendly Societies (private societies for mutual-health and old-age insurance), the protection of seamen, land improvements carried out by tenants, and the adulteration of food. There was no similar burst of social legislation until after 1906.

Foreign policy. If there were significant, if not fully acknowledged, differences between the records of the two governments on domestic issues, there were open, even strident differences on questions of foreign policy. Gladstone had never been a Palmerstonian. He was always anxious to avoid the resort to force, and he put his trust not in national prejudices but in an enlightened public opinion in Europe as well as England. His object was justice rather than power. In practice, however, he often gave the impression of a man who vacillated and could not act firmly. Disraeli was willing to take risks to enhance British prestige and to seek to profit from, rather than to moralize about, foreign dissensions. His first ventures in "imperialism"—a speech at the Crystal Palace in 1872, the purchase of the Suez Canal shares in 1875, and the proclamation of the queen as "Empress of India"—showed that he had abandoned the view, popular during the middle years of the century, that colonies were millstones around the mother country's neck. But these moves did not involve him in any European entanglements, nor did the costly if brilliantly led campaigns of Major General Frederick Roberts in Afghanistan (1878–80) and the annexation of the Transvaal in South Africa in 1877.

Disraeli's "imperialism"

It was the Middle Eastern crisis of 1875–78 that produced the liveliest 19th-century debate on foreign policy issues. In May 1876 Disraeli rejected overtures made by Russia, Austria-Hungary, and Germany to deal jointly with Turkey, which was faced with revolt in Serbia. His pro-Turkish sympathies irritated many Liberals, and, after Turkey had gone on to suppress with great violence a revolt in Bulgaria in 1876, the Liberal conscience was stirred and mass meetings were held in many parts of the country. Gladstone, who had gone into retirement as Liberal leader in 1875, was slower to respond to the issue than many of his followers, but, once roused, he emerged from retirement, wrote an immensely influential pamphlet on the atrocities, and led a public campaign on the platform and in the press. For him the Turks were "inhuman and despotic," and, whatever the national interests involved, Britain, in his view, should do nothing to support them. Disraeli's calculations concerned strategic and imperial necessities rather than ideals of conduct, and his suspicions were justified when the Russians attacked Turkey in April 1877. Opinion swung back to his side, and in 1878 Disraeli sent a British fleet to the Dardanelles. London was seized by war fever—the term jingoism was used to describe it—which intensified when news arrived that a peace treaty had been signed at San Stefano whereby Turkey accepted maximum Russian demands. Reservists were mobilized in Britain, and Indian troops were sent to the Mediterranean. Disraeli's foreign minister, who disapproved of such action, resigned, to be succeeded by Robert Gascoyne-Cecil, Marquess of Salisbury, who was eventually to serve as prime minister in the last Conservative administrations of the 19th century. The immediate crisis passed, and, at an international conference held in Berlin in June and July 1878, which Disraeli attended, the inroads into Turkish territory were reduced, Russia was kept well away from Constantinople, and Britain acquired Cyprus. Disraeli brought back "peace with honour." But the swings of public opinion continued, and in 1879 Gladstone, starting at Midlothian in Scotland, fought a nationwide political campaign of unprecedented excitement and drama. At the general election of April 1880 the Liberals returned to power triumphantly, with a majority of 137 over the Conservatives. Disraeli, who had moved to the House of Lords in 1876, died in 1881.

LATE VICTORIAN POLITICS

Gladstone and Chamberlain. Yet the second Gladstone administration (1880–85) did not live up to the promise of the election victory. Indeed, in terms of political logic, it seemed likely in 1880 that the Gladstonian Liberal Party would eventually split into Whig and radical components, the latter led by Joseph Chamberlain. This development was already foreshadowed in the Cabinet that Gladstone assembled, which was neither socially compact nor politically united. Eight of the 11 members were Whigs, but one of the other three—Chamberlain—was representative of a new and aggressive urban radicalism, less interested in orthodox statements of liberal individualism than in the uncertain aspirations and strivings of the different elements in the mass electorate. Already, as mayor of Birmingham (1873–76), he had embarked upon large-scale schemes of civic improvement, which he did not scruple to call "municipal socialism." The Whigs, at the opposite end of the spectrum, although the largest group in the Cabinet, were the smallest group in the country. Many of them were already abandoning the Liberal Party; all of them were nervous about the kind of radical program that Chamberlain and the newly founded National Liberal Federation (1877) were advocating and about the kind of caucus-based party organization that Chamberlain favoured locally and nationally. For the moment, however, Gladstone was the man of the hour, and Chamberlain himself conceded that he was indispensable.

The government carried a number of important reforms culminating in the Third Reform Bill of 1884 and the Redistribution Act of 1885. The former continued the trend toward universal male suffrage by giving the vote to agricultural labourers, thereby tripling the electorate, and the latter robbed 79 towns with populations under 15,000 of their separate representation. For the first time the franchise reforms ignored the traditional claims of property and wealth and rested firmly on the democratic principle that the vote ought to be given to people as a matter of right, not of expediency.

Third Reform Bill

The most difficult problems continued to arise in relation to foreign affairs and, above all, to Ireland. When in 1881 the Boers defeated the British at Majuba Hill and Gladstone abandoned the attempt to hold the Transvaal, there was considerable public criticism. And in the same year, when he agreed to the bombardment of Alexandria in a successful effort to break a nationalist revolt in Egypt, he lost the support of the aged radical Bright. In 1882 Egypt was occupied, thereby adding, against Gladstone's own inclinations, to British imperial commitments. A rebellion in the Sudan in 1885 led to the massacre of General Charles Gordon and his garrison at Khartoum two days before the arrival of a mission to relieve him. Large numbers of Englishmen held Gladstone personally responsible, and in June 1885 he resigned after a defeat on an amendment to the budget.

The Irish question. The Irish question loomed ominously as soon as Parliament assembled in 1880, for there was now an Irish nationalist group of more than 60 members led by Charles Stewart Parnell, most of them commit-

ted to Irish Home Rule; in Ireland itself, the Land League, founded in 1879, was struggling to destroy the power of the landlord. Parnell himself embarked on a program of agrarian agitation in 1881, at the same time that his followers at Westminster were engaged in various kinds of parliamentary obstructionism. Gladstone's response was an Irish Land Act, based on guaranteeing "three fs"—fair rents, fixity of tenure, and free sale—and a tightening up of the rules of closure in parliamentary debate. The Land Act did not go far enough to satisfy Parnell, who continued to make speeches couched in violent language, and after a coercion act was passed by Parliament in the face of Irish obstructionism, he was arrested. Parnell was released, however, in April 1882 after an understanding had been reached that he would abandon the land war and the government would abandon coercion. Lord Frederick Charles Cavendish, a close friend of Gladstone and brother of the Whig leader, Lord Hartington, was sent to Dublin as chief secretary on a mission of peace, but the whole policy was undermined when Cavendish, along with the permanent undersecretary, was murdered in Phoenix Park, Dublin, within a few hours of landing in Ireland.

The Phoenix Park murders

Between 1881 and 1885 Gladstone coupled a somewhat stiffer policy in Ireland with minor measures of reform, but in 1885, when the Conservatives returned to power under Salisbury, the Irish question forced itself to the forefront again. Henry Herbert, Earl of Carnarvon, the new lord lieutenant of Ireland, was a convert to Home Rule and followed a more liberal policy than his predecessor. At the subsequent general election of November 1885 Parnell secured every Irish seat but one outside Ulster and urged Irish voters in British constituencies—a large group mostly concentrated in a limited number of places like Lancashire and Clydeside—to vote Conservative. The result of the election was a Liberal majority of 86 over the Conservatives, almost exactly balanced by the Irish group, who thus controlled the balance of power in Parliament. The Conservatives stayed in office, but when in December 1885 the newspapers reported a confidential interview with Gladstone's son, in which he had stated (rightly) that his father had been converted to Home Rule, Salisbury made it clear that he himself was not a convert, and Carnarvon resigned. All Conservative contacts with Parnell ceased, and a few weeks later, in January 1886, after the Conservatives had been defeated in Parliament on a radical amendment for agrarian reform, Salisbury resigned and Gladstone returned to power.

Split of the Liberal Party. Gladstone's conversion had been gradual but profound, and it had more far-reaching political consequences for Britain than for Ireland. It immediately alienated him further from most of the Whigs and from a considerable number of radicals led by Joseph Chamberlain. He had hoped at first that Home Rule would be carried by an agreement between the parties, but Salisbury had no intention of imitating Peel. Gladstone made his intentions clear by appointing John Morley, a Home Rule advocate, as Irish secretary, and in April 1886 he introduced a Home Rule bill. The Liberals remained divided, and 93 of them united with the Conservatives to defeat the measure. Gladstone appealed to the country and was decisively beaten at the general election, in which 316 Conservatives were returned to Westminster along with 78 Liberal Unionists, the new name chosen by those Liberals who refused to back Home Rule. The Liberals mustered only 191, and there were 85 Irish nationalists. Whigs and radicals, who had often seemed likely to split Gladstone's 1880 government on left–right lines, were now united against the Gladstonians, and all attempts at Liberal reunion failed.

Chamberlain, the astute radical leader, like many others of his class and generation, ceased to regard social reform as a top priority and worked in harness with Hartington, his Whig counterpart. In 1895 they both joined a Salisbury government. The Liberals were, in effect, pushed into the wilderness, although they held office briefly and unhappily from 1892 to 1895. Gladstone, 82 years old when he formed his last government, actually succeeded in carrying a Home Rule bill in the Commons in 1893, with the help of Irish votes (Parnell's power had been broken as a result of a divorce case in 1890, and he died in 1891), but the bill was thrown out by the Lords. He resigned in 1894, to be succeeded by Archibald Primrose, Earl of Rosebery, who further split the party; at the general election of 1895, the Conservatives could claim that they were the genuinely popular party, backed by the urban as well as the rural electorate. Although Salisbury usually stressed the defensive aspects of Conservatism, both at home and abroad, Chamberlain and his supporters were able to mobilize considerable working-class as well as middle-class support for a policy of crusading imperialism.

Rise of the Conservatives

Imperialism and British politics. The word imperialism was the key word of the 1890s, just as Home Rule had been in the critical decade of the 1880s, and the cause of empire was associated not merely with the economic interests of businessmen looking for materials and markets and the enthusiasm of crowds excited by the adventure of empire but also with the traditional lustre of the crown. Disraeli had emphasized the last of these associations, just as Chamberlain emphasized the first. In the middle years of the century it had been widely held that colonies were burdens and that materials and markets were most effectively acquired through trade. Thus an "informal empire" had been created that was as much dependent on Britain as the formal empire. Nonetheless, even during these years, as a result of pressure from the periphery, the process of establishing protectorates or of acquiring colonies had never halted, despite a number of colonial crises and small colonial wars in Africa, Asia, and the Pacific. Most of the new acquisitions were located in tropical areas of the world and peopled mainly by non-Europeans.

There were further crises during the 1880s and '90s, when the Liberals were divided on both tactics and objectives, and public opinion was stirred. When Chamberlain chose to take over the Colonial Office in 1895, he was acknowledging the opportunities, both economic and political, afforded by a vast "undeveloped estate." The same radical energies that he had once devoted to civic improvement were now directed toward imperial problems. The argument about empire assumed an increasingly popular dimension. Boys' books and magazines, for example, focused on the adventure of empire and the courage and sense of duty of empire builders, and textbooks often taught the same lessons. So, also, did the popular press. In consequence the language of imperialism changed.

The problems of empire

In fact, however, it was difficult to pull the empire together politically or constitutionally. Certainly moving toward federation was a challenging task since the interests of different parts were already diverging, and in the last resort only British power—above all, sea power—held the empire together. The processes of imperial expansion were always complex, and there was neither one dominant theory of empire nor one single explanation of why it grew. White colonies, like Canada or New Zealand and the states of Australia, had been given substantial powers of self-government since the Durham Report of 1839 and the Canada Union Act of 1840. Yet India, "the brightest jewel in the British crown," was held not by consent but by conquest. The Indian "mutiny" of 1857 was suppressed, and a year later the East India Company was abolished and the new title of viceroy instituted. Imperial control was tightened, too, through the construction of a network of railways, Thomas Macaulay's dream that India would one day be free and that such a day would be the happiest in British history seemed to have receded, although the nationalist movement that emerged after the first Indian National Congress in 1885 was eventually to gain in strength. Meanwhile, given the strategic importance of India to the military establishment, attempts were made to justify British rule in terms of benefits of law and order said to accrue to Indians. "The white man's burden," as the poet Rudyard Kipling saw it, was a burden of responsibility.

It was difficult for the British voter to understand or to appreciate this network of motives and interests. Chamberlain himself was always far less interested in India than in the white "kith-and-kin dominions" and in the new tropical empire that was greatly extended in area between 1884 and 1896, when 2.5 million square miles of territory

fell under British control. Even he did not fully understand either the rival aspirations of different dominions or the relationship between economic development in the "formal" empire and trade and investment in the "informal" empire where the British flag did not fly.

Victoria's jubilees in 1887 and 1897 involved both imperial pageantry and imperial conferences, but, between 1896 and 1902, public interest in problems of empire was intensified not so much by pageantry as by crisis. British–Boer relations in South Africa, always tense, were further worsened after the Jameson raid of December 1895, and, in October 1899, war began. The early stages of the struggle were favourable to the Boers, and it was not until spring 1900 that superior British equipment began to count. British troops entered Pretoria in June 1900 and Paul Kruger, the Boer president, fled to Europe, where most governments had given him moral support against the British. Thereafter the Boers followed guerrilla tactics, and the war did not end until May 1902. It was the most expensive of all the 19th-century "little wars," with the British employing 450,000 troops, of whom 22,000 never returned. Just as the Crimean War had focused attention on "mismanagement," so the South African (Boer) War led to demands not only for greater "efficiency" but also for more enlightened social policies in relation to health and education.

The South African War

While the war lasted, it emphasized the political differences within the Liberal Party and consolidated Conservative–Unionist strength. Rosebery's Liberal imperialism was totally uncongenial to young pro-Boer Liberals like Lloyd George. A middle group of Liberals emerged, but it was not until after 1903 that party rifts were healed. The Unionists won the "khaki election" of 1900 and secured a new lease of power for nearly six years, but their unity also was threatened after the Peace of Vereeniging in May 1902. Salisbury retired in 1902, to be succeeded by his nephew, Arthur Balfour, a brilliant man but a tortuous and insecure politician. There had been an even bigger break in January 1901 when the queen died, after a brief illness, in her 82nd year. She had ruled for 64 years and her death seemed to mark not so much the end of a reign as the end of an age.

EDWARDIAN AND PREWAR BRITAIN

Victoria's successor, Edward VII, 59 years old, had never been on good terms with his mother, whose way of life was sharply different from his. He, too, gave his name to an age: flamboyant, ostentatious, at times vulgar and strident, with picturesque contrasts of fortune and circumstance. Yet the sharpness of the contrast between "Edwardian" and "late Victorian" should not be exaggerated. The last decade of the 19th century and the first decade of the 20th century had much in common, and there had been bigger breaks before in mood and preoccupation between the high-Victorian years and the 1890s.

Darwin's disciple, Thomas Huxley, an influential popularizer of science, had noted during the 1870s that everything was in question—opinions, institutions, and conventions—and the questioning thereafter never stopped. "The disintegration of opinion is so rapid," one writer put it in the 1880s, "that wise men and foolish are equally ignorant where the close of this waning century will find us." The writers of the last decades of the 19th century included iconoclasts like George Bernard Shaw and nonconformists like Oscar Wilde; for both, as for many others like them, all that was established was now suspect. Some commentators wrote of "a general revolt" against the accepted canons of the mid-century, a revolt influenced by thinkers outside Britain and challenging not only political or social assumptions (for example, about law and will or self-help and respectability) but also 19th-century culture as a whole, the culture of an industrialized society transformed through individual enterprise.

The economy. Changes in economic conditions during the last decades of the century were obviously of crucial importance. Mid-Victorian prosperity had reached its peak in a boom that collapsed in 1873. Thereafter, although national income continued to increase (nearly four times at constant prices between 1851 and 1911), there was a persistent pressure on profit margins, with a price fall that lasted until the mid-1890s. Contemporaries talked misleadingly of a "great depression," but however misleading the phrase was as a description of the movement of economic indexes, the period as a whole was one of doubt and tension. There was anxious concern about both markets and materials, and the fact that there was a retardation in the national rate of growth to below 2 percent per annum was even harder to bear when the growth rates of competitors were rising, sometimes in spectacular fashion.

Slowing of growth rate

The interests of different sections of the community diverged between 1870 and 1900 as they had diverged before the mid-Victorian period of equipoise. In particular, agrarian and meat-producing farmers felt the full weight of foreign competition in cereals, and many, though not all, industrialists felt the growing pressure of foreign competition in both old and new industries. As a result of improved transport, including storage and refrigeration facilities, and the application of improved agricultural machinery, overseas cereal producers fully penetrated the British market. In 1877 the price of English wheat stood at 56 shillings nine pence a quarter (1846: 54 shillings six pence); for the rest of the century it never again came within 10 shillings of that figure. During the 1890s, therefore, there was a sharp fall of rents, a shift in land ownership, and a challenge to the large estate in the cereal-growing and meat-producing areas of the country. The fact that dairy and fruit farmers flourished did not relieve the pessimism of most spokesmen of the threatened landed interests.

In industry, where there were new forms of power and a trend toward bigger plants and more impersonal organization, there were also moves throughout the period to increase cartels and amalgamations. Britain was never as strong or as innovatory in the age of steel as in the earlier age of iron—by 1896 British steel output was less than that of either the United States or Germany—while the textile industry was declining sharply. Exports fell between 1880 and 1900 from £105 million to £95 million. There were many explanations of what was happening—some concerned education; others were psychological as well as economic—but none of them was encouraging.

Yet the country's economic position would have been completely different had it not been for Britain's international economic strength as banker and financier. During years of economic challenge at home capital exports greatly increased, until they reached a figure of almost £200 million per annum before 1914, and investment income poured in to rectify adverse balances on visible trade accounts. Investing during these years in both "formal" and "informal" empire was more profitable, if more risky, than investing at home. But it also contributed to domestic obsolescence, particularly in the old industries. Thus ultimately there was a price to pay for imperial glory. During the last 20 years of peace before 1914, when Britain's role as rentier was at its height, international prices began to rise again, and they continued to rise, with fluctuations, until after the end of World War I. The City of London was at the centre of international markets of capital, money, and commodities.

The rise of labour. Meanwhile, whether prices were falling or rising, labour in Britain was increasingly discontented, more articulate, and more highly organized. Throughout the period national income per head grew faster than the continuing population growth (which stayed at above 10 percent per decade until 1911, although the birth rate fell sharply after 1900), but neither the growth of income nor the falling level of retail prices until the mid-1890s made for industrial peace. By the end of the century, when pressure on real wages was once again increasing, there were two million trade unionists in unskilled unions as well as in skilled unions of the mid-century type, and by 1914 the figure had doubled.

There were also significant political changes. Some of the new union leaders were confessed socialists, anxious to use political as well as economic power to secure their objectives, and a number of socialist organizations emerged between 1880 and 1900—all conscious, at least intermittently, that, whatever their differences, they were part of a "labour movement." The Social Democratic Federation,

Labour and politics

influenced by Marxism, was founded in 1884; it was never more than a tiny and increasingly sectarian organization. The Independent Labour Party, founded in Bradford in 1893, had a more general appeal, while the Fabian Society, founded in 1883–84, included intellectuals who were to play a large part in 20th-century labour politics. In February 1900 a labour representation conference was held in London at which trade unionists and socialists agreed to found a committee, with Ramsay MacDonald as first secretary, to promote the return of Labour members to Parliament. This conference marked the beginning of the 20th-century Labour Party, which, with Liberal support, won 29 seats at the general election of 1906. Although until 1914 the party at Westminster for the most part supported the Liberals, in 1909 it secured the allegiance of the "Lib-Lab" miners' members. Financially backed by the trade unions, it was eventually to take the place of the Liberal Party as the second party in the state.

The return of the Liberals. The Liberals returned to power in December 1905 after Balfour had resigned. Between the end of the South African War and this date they had become more united as the Conservatives had disintegrated. In 1903 Chamberlain had taken up the cause of protection, thereby disturbing an already uneasy balance within Balfour's cabinet. He failed to win large-scale middle-class or working-class support outside Parliament, as he had hoped, and the main effect of his propaganda was to draw rival groups of Liberals together. At the general election of 1906 the Liberals, led by Sir Henry Campbell-Bannerman, a cautious Scot who had stayed clear of the extreme factions during the South African War, won 377 seats, giving them an enormous majority of 84 over all other parties combined. The new Cabinet included radicals and Liberal imperialists, and when Campbell-Bannerman retired in 1908, H.H. Asquith moved from the Home Office to succeed him.

Social reform had not been the chief cry at the general election, which was fought mainly on the old issues of free trade, temperance reform, and education. In many constituencies there was evidence of Nonconformist grievances against a Balfour education act of 1902 that abolished the school boards, transferred educational responsibilities to the all-purpose local authorities, and laid the foundations of a national system of secondary education. Yet local and national inquiries, official and unofficial, into the incidence of poverty had pointed to the need for public action to relieve distress, and, from the start, the new Liberal government embarked upon a program of social legislation. In 1906 free school meals were made available to poor children; in 1907 a school medical service was founded; in 1908 a Children's Act was passed, along with an Old Age Pensions Act granting pensions under prescribed conditions to people over 70; in 1908 the miners were given a statutory working day of eight hours; and in 1909 trade boards were set up to fix wages in designated industries where there was little or no trade-union strength, and labour exchanges were created to try to reduce unemployment (a subject that was also being investigated locally and nationally) and to increase mobility. The vigour of these reforms owed much to a partnership between Winston Churchill at the Board of Trade and the "Welsh wizard," David Lloyd George, chancellor of the exchequer.

Social legislation

Lloyd George's budget of 1909 set out deliberately to raise money to "wage implacable warfare against poverty and squalidness." The money was to come in part from a supertax on high incomes and from capital gains on land sales. The budget so enraged Conservative opinion, inside and outside Parliament, that the Lords, already hostile to the trend of Liberal legislation, rejected it, thereby turning a political debate into a constitutional one concerning the powers of the House of Lords. Passions were as strong as they had been in 1831, yet, at the ensuing general election of January 1910, the Liberal majority was greatly reduced and the balance of power in Parliament was now held by Labour and Irish Nationalist members. The death of Edward in May 1910 and the succession of the politically inexperienced George V added to the confusion, and it proved impossible to reach agreement between the parties on the outlines of a Parliament bill to define or to curb the powers of the House of Lords. After a Liberal Parliament bill had been defeated, a second general election in December 1910 produced similar political results to those earlier in the year, and it was not until August 1911 that the peers eventually passed the Parliament Act by 131 votes to 114. The act provided that money bills could become law without the assent of the Lords and that other bills would also become law if they passed in the Commons and failed in the Lords three times within two years. The act was finally passed only after the Conservative leadership had repudiated the "diehard peers" who refused to be intimidated by a threat to create more peers.

The Parliament Act of 1911

In the course of the struggle over the Parliament bill, strong, even violent, feelings had been roused among lords who had seldom bothered hitherto to attend their house. Their intransigence provided a keynote to four years of equally fierce struggle on many other issues in the country, with different sectional groups turning to noisy direct action. The Liberals remained in power, carrying important new legislation, but they faced so much opposition from extremists, who cared little either about conventional political behaviour or the rule of law, that these years have been called by the American historian George Dangerfield "the strange death of Liberal England." The most important legislation was once more associated with Lloyd George—the National Insurance Act of 1911, which provided, on a contributory basis, for limited unemployment and health insurance for large sections of the population.

The National Insurance Act, which Parliament accepted without difficulty, was the subject of much hostile criticism in the press and was bitterly opposed by doctors and duchesses. Nor did it win unanimous support from labour. The parliamentary Labour Party itself mattered less during these years, however, than extraparliamentary trade-union protests, some of them violent in character—"a great upsurge of elemental forces." There was a wave of strikes in 1911 and 1912, some of them tinged with syndicalist ideology, all of them asserting, in difficult economic circumstances for the workingman, claims that had seldom been made before. Old-fashioned trade unionists were almost as unpopular with the rank and file as capitalists. In June 1914, less than two months before the outbreak of World War I, a "triple alliance" of transport workers, miners, and railwaymen was formed to buttress labour solidarity. In parallel to labour agitation, the suffragettes, fighting for women's rights, resorted to militant tactics that not only embarrassed Asquith's government but tested the whole local and national machinery for maintaining order. The Women's Social and Political Union, founded in 1903, was prepared to encourage illegal acts, including bombing and arson, which led to sharp police retaliation, severe sentences, harsh and controversial treatment in prison, and even martyrdom.

The issue that created the greatest difficulties, however, was one of the oldest: Ireland. In April 1912, armed with the new powers of the Parliament Act, Asquith introduced a new Home Rule bill. Conservative opposition to it was reinforced on this occasion by a popular Protestant movement in Ulster; and the new Conservative leader, Andrew Bonar Law, who had replaced Balfour in 1911, gave his covert support to army mutineers in Ulster. No compromises were acceptable, and the struggle to settle the fate of Ireland was still in full spate when war broke out in August 1914. Most ominously for the Liberals, the Irish Home Rule supporters at Westminster were losing ground in southern Ireland, where in 1913 a militant working-class movement entered into close alliance with the nationalist forces of Sinn Féin. Ireland was obviously on the brink of civil war.

Irish difficulties

The international crisis. The seeds of international war, sown long before 1900, were nourished between the resignation of Salisbury in 1902 and August 1914. Two intricate systems of agreements and alliances—the Triple Alliance of Germany, Austria-Hungary, and Italy and the Triple Entente of France, Russia, and Britain—faced each other in 1914. Both were backed by a military and naval apparatus (Britain had been building a large fleet, and Richard Haldane had been reforming the army), and

both could appeal to half-informed or uninformed public opinion. The result was that a war that was to break the continuities of history started as a popular war.

The Liberal government under Asquith faced a number of diplomatic crises from 1908 onward. Throughout a period of recurring tension, its foreign minister, Sir Edward Grey, often making decisions that were not discussed by the Cabinet as a whole, strengthened the understanding with France that had been initiated by his Conservative predecessor in 1903. An alliance had already been signed with Japan in 1902, and in 1907 agreements were reached with Russia. Meanwhile, naval rivalry with Germany familiarized Englishmen with the notion that, if war came, it would be with Germany. The 1914 crisis began in the Balkans, where the heir to the Austro-Hungarian throne was assassinated in June 1914. Soon Austria, backed by Germany, and Russia, supported by France, were arrayed against each other. The British Cabinet was divided, but after the Germans invaded Belgium on August 4, thereby violating a neutrality that Britain was committed by treaty to support, Britain and Germany went to war. (As.B.)

Invasion of Belgium

Britain from 1914 to the present

WORLD WAR I

The British declaration of war on Germany on Aug. 4, 1914, brought an end to the threat of civil war in Ireland, which since March had occupied Prime Minister H.H. Asquith's Liberal Cabinet almost to the exclusion of everything else. Formally at least, party warfare came to an end. The Conservatives agreed not to contest by-elections and to support the government in matters pertaining to the war.

The Asquith coalition. Such compromises were easy to make in the autumn of 1914 when the excitement over the outbreak of war was high, causing a crush of enlistments, and when it was still generally believed that the war would be over within six months. By the spring of 1915, however, enthusiasm for the war began to cool and recruiting fell off. Moreover, Asquith's government seemed to have lost its grip on affairs; newspapers carried reports of an inadequate supply of ammunition on the Western Front, and on May 15 the first sea lord, Admiral John, Lord Fisher, resigned. The Conservative leader, Andrew Bonar Law, under pressure from his followers to take a stronger stand, announced that the party would demand a debate on the conduct of the war. Asquith quickly offered forming a coalition, thereby ending the last Liberal government. The coalition consisted of Liberals, Conservatives, and one Labourite.

In the new Cabinet, announced on May 25, Arthur James Balfour replaced Winston Churchill as first lord of the admiralty. More important, a new department, the ministry of munitions, was established with the Liberal David Lloyd George at its head.

The coalition, supposed to allay tension among parties over the conduct of the war, worked badly. Although the ministry of munitions did indeed resolve the armament crisis surprisingly quickly, dissatisfaction with Asquith's relaxed management of affairs continued and centred in the autumn of 1915 upon the rising demand, in the press and among the Conservatives, for compulsory military service. With apparent reluctance the prime minister allowed an inadequate measure for the conscription of unmarried men to be passed in January 1916. But it was not until May 1916, after more controversy and threats of resignation, that a comprehensive bill passed for compulsory enlistment of all men between the ages of 18 and 41.

The Easter Rising

Meanwhile, on April 24, 1916, Monday of Easter Week, not unconnected with the approaching conscription, a rebellion broke out in Dublin directed at securing Irish independence. Violence was suppressed within six days, and the surviving rebels were arrested amid general derision from the Irish population. But Britain's punishment of the rebels, including 14 summary executions, quickly turned Irish sympathy toward the men, who were now regarded as martyrs. The Easter Rising was the beginning of the Irish war for independence.

Even though the rebellion was quelled, the problems of Ireland needed to be addressed. Prime Minister Asquith called upon Lloyd George to try to arrange for an immediate grant of Home Rule between the Irish Nationalist and Unionist parties (the former being fully committed to the principle of Home Rule, the latter only partially). Although a compromise was in fact reached, discontent among senior Unionists prevented a bill from going forward. Thereafter Home Rule ceased to be an issue because southern Ireland now wanted nothing but independence. Asquith was further weakened.

The government also drew criticism for its war policies. For one, Britain was unable to help Romania when it declared war upon the Central Powers in the summer of 1916. More significantly, Britain launched its first major independent military operation, the Battle of the Somme (July 1 to Nov. 13, 1916), with disastrous results. On the first day of battle the British suffered almost 60,000 casualties. Although little of strategic significance was accomplished, the battle brought the reality of war home to Britain. (For details on the military aspects of World Wars I and II, see the article WORLD WARS.) Dissatisfaction with the government mounted, until in the first week of December Asquith and most of the senior Liberal ministers were forced to resign. Lloyd George became prime minister with a Cabinet consisting largely of Conservatives.

The Battle of the Somme

Lloyd George. Lloyd George governed Britain with a small "War Cabinet" of five permanent members, only one of whom was a politician of standing. Although Lloyd George had to take note of the opinions of Parliament and of the men around him and to pay attention to the tides of public political sentiment, the power of making decisions rested entirely with him. He was faced with the same sentiments of apathy, discontent with the country's leadership, and war weariness that had brought down the Asquith government. Not only had Britain's supreme military effort in 1916 failed, but the war had lost its meaning. Belgium was forgotten, still more Serbia. Thus, in the next two years, Lloyd George set out to reinvest the war with meaning. Its purpose would be to create a better Britain and a safer world. Victory promised hope for the future. Toward that goal he established new ministries and brought workingmen into government. Lloyd George's reconstruction program was built on principles that were later enunciated by Woodrow Wilson in his Fourteen Points and his slogan of making the world safe for democracy. Lloyd George's own slogan of 1918 was to forge a nation fit for heroes to live in.

Lloyd George controlled the government but not the Liberal Party; only about 100 Liberals in the House of Commons supported him, the rest remaining loyal to Asquith. Worse, Lloyd George had no party organization in the country. The division within the Liberal Party hardened during the controversy over a statement made by him in April 1918 concerning the strength of troops in France. Although this controversy, the so-called Maurice Debate on May 9, strengthened Lloyd George temporarily, it also made clear his dependence upon the Conservatives. Soon afterward, in the summer of 1918, he began to plan what he expected to be a wartime general election to be entered into in coalition with the Conservatives. The sudden armistice of Nov. 11, 1918, however, intervened and the wartime election became a victory election. Meanwhile the Labour Party had withdrawn its support from the coalition and called upon Labour members to resign. Most, but not all, did.

The election of 1918. The general election of Dec. 14, 1918, was a landmark in 20th-century British history and may have helped to set the course of politics through the interwar period. To begin, the Representation of the People Act of 1918, which gave the vote to all men over 21 and all women over 30 and removed the property disqualifications of the older household franchise, tripled the electorate. Ironically, the election registered the lowest voter turnout of any election in the 20th century. Further, 37 seats were added to the House of Commons.

Even though the coalition was returned to office, the real winners of the election were the Conservatives. Lloyd George's Liberals and the Conservatives, who had arranged not to contest seats against each other, together

won 478 of the 707 seats. Even though Lloyd George had the support of 133 Liberals in the House of Commons, the Asquithian Liberal Party was nearly wiped out, returning only 28 members as opposed to the Labour Party's 63. (Similarly, the old Irish Nationalist Party was destroyed and replaced by Sinn Féin, the party of independence.) Thus, despite the coalition's overwhelming victory, Lloyd George remained dependent on the Conservatives. The Liberal organization in the country was in shambles.

Finally the election had focused not upon the reconstruction of Britain, as the leaders of each party had intended, but on the punishment of Germany after the war, a matter the government had hoped to defer. The election had committed the British government to a harsh peace.

BETWEEN THE WARS

Economic and social developments. Although Britain suffered far less physical damage than France and, unlike Italy, underwent no political revolution, World War I may have affected Britain more fundamentally than any other of the Western allies, and far more, it would appear, than did World War II. It had provided the occasion for massive governmental experiments in economic enterprise, in insurance, in the management of the railroads and coal mines, and above all in the huge ministry of munitions. The government had become directly involved in the manufacture of every conceivable weapon of war, from airplanes to steel helmets, while embarking also on important social experiments in the provision of accommodations, recreation, and medical care for hundreds of thousands of war workers, both men and women. These endeavours, for the most part highly successful, were commented upon and remembered after the war when it was argued that a nation that could build guns could also build houses and schools and that the ideal employer was the government. Thus the war, first of all, was a massive catalyst for social and economic change.

Second, the war blurred the frontiers between the classes and the sexes. Although the comradeship of the trenches may not have been as effective in reducing the chasm between gentleman and common man as once thought, the undeniably enhanced economic position of the factory worker caused by the demand for labour and the growing prosperity and political power of trade unions had the same effect. Whether a new bond of affection had developed between the common man and the gentleman is unclear; in any case, the distinction between them was old-fashioned. But it was quite clear that the worker feared his employer far less and liked him no more than before the war and that the old deferential Britain was nearly gone.

Emancipation of women

The war also changed the position of women, bringing political, and to some extent economic and social, emancipation. With the outbreak of war the woman's suffrage movement had turned its attention wholeheartedly to the military effort. Large numbers of women were employed by the ministry of munitions, smaller numbers by private armament makers against serious opposition by unions, and still fewer in government and private offices.

Economically Britain had been hurt severely. The huge balances of credit in foreign currencies that had provided the capital for the City of London's financial operations for a century were spent. Britain had moved from the position of a creditor to that of a debtor nation. Moreover, its industrial plant, already out of date at the start of the war, had been allowed to depreciate and decay further. The industries of the Industrial Revolution, such as coal mining, textile production, and shipbuilding, upon which British prosperity had been built, were now either weakened or redundant. The Japanese had usurped the textile export market. Coal was superseded by other forms of energy. Shipping lost during the war had been almost fully replaced with more modern and more efficient vessels; in addition, with central Europe suffering severely from war damage, the need for transportation declined.

Finally, the Treaty of Versailles, particularly its harsh demands on Germany for financial reparations, insured that foreign markets would remain depressed. Germany had been Britain's largest foreign customer. The export of German coal to France, as stipulated by the treaty, upset world coal markets for nearly a decade. Depression and unemployment, not prosperity and a better Britain, characterized the interwar years.

Versailles reconsidered

The peace treaty with Germany, drawn up far too rapidly, without German participation, between January and May 1919, went into effect on June 28. Even as peace with Germany was declared, the British nation, as well as members of the government, was beginning to realize that the punitive treaty, burdening Germany with the responsibility and much of the cost of the war, was a mistake. Accordingly British foreign policy for much of the decade of the 1920s aimed at rehabilitating Germany and bringing it back into the family of nations. In general, this attempt was opposed by France and resulted in a rupture between Britain and its wartime ally, forcing France into a position of isolation that would have prodigious consequences for Europe and indeed for the rest of the world with the rise of Adolf Hitler in the early 1930s.

Lloyd George spent a great deal of time in the four postwar years of his administration on foreign affairs. As a consequence problems within the United Kingdom, such as unemployment, poor housing, Irish separatism, and the revival of industry, were too frequently neglected. Many of the promises for reconstruction made in speeches and papers during the war were never carried out. The government, however, tried to diminish the habitual confrontation between newly powerful organized labour and industry. Unemployment insurance was extended to virtually all workers, and a serious attempt was made to begin a public housing program. Railroads were reorganized, and for three years after the war coal mines remained in public hands. This restructuring of industry, however, came to an end with the serious rise in unemployment that began in 1920 and culminated in 1921 in a full-scale industrial depression with nearly a quarter of the labour force out of work. One of the factors in the depression was a disastrous coal strike in April 1921, caused in a considerable measure by the collapse of world coal prices resulting from German coal reparations to France. The immediate effect of the economic depression was a demand by the Conservatives for government economy that the prime minister could not ignore.

Ireland and the return of the Conservatives. In 1919 revolutionary disorder broke out in the south of Ireland when the provisional government of Ireland, organized by the Sinn Féin party, began guerrilla military operations against the British administration. Through 1920 the British government attempted to put down violence with violence, while passing an act allowing Home Rule to both the south of Ireland and to Ulster. The six Protestant Unionist counties of the north accepted Home Rule and in 1921 set up in Belfast an autonomous government. In the 26 counties of the south, Home Rule was defiantly rejected. By the spring of 1921, however, with the Belfast government in operation and with demands both in Britain and in the rest of the world that the fighting in Ireland come to an end, compromise became possible. In the summer a truce was arranged, and on Dec. 6, 1921, after prolonged negotiations, the British government and the Irish rebels signed a so-called treaty allowing the establishment of what was, in effect, a dominion government in Dublin.

Home Rule for Northern Ireland

Lloyd George's insistence that the Irish be granted the substance, if not the letter, of their demands as well as the clearly declining popularity of the coalition government with which they were associated caused general unhappiness, not among the Conservative leadership but among the members of the Conservative back bench in the House of Commons. Finally, in October 1922, when the proposal to join forces in a second coalition election was decisively rejected, largely by the Conservative rank and file, the Conservative Party withdrew from the coalition. Lloyd George resigned on October 20, and George V invited the Conservative leader, Andrew Bonar Law, to form a government. On Nov. 15, 1922, the hastily established Conservative government won a solid victory in a general election. The decline of the Liberal Party was confirmed by the fact that the two wings of the party together returned only 116 members of Parliament to Labour's 142.

The Baldwin era. Law remained prime minister only until May 20, 1923, when, ill with cancer, he resigned. He was succeeded by an almost unknown politician, Stanley Baldwin, who would nonetheless dominate British politics until his resignation from his third government in May 1937. Baldwin seemed an unlikely leader for a major party; he had been in Parliament for 15 years without making a mark. Yet behind the unassuming demeanour was a crafty politician. Baldwin understood, as perhaps his predecessors had not, that the British voter, certainly the middle-class voter, desired not excitement and reform but tranquillity. Nostalgia for the assumed stability of prewar Britain was strong and indeed a key to the politics of the 1920s. This frame of mind would contrast sharply with Britain's mood after World War II.

The new Conservative government was faced with high unemployment, industrial stagnation, foreign debts, and continuing demand for economy in government. Baldwin's response was to abandon Britain's historic policy of free trade and to return to import duties. Although he was supported in this by a majority of his party, he nonetheless promised to hold an election on the subject before implementing such a policy. Consequently, on Dec. 6, 1923, a second election was held in which the comfortable Conservative majority of 345 was reduced to 258, with the now united Liberal Party electing 159 and Labour 191. As a result, on Jan. 22, 1924, the first Labour government in British history under Prime Minister James Ramsay MacDonald came to power with Liberal support.

James Ramsay MacDonald

MacDonald remained in office only nine months and accomplished little except the revival of the public housing program abandoned by Conservative pressure in the Lloyd George administration. During his time in office he was continually charged in the House of Commons and in the newspapers with unseemly weakness toward the Bolshevik government of the Soviet Union and with an unwillingness to deal firmly with purported revolutionary socialist conspiracies within the United Kingdom. Over this matter the Liberals finally turned against him and on Oct. 29, 1924, in an election dominated by charges of Soviet influence, MacDonald was heavily defeated. Stanley Baldwin returned to the prime ministership, backed by a more than two-to-one majority over Labour and the Liberals combined. The Liberal representation in the House of Commons was reduced to 40.

Baldwin's return to office coincided with the French evacuation of the Ruhr Valley in Germany and the revival of Germany as an economic power. Partly as a result Britain in the nearly five years of the second Baldwin government experienced a relative economic prosperity, although unemployment never went below 10 percent of the working population covered by unemployment insurance. A new collapse in domestic coal prices, however, caused by the revival of German coal mining, produced the threat of a second strike in British coal. It erupted in May 1926 with a walkout in the coal industry and a sympathy strike in the rest of Britain's organized labour. Except as a monument in the history of British labour, however, this so-called General Strike is as unimportant as it was unsuccessful. As a general strike, it lasted only nine days, from May 3 to May 12. The miners themselves held out for nearly eight months and were finally starved into returning as winter began, at lower wages and with longer hours. Economically, the chief effect of the strike was to hasten the decay of the huge British coal industry. However, Baldwin's handling of it—he prepared emergency services but then did nothing—greatly increased his popularity; indeed, he is remembered as a peacemaker, although his government passed an act declaring general strikes to be revolutionary and hence illegal. Yet beyond that his administration, particularly the ministry of health under Neville Chamberlain, accomplished a good deal; it vastly extended old-age pensions and pensions for widows and orphans, reformed local government, and finally, in 1928, extended the franchise to women aged 21 to 30 on the same terms as those for men.

Baldwin dissolved the House of Commons in the spring of 1929, expecting to be returned. Instead on May 30 MacDonald's Labour Party received 288 seats compared to the Conservative Party's 260, with the Liberals holding again, with 59 seats, the balance of power. Thus MacDonald formed his second government, again with Liberal consent, if not support. The Liberals could do little else. In 1924 Labour, by its inaction, had proved itself as a responsible rather than revolutionary party. In the minds of Britons Labour had replaced the Liberals as the natural alternative party.

The 1920s and early '30s have come to be known in Britain as the Baldwin Age, with the "flapper"—the emancipated young woman—as its symbol. Perhaps the reality behind this symbol was the women's vote and women's entry into the economic world made possible by fundamental changes in England's economy. Generally, the economic centre of England was shifting from the north to the south, reversing the 19th-century trend from south to north. The old industries of the north, such as steel, shipbuilding, coals, and machine building, were in depression. In the Midlands appeared electrical manufacturing and automobile industries. In the south, in addition to construction industries, new service industries such as hotels and the shops of London flourished. These in particular offered employment opportunities for women at a time when the demand for domestic servants was in decline. London grew enormously, and the unemployment rate there was half that of the north of England and of Wales, Scotland, and Northern Ireland. The effect of these developments was to divide Britain politically and economically into two areas, a division that with the exception of an interval during World War II and its immediate aftermath still exists.

The Baldwin Age

The Great Depression and its aftermath. An era in international history ended with MacDonald's accession in 1929. Within months the British economy, as well as that of the rest of the world, was devastated by the Great Depression. The postwar world of reconstruction became a prewar world of deep depression, radicalism, racism, and violence. MacDonald, a well-meaning, highly intelligent, but uneducated Scottish cottage boy, was badly equipped to handle the depression. By the end of 1930 unemployment was nearly double the figure of 1928 and would reach 25 percent of the work force by the spring of 1931. It was accompanied, after the closing of banks in Germany in May, by a devastating run on gold in British banks that threatened the stability of the pound.

MacDonald's government fell in August over the protection of the pound: Britain needed to borrow gold, but foreign bankers would lend gold only on the condition that domestic expenditures would be cut, and this meant, among other things, reducing unemployment insurance payments. A Labour Party, however, whose central commitment was to the welfare of the working people, could not mandate such a course of action even in an economic crisis. Thus the Labour Cabinet resigned. MacDonald with a few colleagues formed a coalition with the Conservative and Liberal opposition on Aug. 24, 1931. This new "national" government, which allowed Britain to go off the gold standard on September 21, was confirmed in office by a general election on October 27, in which 473 Conservatives were returned, while the Labour Party in the House of Commons was nearly destroyed, with only 52 seats. MacDonald nonetheless, with 13 so-called National Labour colleagues, remained prime minister. The new government was in fact a conservative government, and MacDonald, by consenting to remain prime minister, became and remains in Labour histories a traitor.

Under Neville Chamberlain, who became chancellor of the exchequer in November 1931, the coalition government pursued a policy of strict economy. Housing subsidies were cut; Britain ended its three-quarter-century devotion to free trade and began import protection; and interest rates were lowered. Manufacturing revived, stimulated particularly by a marked revival in the construction of private housing made possible by reduced interest rates and by a modest growth in exports as a result of the cheaper pound. Similarly, unemployment declined, although it never reached the 10 percent level of the late 1920s until after the outbreak of war.

Baldwin and the abdication crisis. In June 1935, Bald-

win rather abruptly took over the prime ministership from MacDonald, whose health was clearly failing. A general election followed on November 14, in which the Conservatives returned 432 members to Parliament to Labour's 154. But because the so-called National Liberals and a few remaining National Labour members still participated in the government, it was technically a coalition. This election was the last British general election for nearly 10 years. Hence Baldwin in his final 18 months of office presided over the beginnings of Britain's appeasement policy and over the more spectacular but less important abdication of the new king, Edward VIII, who ascended the throne on Jan. 20, 1936, upon the death of his father, George V.

In the quarter century since his father's accession, Edward as Prince of Wales had become the most public and best-known heir to the throne since his grandfather, Edward VII. But, unknown to the British public, some years before his accession he had fallen in love with an American divorcée, Wallis Simpson, who was then married to a British subject, Ernest Simpson. Edward determined to marry her, and in 1936, after his accession, Wallis Simpson began divorce proceedings against her husband. Baldwin, well before his actual confrontations with the king, had determined that Edward could not marry Mrs. Simpson and remain monarch. He warned him not to attempt to influence public opinion or to try to remain on the throne. The temper of the people and of Parliament was against him. Eventually, on Dec. 11, 1936, Edward announced his abdication in a poignant radio broadcast and left Great Britain. Baldwin had triumphed. The king was succeeded by his younger brother, who became George VI and who had an eminently suitable family, including two young daughters. After George VI's coronation on May 12, 1937, Baldwin resigned, amid every sign of popular affection; he was succeeded on May 28 by Neville Chamberlain.

The abdication crisis

Foreign policy and appeasement. Chamberlain, rather than Baldwin, has always been regarded as the man of appeasement. Historically this is correct only in the sense that Chamberlain formulated a policy of accommodation with Germany and Italy. But Chamberlain was also the man who began British rearmament, pronounced appeasement a failure, and declared war upon Germany. Baldwin was equally zealous to avoid any sort of confrontation with the European dictators while doing as little as possible to strengthen Britain's armed forces.

Hitler's accession to power in Germany on Jan. 30, 1933, occasioned only the slightest interest in Britain. Little was known of him. It was usually assumed that he was a tool of the right or the army and in any case would not remain in office long. This illusion began to be shattered in January 1935 when Germany overwhelmingly won a plebiscite in the Saar basin; the Saarlanders voted to return their area to Germany, from which it had been separated by the Treaty of Versailles as part of German reparations, rather than remaining with France. This was an enormous boost to Hitler's prestige as well as a confirmation of the attraction of Nazi Germany and, by the same token, a setback for France and the idea of democracy.

On the wave of popularity the plebiscite brought, Hitler reintroduced military conscription in Germany and announced the creation of the Luftwaffe, both in violation of the Treaty of Versailles. In response the former wartime allies and guarantors of the peace treaty, Britain, France, and Italy, met at Stresa in Italy in April and there discussed collective action to uphold the disarmament terms of the treaty; this became the "Stresa front." Its maintenance, specifically the challenge of keeping Italy a foe of Germany, formed the motivation for Britain's foreign policy for the next 18 months; in effect it was the beginnings of appeasement. In August 1935 Italy attacked the empire of Ethiopia in Africa, announcing that it had apprised Britain and France at Stresa of its intentions of doing so. British public opinion was torn between a desire to avoid war and an unwillingness to sanction unprovoked aggression. The compromise was a retreat to the fiction of "collective security," which meant a dependence upon action by the League of Nations at Geneva. Support for the League of Nations became the Conservative position on foreign policy in the general election of November 1935.

The Stresa front

Britain at this time was interested in pursuing friendship with Italy. Immediately after the election the British foreign secretary, Sir Samuel Hoare, and the French premier, Pierre Laval, put together a plan for the rescue of part of Ethiopia that required the cession of certain areas to Italy. This plan found its way into the press, provoking a general denunciation of compromise with evil. Hoare had to resign, and the first attempt at appeasement failed. By the spring of 1936, with the League of Nations still debating what to do about Italian aggression—specifically, whether to impose sanctions on oil—resistance in Ethiopia collapsed. Meanwhile, on March 7, Hitler took advantage of the disarray in the west and broke the first of the territorial clauses of the Treaty of Versailles by sending troops into the Rhineland, the German territory to the west of the Rhine River bordering on Belgium and Holland.

The Rhineland occupation turned the balance toward Germany and against the west. Although in Britain there was virtually no reaction—after all it was German territory—the effect on France, and particularly on the French army command, was devastating. As a consequence France virtually gave up the direction of its foreign affairs. Diplomatic initiative rested entirely in London. Now that it was too late, the 15-year rupture between Britain and France came to an end.

In July 1936 revolution against the Republican government of Spain broke out, led by conservative forces within the Spanish army under the command of General Francisco Franco. It quickly became apparent that the revolutionaries were supported by Italy, and to a lesser extent Germany, not only with money and arms but also with men. The British reaction, adopted also by the French, was peculiar. Although, according to public opinion polls begun in 1937, less than 3 percent of the British population favoured a Francoist victory, British policy was to forbid the supply of arms to either side. By this policy of nonintervention the British and the French avoided involvement in war against Franco and by implication against the Italian government. The pursuit of friendship with Italy could continue. Meanwhile the democratic Spanish government was unable to buy arms from the Western democracies. Franco finally triumphed in the spring of 1939.

Chamberlain was determined to continue the policy of accommodation with Italy. He was convinced that at some point it could be reunited with the Western allies and the Stresa front be recreated. Mussolini and his officials gave many private intimations that this might be possible. But at the same time Chamberlain determined to pursue a general policy of European settlement that would include Germany. The prime minister, and many Britons, felt that Germany had been badly treated by the Treaty of Versailles and that the principle of self-determination dictated that German minorities in other countries should not be prevented from joining Germany if they clearly chose to do so. Hence, when Germany overran the Austrian republic in March 1938 and incorporated the small state into the Reich, Britain took no action. Similarly, when almost immediately Hitler began to denounce what he styled the Czech persecutions of the militant German minority in the Sudetenland of Czechoslovakia, Chamberlain searched for a means, not to prevent the Czech borderland from being transferred to Germany, but to ensure that it was accomplished peacefully. Because Czechoslovakia had a military alliance with France, war would surely result if it resisted the Germans and called upon French aid.

Chamberlain's policy of appeasement

The attempted settlement of the Sudeten crisis, culminating in the so-called Munich agreement, was the climax of the appeasement policy. Between Sept. 15 and 29, 1938, Chamberlain traveled to Germany three times to meet Hitler. From the last meeting, held at Munich on September 30, he took back what he believed to be an agreement that the German portions of Czechoslovakia constituted Hitler's last territorial claim in Europe and that Germany, as well as Britain, would renounce war as a means of settling international claims. He had, he said with some pride, brought "peace for our time."

Chamberlain's policy failed because he had believed that Hitler sincerely aimed only at reuniting Germans, whereas in fact Hitler's appetite for territory, particularly

to the east, was unlimited. On March 15, 1939, the German army, virtually without warning, occupied the rest of Czechoslovakia even though it was not inhabited by Germans. On March 18 Chamberlain, distinctly angry, made an announcement that amounted to the end of appeasement; in the following weeks Britain offered a guarantee of Polish territory (where Hitler would clearly be looking next), signed a military alliance with Poland, and undertook serious preparation for war, including the first peacetime military conscription.

WORLD WAR II

The Polish crisis precipitated the war. Through the summer of 1939 German propaganda grew more strident, demanding cession to Germany of the city of Gdańsk (Danzig) while gradually escalating demands for special rights in, and finally annexation of, the Polish corridor. Because the only nation able to defend Poland was the Soviet Union, a British–French mission in the summer of 1939 began negotiations for a treaty with Joseph Stalin. Poland, however, announced that it would not allow Soviet troops to enter Polish territory, even for the purpose of defending the country against Germany. Hitler put a stop to these negotiations on August 23 when he announced a nonaggression pact with the Soviet Union. On September 1 German troops invaded Poland. Britain and France declared war on Germany on September 3.

The outbreak of war

The phases of war. From the British perspective World War II fell readily into three distinct phases: the first, the so-called phony war and the period of German victories in the west, ended with the decision of France on June 18, 1940, to ask for an armistice with Germany. The second, heroic phase, when Britain stood alone, began with the battle for survival in the air over the British Isles and ended in the first week of December 1941 with the successful Soviet defense of Moscow after Hitler's attack on June 22 and with the Japanese declaration of war on the United States and the British Empire on December 7. Then followed what Churchill termed the period of the Grand Alliance, lasting from December 1941 until Germany's capitulation in May 1945.

Perhaps the most important event of the first phase was the announcement on Sept. 3, 1939, that Churchill, assumed to have reached the end of his career in 1936 as a result of his having embraced the king's cause during the abdication crisis, would reenter the government as first lord of the admiralty. Churchill thus was in charge of the Royal Navy on April 9 and 10, 1940, when Hitler without warning overran Denmark and Norway, greatly extending his northern flank and virtually destroying the naval blockade of Germany that had been established at the beginning of the war.

The Norwegian campaign destroyed also the Chamberlain government. The obviously poor planning and the incapacity of the British forces in an area where the Germans were at a serious disadvantage caused a rebellion within the Conservative Party. A bitter debate lasting from May 7 to May 9, 1940, resulted in Chamberlain's resignation the next day. Although Churchill himself, as first lord of the admiralty, was heavily involved and did not attempt to deny his responsibility, Chamberlain quickly discovered that the coalition government he hoped to establish with either himself or Lord Halifax as prime minister, could, at the insistence of the Labour Party, be headed only by Churchill. Thus, on May 10 Churchill was announced as prime minister. Chamberlain, to his immense credit, consented to remain in the Cabinet and to control, on Churchill's behalf, the Conservative Party.

On the same day, May 10, 1940, the German army struck in the west against The Netherlands, Belgium, and Luxembourg. France held out for just 38 days. When on June 18 the French government resolved to ask for an armistice, Churchill announced on the radio that Britain would fight on alone; it would be the nation's "finest hour." So began the second phase of World War II for Britain. Through August and September 1940 the fate of the nation depended upon 800 fighter airplanes, and upon Churchill's resolution, in the terrific bombardment that became the Battle of Britain. In the last six months of 1940 some 23,000 civilians were killed, and yet the nation held on.

The Battle of Britain

Perhaps the important political lesson of World War II lay in the realization that a democratic nation, with a centuries-old tradition of individual liberty, could with popular consent be mobilized for a gigantic national effort. The compulsory employment of labour became universal, for both men and women. In 1943 Britain was devoting 54 percent of its gross national product to the war. Medical services were vastly extended. Civilian consumption was reduced to 80 percent of the prewar level. Yet by and large the political tensions that had accompanied an equally desperate war 25 years before did not appear. Britain was unified in a way it had seldom been. Politics, as opposed to the direction of the war, certainly for the voters, became almost irrelevant. There was some parliamentary criticism of Churchill's leadership, but public approval, measured by repeated opinion polls, hardly wavered.

Political developments. The one political event of significance before 1945 was the publication of a report in November 1942 by Sir William Beveridge entitled *Social Insurance and Allied Services*. It contained a plan for the unification of Britain's welfare schemes, basically the existing schemes of national health insurance, unemployment insurance, and old-age pensions, with a new, parallel but separate plan for the insurance of industrial injuries. The "Beveridge Report" was at once immensely popular among the people at large and highly controversial within Churchill's government. The government had withdrawn its sponsorship of Beveridge's investigation when it discovered that he was preparing a full-scale overhaul of Britain's welfare services, and so the report appeared solely as the proposals of Beveridge himself. The government, nonetheless, distributed the report widely as a propaganda tool without, awkwardly, making any commitment to implement it. The report would become an issue in the general election of 1945.

The Beveridge Report

German hostilities in the west ended at midnight on May 8, 1945. Six months earlier Churchill had promised in the House of Commons that he would ask the king to dissolve the sitting Parliament, elected in 1935, soon after the German surrender unless the Labour and Liberal parties seriously desired to continue the coalition government. Accordingly, he began conversations with Clement Attlee, the leader of the Labour Party, in the middle of May, proposing that Labour remain in the coalition until Japan surrendered, an event he estimated to be at least 18 months away. Churchill believed Attlee to have been initially sympathetic, but other members of the Labour Party pressed for departure. As a result Churchill dissolved the government on May 23, appointed a new, single-party Conservative government, and set election day for July 5. Because it was necessary to count the military vote, the results could not be announced until July 26.

Considering that the leading figures in each party had been Cabinet colleagues only a few weeks before, the electoral campaign was remarkably bitter. Largely on the advice of William Maxwell Aitken, Baron of Beaverbrook, the Conservatives focused chiefly on Churchill himself as the man who had won the war. Churchill denounced Labour as the party of socialism and perhaps of totalitarianism while promising strong leadership and grand but unspecific measures of social reform. Labour, even though the war in the Pacific continued, concentrated on peacetime reconstruction and fair shares for all.

Quite clearly Churchill's rhetoric, his attacks on former comrades, angered many voters. But the mood in the country that gave Labour its overwhelming victory was obviously determined by the recollection of the hardships of the 1920s and 1930s; Britons voted against Stanley Baldwin and Neville Chamberlain. In the end Labour won 393 seats, almost double the Conservative total of 213 and far more than it had expected. On July 26, 1945, as soon as the results were clear, Churchill resigned and Attlee became prime minister.

BRITAIN SINCE 1945

Labour and the welfare state (1945–51). Labour rejoiced at its political triumph, the first independent par-

liamentary majority in the party's history, but it faced grave problems. The war had stripped Britain of virtually all its foreign financial resources, and the nation had built up "sterling credits"—debts owed to other countries that would have to be paid in foreign currencies—amounting to several billion pounds. Moreover, the economy was in disarray. Some industries, for example aircraft, were far larger than needed while others, railways and coal mines, were desperately short of new equipment and in bad repair. With nothing to export, Britain had no way to pay for imports or even for food. To make matters worse, within a few weeks after the surrender of Japan on Sept. 2, 1945, U.S. President Harry S. Truman, as he was required to do by law, ended lend-lease, upon which Britain had depended for its necessities as well as its arms. John Maynard Keynes, as his last service to Great Britain, had to negotiate a $3,750,000,000 loan from the United States and a smaller one from Canada. In international terms Britain was bankrupt.

Nationalization programs

Labour, nonetheless, set about enacting the measures that in some cases had been its program since the beginning of the century. Nationalization of railroads and coal mines, which were in any case so run down that any government would have had to bring them under state control, and of the Bank of England began immediately. In addition road transport, docks and harbours, and the production of electrical power were nationalized. There was little debate. The Conservatives could hardly argue that any of these industries, barring electric power, was flourishing or that they could have done much differently.

More debate came over Labour's social welfare legislation, which created the "welfare state." Labour enacted a comprehensive program of national insurance, based upon the Beveridge Report but differing from it in important ways. It regularized the de facto nationalization of public assistance, the old poor law, in the National Assistance Act of 1946, and in its most controversial move it established the gigantic framework of the national health service, which provided free, comprehensive medical care for every citizen, rich or poor. The pugnacious temper of the minister of health, Aneurin Bevan, and the insistence of the purists in the Labour Party upon the nationalization of all hospitals provoked the only serious debate accompanying the enactment of this immense legislative program, most of which went into force within two years after Labour's accession to office. Bevan emerged at this time as an important figure on the Labour left and would remain its leader until his death in 1960. He was responsible for perhaps Labour's greatest, although unnoticed, domestic achievement during the Attlee government, the construction, or reconstruction, of nearly one million houses and apartments. In this time of straitened resources, housing, all the public opinion polls showed, was by far the greatest worry of the British population.

Economic crisis and relief (1947). Labour's record in its first 18 months of office was distinguished. In terms of sheer legislative bulk the government accomplished more than any other government in the 20th century save perhaps Asquith's pre-World War I administration or the administration of Margaret Thatcher beginning in 1979. Yet by 1947 it was overtaken by the economic crisis that had not abated. The loan from the United States that was supposed to last four years was nearly gone. Imports were cut to the bone. Bread, never rationed during the war, had to be controlled. Britain had to withdraw support from Greece and Turkey, reversing a policy more than a century old, and call upon the United States to take its place. Thus, at Britain's initiative, came into existence the Truman Doctrine.

The Marshall Plan

Relief came with the U.S. secretary of state George C. Marshall's announcement at the commencement at Harvard University on June 5, 1947, that the United States would undertake a massive program of financial aid to the European continent. Any nation in the Eastern or Western bloc was entitled to take part. Although the Soviet Union immediately denounced the "Marshall Plan" as the beginning of a division between the East and the West, all western European countries, including Britain, hastened to participate. It can be argued that the Marshall Plan and the Truman Doctrine represent the permanent involvement of the United States in Europe.

Withdrawal from the empire. Britain, not entirely by coincidence, was also beginning its withdrawal from the empire. Most insistent in its demand for self-government was India. The Indian independence movement had come of age during World War I and had gained momentum with the inexcusable Amritsar Massacre of 1919. The All-India Congress Party, headed by Mohandas K. Gandhi, the "Mahatma," evoked sympathy throughout the world with its policy of nonviolent resistance, forcing Baldwin's government in the late 1920s to seek compromise. The eventual solution, embodied in the Government of India Act of 1935, provided responsible government for the Indian provinces, the Indianization of the civil service, and an Indian parliament, but it made clear that the Westminster Parliament would continue to legislate for the subcontinent. The act pleased no one, neither the Indians, the Labour Party, which considered it a weak compromise, nor a substantial section of the Conservative Party headed by Churchill, which thought it went too far. Agitation in India continued.

Further British compromise became inevitable when the Japanese in the spring of 1942 swept through Burma to the eastern borders of India while also organizing in Singapore a large Indian National Army and issuing appeals to Asian nationalism. During the war Churchill reluctantly offered increasing installments of independence amounting to dominion status in return for all-out Indian support for the conflict. These offers were rejected by both the Muslim minority and the Hindu majority.

Withdrawal from Asia

The election of a Labour government at the end of World War II coincided with the rise of sectarian strife within India. The new administration determined quickly that Britain would have to leave India. This decision was announced on June 3, 1947, at the height of the postwar financial crisis, and British administration in India ended 10 weeks later, on August 15. Burma and Ceylon, which later became Sri Lanka, received independence by early 1948. Britain, in effect, had no choice but to withdraw from colonial territories it no longer had the military and economic power to control.

The same circumstances that dictated the withdrawal from India required, at almost the same time, the termination of the mandate in Trans-Jordan, the evacuation of all Egypt except the Suez Canal territory, and in 1948 the withdrawal from Palestine, which coincided with the proclamation of the state of Israel. It can be argued that the orderly and dignified ending of the British Empire, which made it possible for most of the former colonial nations to remain friendly with Britain, was Labour's greatest international achievement.

Conservative government (1951–64). The last years of Attlee's administration were troubled by economic stringency and inflation. The pound was sharply devalued in 1949, and a general election on Feb. 23, 1950, reduced Labour's majority over the Conservative and Liberal parties to only eight. Attlee himself was in poor health, and Ernest Bevin, formerly the strongest man in the Cabinet, had died. More radical members of the party, led by Aneurin Bevan, were growing impatient with the increasingly moderate temper of the leadership. On Oct. 25, 1951, a second general election in a House of Commons not yet two years old returned the Conservatives under Churchill to power with a majority of 22.

The Conservatives remained in power for the next 13 years, from October 1951 until October 1964, first under Churchill, who was able to preside over the accession of the new monarch, Queen Elizabeth II, on Feb. 6, 1952, but who was forced to resign on account of age and health on April 5, 1955, and then under Churchill's long-time lieutenant and foreign secretary, Anthony Eden. Eden resigned in January 1957, partly because of ill health, but chiefly because of his failed attempt to roll back the retreat from empire by a reoccupation of the Suez Canal Zone after the nationalization of the canal by the Egyptian president, Gamal Abdel Nasser, in the summer of 1956. This belated experiment in imperial adventure drew wide criticism from the United States and the British dominions,

and indeed within Britain itself. Although it was cut short in December 1956 when UN emergency units supplanted British (and French) troops, it divided British politics as few foreign issues have done since. Eden was succeeded by his chancellor of the exchequer, Harold Macmillan, even though he had supported the Suez invasion. Macmillan remained in office until October 1963 when he too retired because of ill health, to be succeeded by Sir Alexander Douglas-Home, then foreign secretary.

In this period of single-party government, longer than any similar tenure since the 1820s, the themes were economic change and the continued retreat from colonialism. As the 1940s had seen Britain's withdrawal from Asia and the Mediterranean basin, the 1950s and '60s witnessed Britain's withdrawal from Africa.

On the surface the 1950s and early '60s were years of economic expansion and prosperity. The economic well-being of the average Briton rose dramatically and visibly. But when prosperity created a demand for imports, large-scale buying abroad hurt the value of the pound. A declining pound meant higher interest rates as well as credit and import controls, which, in turn, caused inflation. Inflation hurt exports and caused strikes. These crises occurred in approximately three-year cycles.

The economic concern then of the British government in the 1950s and '60s and indeed through the 1970s was to increase productivity and ensure labour peace so that Britain could again become an exporting nation able to pay for public expenditure at home while maintaining the value of its currency and its place as a world banker. A drastic run on the pound had been one of the pressing reasons for the quick withdrawal from Suez in 1956, and throughout the 1950s and '60s Britain's share of world trade fell with almost perfect consistency by about 1 percent per year. On the other hand Britain benefited from an unprecedented rise in tourism occasioned mostly by the attraction of "swinging London."

Refusal to join the EEC

All of this made Britain's decision, after fierce political discussion, not to join the planned European Economic Community (EEC), finally established by the Treaty of Rome on March 25, 1957, an event of signal importance. It meant that although economic conditions in Britain did indeed improve in the last years of the 1950s and through 1960—Prime Minister Harold Macmillan could remark with only slight irony that the British people had never "had it so good"—Britain nevertheless did not share in the astonishing growth in European production and trade, led by the "economic miracle" in West Germany. By the middle 1960s there were signs that British prosperity was declining. Increases in productivity were disappearing and labour unrest was marked. Prime Minister Macmillan quickly realized that it had been a mistake not to join the EEC, and in July 1961 he initiated negotiations to do so. By this time, however, the French government was headed by Charles de Gaulle, and he chose to veto Britain's entry. Although the Labour government that came to power in 1964 continued to press for British inclusion, Britain did not finally join the EEC until 1973.

Labour interlude (1964–68). The long Conservative tenure came to an end on Oct. 16, 1964, with the appointment of a Labour administration headed by Harold Wilson. Wilson had been Labour leader only a little more than a year and a half, since the death of the widely admired Hugh Gaitskell. Although he was thought to be a Labour radical and had attracted a substantial party following on this account, he was in fact a moderate.

Wilson's government inherited the problems that had accumulated during the long period of Conservative prosperity: poor labour productivity, a shaky pound, and trade union unrest. His prescription for improvement included not only a widely heralded economic development plan, to be pursued with the introduction of the most modern technology, but also stern and unpopular controls on imports, the devaluation of the pound, wage restraint, and an attempt, in the event these measures proved unsuccessful, to reduce the power of the trade unions. Eventually the Wilson government became unpopular and was kept in power primarily, as even Cabinet members understood, by weakness and division in the Conservative Party. Finally, in 1968, Wilson was confronted with an outbreak of civil rights agitation in Northern Ireland that quickly degenerated into armed violence.

The return of the Conservatives. The Conservatives returned in a statutory general election on June 18, 1970, with a majority of 32. The new prime minister was Edward Heath, the first Conservative leader to be chosen by a formal ballot of the parliamentary party. Heath set himself three goals: to take Britain into the EEC, to restore economic growth, and to break the power of the trade unions. In his short term in office he succeeded only in the first. In fact, Heath was defeated by the trade unions, who simply boycotted his industrial legislation, and by the Arab oil embargo, which began in 1973 and which made a national coal miners' strike in the winter of 1973–74 particularly effective. Quite properly Heath used the strongest weapon available to a prime minister—a general election, on Feb. 28, 1974—to settle the issue of who governed Britain. The election, held when factories were in operation only three days a week and civilian Britain was periodically reduced to candlelight, was a repudiation of the policy of confrontation with labour. Wilson returned to office and promptly made peace by granting the miners' demands. The concessions to the miners and the abolition of wage controls were confirmed on Oct. 10, 1974, with a second election when Wilson's tiny majority, based upon cooperation from the Scottish and Welsh Nationalist parties as well as the Liberals, was increased to an almost workable margin of 20.

Edward Heath

The last years of the fifth Labour government of the 20th century were unhappy. In a referendum on April 9, 1975, on the EEC, promised by Wilson when in opposition, Britain's membership (which Wilson supported) was confirmed. But neither Wilson nor James Callaghan, who succeeded him on April 5, 1976, was able to come to terms with labour unions, which were as willing to embarrass a Labour government as a Conservative one. Union unrest, induced by rapidly increasing prices, made the late 1970s a period of almost endless industrial conflict, culminating in the winter of 1978–79 in a series of bitter disputes, which the government seemed unable to control and which angered the voters. Meanwhile Labour's slender majority in the House of Commons eroded with the defection of the Liberal and Nationalist parties. On March 28, 1979, Callaghan was defeated by a vote of no confidence, the first such dismissal of a prime minister since Baldwin in 1924. In the subsequent election on May 3, 1979, the Conservatives, under the leadership of Margaret Thatcher, were returned with a majority of 59 over the Labour and Liberal parties.

Margaret Thatcher

Thatcher set out to end socialism in Britain. Her most dramatic acts consisted of a continuing series of statutes to denationalize nearly every industry that Labour had taken under government control in the previous 40 years as well as some industries, such as telecommunications, that had been in state hands for a century or more. But perhaps her most important achievement, helped by high unemployment in the old heavy industries, was in winning the contest for power with the trade unions. Instead of attempting to put all legislation in one massive bill, as Heath had done, Thatcher proceeded step by step, making secondary strikes and boycotts illegal, providing for fines, as well as allocation of union funds, for the violation of law, and taking measures for ending the closed shop. Finally in 1984 she won a year-long struggle with the coal miners, which had been accompanied by continuing violence. Thatcher was helped by a revival of world prosperity and lessening inflation, by the profits from industries sold to investors, and by the enormous sums realized from the sale abroad of North Sea oil. She achieved popularity by sending the armed forces to expel an Argentine force from the Falkland Islands in the spring of 1982, on the strength of which she won triumphant reelection on June 9, 1983. Thatcher was, however, unable to end the strife in Northern Ireland. Nonetheless, she succeeded in 1987 in winning an unprecedented third general election, and in January 1988 Thatcher surpassed Asquith as the longest continually serving prime minister since Lord Liverpool in the 19th century. (B.B.G.)

ENGLAND

England is the predominant constituent unit of the United Kingdom. Outside the British Isles, England is often erroneously considered synonymous with the island of Great Britain (England, Scotland, and Wales) and even with the entire kingdom. Despite the political, economic, and cultural legacy that has secured the perpetuation of its name, England no longer officially exists as a country. Unlike Scotland and Wales, which have their own departments of state and Cabinet ministers, and Northern Ireland, which has self-government in domestic affairs, England enjoys no separate political status within the United Kingdom. It is rare for institutions to operate for England alone. A notable exception is the Church of England, the Anglican Communion having separate churches in Wales and Scotland. In sports, cricket remains strictly a matter for England and English counties (apart from the Glamorgan County Cricket Club in Cardiff, South Glamorgan, Wales), and England fields its own teams in rugby and association football (soccer). Generally, however, England gives the appearance of having been swallowed up in the larger mass of Great Britain since the Act of Union of 1707.

Dominance of England

England's dominance in the United Kingdom, nevertheless, is unquestionable. Occupying a little more than half of the landmass, England has more than four-fifths of the population. The bulk of the most fertile lowlands and six of the United Kingdom's seven conurbations are in England, which also has a higher proportion of wealth and natural resources than the rest of the United Kingdom. Population and wealth increase with proximity to London, the national capital. Greater London (since 1986 a geographic entity only), with its ring of densely populated boroughs, dominates England, as England dominates Great Britain.

The "Englishness" of England is a quality generally acknowledged but difficult to define. Most institutions are British, but poets have rarely saluted Britain. Poets by the score, however, have celebrated England and the English, from William Langland in the medieval English of the 14th century onward. Shakespeare is self-consciously English. Art historian Nikolaus Pevsner wrote of "The Englishness of English Art." Yet this is not an insular character. The English of Anglo-Saxon England absorbed the Danish and Norse invaders and their Norman French conquerors. When the English went on to conquer the remainder of the British Isles and then a worldwide empire, they did so without the dilution or dispersal of their essential character. The dominant role of the English is recognized implicitly in the numerous histories of Britain and the British Empire that are called "A History of England."

Another fundamental English characteristic is diversity within a small compass. England (with a total area of 50,363 square miles [130,439 square kilometres]) occupies only about one-thousandth of the world's land area but contains one-hundredth of the world's population and is, therefore, one of the most densely populated countries in the world. Even the farthest points of England are no more than a day's journey from London, and no place in England is more than 75 miles from the sea.

Physical and human geography

THE LAND

Relief. The diversity of the English landscape is based on a complex geologic structure. Almost every phase of geologic history is illustrated in the intricate patterns of England's geologic map. The oldest sedimentary rocks and some igneous rocks (in isolated hills of granite) are in Cornwall and Devon on the southwestern peninsula, volcanic rocks underlie parts of Cumbria, and the most recent alluvial soils cover the Fens of Cambridgeshire and Norfolk. Between these regions lie bands of sandstones and limestones of different geologic periods, witness to primeval times when large parts of central and southern England were submerged below warm seas. In the period of Alpine mountain-building, these sedimentary rocks were lifted and folded to yield chains of hills ranging from 965 feet (294 metres) in the North Downs to 1,083 feet in the Cotswolds. Other important hills are the Chilterns and those in Lincolnshire and North Yorkshire. These hills were rounded into characteristic plateaus with west-facing escarpments by the glaciers of three successive Ice Ages. When the ice melted, the sea level rose, which submerged the land bridge with Europe. The landscape was altered by deep deposits of sand, gravel, and glacial mud.

Hills and downs

Erosion by rain, river, and tides and subsidence in parts of eastern England have further shaped the hills and the coastline. The Pennines, with moorland tops of 2,000 to 3,000 feet, split northern England into northwest and northeast. These plateaus of limestone, grit-stone, and carboniferous strata are associated with major coalfields, some outcropping to the surface. This geologic history underlies the variety of the English landscape, for in journeys of only a few miles it is possible to pass through a succession of different soil structures—such as from chalk down to alluvial river valley, from limestone to sandstone and acid heath, and from clay to sand—each type of soil bearing its own class of vegetation. The Cumbrian Mountains, which include the famous Lake District, are just west of the Pennines. Scafell Pike (3,210 feet), in the central part of the Lake District, is the highest point in England. Slate covers most of the northern portion of the mountains, and thick beds of lava are found in the southern part.

The English shoreline

The geologic complexity of England is strikingly illustrated in the cliff structure of its shoreline. Along the southern coast from the ancient granite cliffs of Land's End in the extreme southwest is a succession of sandstones of different colours and of limestones of different ages, culminating in the white chalk from the Isle of Wight to Dover. A constantly changing panorama of cliffs, bays, and river estuaries distinguishes the English coastline, which, with its many indentations, is some 2,000 miles long.

Drainage. The drainage pattern of England is influenced by its relief. The Pennines, the Cotswolds, and the moors and chalk downs of southern England serve as watersheds for most of the rivers. The Eden, Ribble, and Mersey rise in the Pennines, flow westward, and have a short course to the Atlantic Ocean. The Tyne, Tees, Swale, Aire, Don, and Trent rise in the Pennines, flow eastward, and have a long course to the North Sea. The Welland, Nene, and Great Ouse rise in the northeastern edge of the Cotswolds and drain into the Wash estuary, which forms part of the North Sea. The Welland River valley forms part of the rich agricultural land of Lincolnshire. The Thames, the longest river in England, also rises in the Cotswolds and drains a large part of southeastern England. From the moors and chalk downs of southern England rise the Tamar, Exe, Stour, Avon, Test, Arun, and Ouse. All drain into the English Channel and in some instances help to form a pleasing landscape along the coast. England's largest lake is Windermere, with an area of six square miles (16 square kilometres), located in the county of Cumbria.

Soils. The Cumbrian Mountains and most of the southwestern peninsula have acid brown soils. The eastern section of the Pennines has soils ranging from brown earths to podzols. In much of southern England, which has less than 40 inches (1,000 millimetres) of rainfall, leached brown soils predominate. Acid soils and podzols occur in the southeast. Regional characteristics, however, are important. Black soil covers the Fens in Cambridgeshire and Norfolk; clay soil predominates in the hills of the Weald (in East Sussex and West Sussex); and the chalk downs, especially the North Downs of Kent, are covered by a variety of stiff, brown clay, with sharp angular flints. Fine-grained deposits of alluvium occur in the flood plains, and fine marine silt occurs around the Wash estuary.

Climate. Weather in England is as variable as the topography. As in other temperate maritime zones, the averages are moderate, ranging in the Thames Valley from about 35° F (2° C) in January to 72° F (22° C) in July; but the extremes in England go below 0° F (−18° C) and above

90° F (32° C). The Roman historian Tacitus recorded that the climate was "objectionable, with frequent rains and mists, but no extreme cold." Yet there is snow cover in the higher parts of England about 50 days a year. Though known as a wet country, northeastern and central England has less than 40 inches (1,000 millimetres) of rainfall annually and frequently suffers from drought. In parts of the southeast, the annual rainfall averages only 20 inches.

Rainfall

Charles II thought that the English climate was the best in the world—"a man can enjoy outdoor exercise in all but five days of the year." But no one would dispute that it is unpredictable: hence Dr. Samuel Johnson's observation that "when two Englishmen meet their first talk is of the weather." This changeability of the weather, not only season by season but day by day and even hour by hour, has had a profound effect on English art and literature.

Plant and animal life. England shares with the rest of Britain a diminished range of vegetation and living creatures, partly because the island was separated from the mainland of Europe soon after much of it had been swept bare by the last Ice Age and partly because the land has been so industriously worked. Woodland covers merely 7 percent of England's total land area. A drastic depletion of mature broad-leaved forests, especially oak, was a result of the overuse of timber in the iron and shipbuilding industries. A modified pattern of vegetation has now emerged through overgrazing, forest clearance, reclamation and drainage of marshlands, and the introduction of exotic plant species. Though there are fewer species of plants than in the European mainland, they span a wide range and include some rarities. Certain Mediterranean species exist in the sheltered and almost subtropical valleys of the southwest, while vestiges of the Ice Age tundra survive in parts of the moorland of the northeast. England has a profusion of summer wildflowers in its fields, lanes, and hedgerows. In some parts these have been severely reduced by the use of herbicides on farms and roadside verges. Cultivated gardens, which contain many species of trees, shrubs, and flowering plants from all over the world, account for much of the varied vegetation of the country.

Several animal species such as the bear, wolf, and beaver were exterminated in historic times, but others such as the fallow deer, rabbit, and rat have been introduced. More recently birds of prey have suffered at the hands of farmers protecting their stock and their game birds. Under protective measures, including a law restricting the collecting of birds' eggs, some of the less common birds have been reestablishing themselves. The bird life is unusually varied, mainly because the country lies along the line of bird migrations. Some birds have found town gardens, where they are often fed, a favourable environment, and in London about 100 different species are recorded annually. London also is a favourable habitat for foxes, which in small numbers have colonized woods and heaths within a few miles of the city centre. There are few kinds of reptiles and amphibians—about half a dozen species of each—but they are nearly all plentiful where conditions suit them. Freshwater fish are numerous; the char and allied species of the lakes of Cumbria probably represent an old group related to the trout that migrated to the sea before the tectonic changes that formed these lakes cut off their outlet. The marine fishes are abundant in species and in absolute numbers. The great diversity of shores produces habitats for numerous species of invertebrate animals.

Settlement patterns. Though the significance of the natural changes in England can hardly be exaggerated, the current landscape has been significantly changed by humans. Only the remotest moorland and mountain tops have been untouched; even the bleak Pennine moors of the north are crisscrossed by dry stone walls and their vegetation modified by the cropping of mountain sheep. There is virtually no genuine wilderness left in England. The face of the country has been constantly worked over for centuries, and the marks of its exploitation and use dominate the contemporary landscape. The oldest traces are the antiquarian survivals, such as the Bronze Age forts studding the chalk downs of the southwest, and the corrugations left by the strip farming of medieval open fields.

Towns and villages

More significant is the structure of town and village, which was established in Roman-British and Anglo-Saxon times and has persisted as the basic pattern. The English live in scattered groups of high-density occupation, whether in villages or towns or, in modern times, cities. Despite its badly managed urban sprawl, as cities spilled out into conurbations, England is notable for extensive tracts of farming countryside that lie between its towns, its smaller villages often engulfed in the vegetation of trees, copses, hedgerows, and fields: in a phrase of the poet Gerard Manley Hopkins, "the sweet especial rural scene," which is so prominent in English literature and English art.

The visual impact of a mostly green and pleasant land can be seriously misleading. England is primarily an industrial country, built up during the Industrial Revolution by ruthless exploitation of the coalfields and of cheap labour, especially in the cotton-textile areas of Lancashire, in the woolen-textiles areas of Yorkshire, and in the engineering centres of the Midlands and the northeast. England has large tracts of derelict areas, scarred by the spoil heaps of the coal mines, by quarries and clay pits, by abandoned industrial plants, and by rundown slums.

One of the earliest initiatives to preserve the heritage of the past was the establishment in 1895 of the National Trust, a private organization dedicated to the preservation of historic places and natural beauty in England, Wales, and Northern Ireland. (There is a separate National Trust for Scotland.) In 1957 the Civic Trust was established to promote interest in and action on issues of the urban environment. Hundreds of local societies dedicated to the protection of the urban environment have been set up, and many other voluntary organizations as well as government agencies are working to protect and improve the English landscape. Greenbelts have been mapped out for London and other conurbations. The quality of town life has been improved by smoke control and checks on river pollution, so effectively that the recorded sunshine in London and other major urban centres has greatly increased and the "pea soup" fogs that once characterized London have become memories of the past. Fish are returning to reaches of rivers such as the Thames, the Tyne, and the Tees from which they had been driven by industrial pollution.

Protection of the urban environment

THE PEOPLE

Ethnic background. By ethnic origin the English are a mongrel breed. Their language is polyglot, drawn from a variety of sources, and its vocabulary has been augmented by importations from all over the world. The English language does not identify the English, for it is the main language of Wales, Scotland, Ireland, many Commonwealth countries, and the United States. The primary source of the language, however, is the main ethnic stem of the English, the Anglo-Saxons, who invaded and colonized England in the 5th and 6th centuries. Their language provides about half the words in modern English vocabulary.

In the millennia following the last Ice Age, the British Isles were peopled by migrant tribes from the continent of Europe and, later, by traders from the Mediterranean area. During the Roman occupation, England was inhabited by Celtic Brythons, but the Celts withdrew before the Teutonic Angles, Saxons, and Jutes (from northwestern Germany) into the mountainous areas of western and northern Britain. The Anglo-Saxons neither preserved nor absorbed the Roman-British culture they found in the 5th century. There are few traces of Celtic or Roman Latin in the early English of the Anglo-Saxons, though some words survive in place-names, such as the Latin *castra* for camp providing the suffix "cester" and *combe* and *tor,* Celtic words for "valley" and "hill." Old Norse, the language of the Danes and Norsemen, left more extensive traces, partly because it had closer affinities to Anglo-Saxon and because the Danish occupation of large tracts of eastern and northern England was for a time deeply rooted, as some place-names show.

Roman and Anglo-Saxon occupation

The history before the Norman Conquest is poorly documented, but what stands out is the tenacity of the Anglo-Saxons in surviving a succession of invasions. They united most of what is now England in the 9th, 10th, and 11th centuries, only to be overthrown by the Normans in 1066. For two centuries French became the language of the

court and the ruling nobility; yet English prevailed and by 1362 had reestablished itself as an official language. Church Latin, as well as a residue of Norman French, was incorporated into the language during this period. It was subsequently enriched by the Latin and Greek of the educated scholars of the Renaissance. The seafarers, explorers, and empire builders of modern history have imported foreign words, most copiously from Europe but also from Asia. These words have been so completely absorbed into the language that they pass unselfconsciously as English. The English, it might be said, are great anglicizers.

The English have also absorbed and anglicized people of alien race, from Scandinavian pillagers and Norman conquerors to Latin churchmen. In the royal line, a Welsh dynasty of monarchs, the Tudors, was succeeded by the Scottish Stuarts, to be followed by the Dutch William of Orange and the German Hanoverians. For the Scots, the Welsh, and the Irish, English became their main language.

Refugees and immigrants

England provided a haven for refugees from the time of the Huguenots in the 17th century to the totalitarian persecutions of the 20th century. Many Jews have settled in England. In recent decades there has been large-scale immigration from Asia, Africa, and the Caribbean, posing seemingly more difficult problems of assimilation, and restrictive immigration regulations have been imposed that are out of key with the open-door policy that had been an English tradition for many generations. To be counted English, it has never been necessary to be of purebred English stock, and indeed, there can be few English who are.

Traditional regions. Though England is a small and homogeneous country bound together by law, administration, and a comprehensive transport system, distinctive regional differences arise from geography and history. It was natural for different groups of the population to establish themselves in recognizable physical areas. In the north, for example, the east and west are separated by the Pennines, and the estuaries of the Humber, the Thames, and the Severn form natural barriers. The geographic regions are often referred to as the standard regions of England, but they do not serve administrative functions. These eight regions are the South West, the South East, the West Midlands, the East Midlands, East Anglia, the North West, Yorkshire and Humberside, and the North.

The South West. The South West contains the last Celtic stronghold in England, Cornwall, where a Celtic language was spoken until the 18th century. There is even a small nationalist movement, Mebyon Kernow (Sons of Cornwall), seeking to revive the old language. Though it has no political significance, the movement reflects the disenchantment of a declining area, with the exhaustion of mineral deposits toward the end of the 19th century. The neighbouring county of Devon also shares this economic decline. These southwestern counties, however, have a large share of the rapidly expanding tourist industry, with a splendid coastline and the Dartmoor and Exmoor national parks. Farther east the counties of Dorset, Hampshire, Avon, and Gloucestershire have diversified manufacturing and service industries. The county of Wiltshire is mainly agricultural. It is famous for the prehistoric stone circles at Stonehenge and the remains of a prototype at Woodhenge. Growth in the manufacturing sector has been a contributing factor for the region's significant population increase.

The South East. The South East has a population and wealth to match many nation states. This is the dominant area of England and the most rapidly growing one despite planning controls directed toward restricting its growth. London itself is a victim of urban sprawl that has been checked too late. It is the administrative headquarters of not only government but also many of Britain's industrial, financial, and commercial undertakings. The area as a whole has an extensive range of manufacturing industry (brewing, clothing, furniture, and printing). With improvements in the transportation systems, however, nuclear and space research facilities, retailing, advertising, and service and high-technology industries are gradually moving to the counties outside London. A greenbelt policy has slowed down the expansion of London, and the region as a whole still has one-third of its area devoted to farming or horticulture. London is the focus of the national transport system, with the mainline railway network based on the London termini, two international airports, at Heathrow and Gatwick, and a third airport at Stansted, in Essex. The Port of London is the largest and commercially most important in Britain. Eight new towns have been established in the South East, and four existing towns have been expanded to take the overflow of London population. With its theatres, concert halls, museums, and art galleries, London is the cultural capital of the country. Whether the people of the South East feel a regional identity is questionable. The Sussex Downs and the Bedfordshire plain or Oxford and Canterbury have nothing much in common apart from being within the magnetic pull of London. Loyalties are more specifically to towns, such as Oxford, Cambridge, or Brighton, and within London there is a sense of belonging more to localities—such as Chelsea or Hampstead, which acquire something of the character of urban villages—than to the metropolis as a whole.

The expansion of London

The West Midlands. Regional characteristics are stronger outside the South East. The West Midlands, comprising the counties of Hereford and Worcester, Shropshire, Staffordshire, and Warwickshire, and also the cities of Birmingham, Coventry, and Wolverhampton, includes the Black Country (an urban agglomeration whose name reflects the coating of grime and soot afflicting the buildings of the region), with various industrial trades, mainly in the metal industries. The major industry is automobile manufacture. With a history going back to the beginnings of the Industrial Revolution, the West Midland towns are ugly but prosperous. It is an area with remarkable specialization place by place: locks and keys in Wolverhampton and Willenhall; the bulk of the country's pottery and china in five towns grouped around Stoke-on-Trent; chocolate at Bournville; and carpets at Kidderminster. The region

The Black Country

Milt and Joan Mann—CAMERAMANN INTERNATIONAL

Jaguar automobile factory, Coventry, Eng.

is not exclusively industrial, with the Shakespeare country around Stratford-on-Avon, the fruit orchards of the Vale of Evesham, and hill country on the Welsh border.

The East Midlands. The East Midlands are less coherent as a region, taking in the manufacturing towns of Northampton (footwear); Leicester (hosiery, knitwear, footwear, engineering); Nottingham (hosiery, lace, bicycles, pharmaceuticals, tobacco); and Derby (railway locomotives, a Rolls-Royce factory, fine porcelain ware). The region includes some of the most productive coalfields and important ironstone mining, and, in broad swathes between the industrial towns, much of England's best farmland. Several canals in the region, including the Grand Junction and Trent and Mersey, were primarily used for commerce from the late 18th to the early 20th century. They are now being revived, mainly for recreational use.

East Anglia. This region retains an air of remoteness that belongs to its history. With the North Sea on its northern and eastern flanks, it was at one time almost cut off by fenland to the west (now drained) and forest (long ago cleared) to the south. In medieval times it was one of the richest wool regions and, in some parts, was depopulated to make way for sheep. It is now the centre of some of the most industrialized farming in England. Vast fields are devoted to cereals, potatoes, and sugar beets. Compared with other regions, East Anglia has a low population density; with rapid industrialization, however, this pattern is changing. The county of Suffolk has a wide industrial base, with printing and production of agricultural machinery, electronics, and automobile components. Norwich, in Norfolk, has diverse manufacturing and service industries. Bacton, on the coast of Norfolk, provides oil and gas terminals for the fields in the North Sea. Newmarket, in Suffolk, is a world-famous centre for horse racing.

Industrialized agriculture

The North West. Regions become more distinctive the farther they are from London. The North West comprises the counties of Cheshire and Lancashire, with the conurbations of South East Lancashire (including Manchester) and Merseyside (including Liverpool), chronically wet and murky. This region is the home of the declining cotton-textiles industry—rapidly being replaced by diversified manufacturing (food processing, chemicals, computers, glass, and rubber products) but still ravaged by the Industrial Revolution. The region expresses itself in an accent of its own, with strongly flavoured variety-hall humour (of which George Formby and Gracie Fields were classic examples), and gave birth to British rock music, with the Beatles and other groups in Liverpool. It is a region still breaking into the new territories of modern industry, its old cotton towns symbolically overshadowed by the grim gritrock Pennine escarpments, which have been stripped of their trees by two centuries of industrial smoke. The former textile towns of Blackburn, Bolton, Oldham, and Rochdale have a wide range of industries, including brewing, electronics, chemicals and paint, textile machinery, and boatbuilding. Several canals, including the Manchester Ship Canal and the Leeds and Liverpool canals, all heavily used for commerce, dominate the landscape of the region.

Yorkshire and Humberside. The industrial conurbation of West Yorkshire on the east side of the Pennines watershed has a character similar to that of the North West—the valley mills are engaged in the manufacture of woolen textiles, clothing, machine tools, and electronics. This region also shows a rugged independence of character expressed in a tough style of humour. The conurbation of West Yorkshire (including the cities of Leeds and Bradford) takes in the woolen textile, financial, and commercial centres. Farther south, steel is concentrated at Sheffield, world-famous for its cutlery and silver plate. West Yorkshire, which formerly based its prosperity on coal, has been diversifying, notably into chocolate, glass, chemicals, and engineering. There, too, industrialism has left ugly scars. Nonetheless, this region has extensive areas of farming, a deep-sea fishing industry operating from Hull, and tourist country along a fine coast in the east (North York Moors National Park) and in the beautiful valleys of the Yorkshire Dales National Park in the west.

The North. This region extends to the Scottish border, taking in the counties of Northumberland, Cumbria, North Yorkshire, Cleveland, and Durham. It also includes the conurbation of Tyneside and is therefore unusually diverse. The Lake District—a distinctive region within the Cumbrian Mountains—is so shut off that it retains a strong sense of local identity quite unlike the North East, which, from the River Tees to the Scottish border, is by cultural climate a region of its own. The Lake District National Park, the Solway coast, the northern Pennines, Hadrian's Wall, and part of the Yorkshire Dales National Park contribute to the scenic landscape of Cumbria. Important industries in Barrow-in-Furness, in the county of Cumbria, are shipbuilding, construction, and the production of nuclear submarines, chemicals, paper, and electronics. The county of Cleveland was heavily industrialized (iron and steel and shipbuilding) during the 19th century, but it has become an important tourist centre along the North Sea and includes the North York Moors National Park. Cleveland also has the largest petrochemical complex in Europe, and oil from the Ekofisk field in the North Sea is piped ashore at Teesport. Coal mining is still important in the county of Durham, but in recent years the emphasis has been on engineering, the manufacture of pharmaceuticals, and service industries. The local flavour of life can be found in the folk songs of Tyneside and the coal-mining villages and in the dialect known as Geordie. The city of Newcastle upon Tyne is an important industrial centre and serves as a commercial centre for the conurbation of Tyneside. The region also contains some of the most desolate land in England, in the Cheviot Hills along the Scottish border.

The Lake District

Religion. Increasingly, England is a secular country. Though nearly three-fifths of the population are baptized in an Anglican church and for form-filling purposes would say they belonged to the Church of England, fewer than one in 40 of the baptized are communicant churchgoers. The Church of England still has some 16,500 churches, but it has been in financial difficulties. The nonconformist Free Churches have nominally fewer members, but there is probably greater dedication among them, as with the Roman Catholic church. Apart perhaps from some isolated centres of Irish settlement in the North West, there is complete religious tolerance in England and no overt prejudice against Catholics. The decline in churchgoing has been thought to be an indicator of decline in religious belief, but opinion polls substantiate the view that belief in God and the central tenets of Christianity survives the flagging fortunes of the churches. There are also large communities of Muslims, Jews, Sikhs, and Hindus.

Demographic trends. England comprises more than four-fifths of the total population of the United Kingdom. Although during the 1970s and '80s the overall birth rate remained constant, the number of births per thousand women between the ages of 20 and 24 fell by two-fifths; this drop reflected a trend among women in that age group to delay both marriage and childbirth. Although the overall death rate remained constant, the mortality rate among young children and young adults decreased. In the 1980s there was a decrease in the proportion of young adults aged 21, and from 1951 through the 1980s the number of people aged 65 and older grew by more than one and one-half times. During the 1980s the population of conurbations, especially Greater London and Merseyside, decreased, as people moved to the rural areas. The standard regions of East Anglia, the South West, and the South East (excluding Greater London) gained population, while the other standard regions all lost population.

Birth and death rates

THE ECONOMY

Until the 18th century, the economy of England was mainly agricultural. With the advent of the Industrial Revolution, however, England gradually evolved into a highly urbanized and industrial region. During the late 18th and the 19th centuries, the growth of heavy industries (iron and steel, textiles, and shipbuilding) in the northeastern counties was based on the proximity of coal and iron-ore deposits. During the 1930s, the Depression and foreign competition contributed to a decrease in the production of manufactured goods and an increase in unemployment in the factories of Lancashire, Cheshire, and Staffordshire.

The unemployed from these northern counties moved south to London and the surrounding counties. The southeast became urbanized and industrialized, with automotive, chemical, electrical, and machine tool manufactures as the leading industries. An increase in population and urban growth during the 20th century caused a significant drop in the acreage of farms in England, but the counties of Cornwall, Devon, Kent, Lincolnshire, Somerset, and North Yorkshire have remained agricultural.

Resources. Coal is England's richest natural resource and has always met the nation's requirement for energy. Important coal reserves (mainly bituminous) are found in Northumberland, Durham, Nottinghamshire, and Derbyshire. The Vale of Belvoir, in Leicestershire, has recoverable reserves estimated at 510 million tons. Exploitation at the deep coal mine at Selby, in North Yorkshire, is planned to yield about 10 million tons of coal per year when completed. Onshore crude oil fields are located in Dorset, Hampshire, Lincolnshire, Nottinghamshire, and Humberside. Seven major and several minor natural gas fields, located in the North Sea, supply England with natural gas. The sources of England's water are both underground (wells, adits, and boreholes) and surface (lakes and rivers). In southeastern England the Downs, the Chiltern Hills, and the major part of the Isle of Wight have chalk soil, underneath which lie the sands and clays that seal the huge water reservoir of the London Basin.

Only 7 percent of England's landscape is woodland. Broad-leaved (oak, beech, ash, birch, and elm) and conifer (pine, fir, spruce, and larch) trees dominate the landscapes of Kent, Surrey, East Sussex, West Sussex, Suffolk, and Hampshire. Important forests include Ashdown in East Sussex; Epping and Hatfield in Essex; Dean in Gloucestershire; Sherwood in Nottinghamshire; Redesdale, Kielder, and Wark in Northumberland; and Grizedale in Cumbria. A substantial acreage of forests is privately owned.

Important forests

Agriculture, forestry, and fishing. The physical environment and natural resources of England are more favourable to agricultural development than those of other parts of the United Kingdom. A greater proportion of the land consists of lowlands with good soils where the climate is conducive to grass or crop growing. The majority of English farms are small, most holdings being less than 250 acres (100 hectares); nonetheless, they are highly mechanized.

Major crops. Wheat, the chief grain crop, is grown in the drier, sunnier counties of eastern and southern England, where new, stronger varieties have become increasingly widespread and average yields, with the aid of chemical fertilizers and weed killers, have risen significantly since the 1960s. Barley is grown mainly for livestock feed and for malting and other industrial markets. The acreage under oats is gradually declining. Corn (maize) and rye are also grown. Principal potato-growing areas are the fenlands of Norfolk, Cambridgeshire, and Lincolnshire; the clay soils of Humberside; and the peats of North Yorkshire. Sugarbeet production depends heavily on government subsidy because of competition from imported cane sugar. In recent years, acreage and yield for rapeseed have increased. Grass and its variants are grown for feeding livestock.

Horticulture. The growing of vegetables, fruit, and flowers, known in England as market gardening, is often done in greenhouses and is found within easy trucking distance of large towns, the proximity of a market being of more consequence than climatic considerations. The fertile (clay and limestone) soil of Kent has always been conducive to fruit growing. Cultivation was first established there on a commercial scale in the 16th century. The county of Kent is a major supplier of fruits and vegetables (apples, pears, black currants, cauliflowers, and cabbages). Hereford and Worcester is noted for its plums, while Somerset and Devon specialize in cider apples.

Livestock. The agriculture of England, though to a lesser extent than in Wales and Scotland, is primarily concerned with livestock husbandry and, in particular, with milk production. Dairying is important in every county, though the main concentrations are in western England. The quality of dairy cattle was improved considerably after World War II by the artificial insemination service, operated largely by the Milk Marketing Board, which makes available the best dairy strains to all farmers. The higher-yielding dairy breeds, including the Frisian and Ayrshire, have become more numerous than the once-dominant Shorthorn.

Dairy farming

Domestic production supplies most of the country's beef needs. Calves born to dairy cattle but sired by beef breeds are increasingly important sources of veal and beef. Special beef breeds, for which Britain is famous, are bred throughout the country, but long-established specialist areas retain their importance. Cattle are often moved from one region to another for raising, storing, and final fattening.

Hill sheep are bred in the Pennines, the Lake District, and the southwestern peninsula, areas where sheep are occasionally the main source of a farmer's income but frequently of subsidiary importance to cattle. The production of fat lambs rather than wool is the main concern of English sheep farmers. Grass-fed breeds, yielding lean meat, are much more important than the large breeds, raised on arable land, that were characteristic of the 19th century.

While specialist pig farms are rare, they do exist, supplying the large sausage and bacon companies. Poultry are kept in small numbers on most farms, but specialist poultry farms, notably in Lancashire and in the southeastern counties serving the London market, have increased.

Forestry. Many forests in England are managed by the Forest Commission, which besides promoting timber production also emphasizes wildlife preservation. During the 18th and 19th centuries, timber was heavily used by the iron and steel and shipbuilding industries. Presently, demand for timber continues in construction and furniture industries, but with the government's afforestation program in effect, new forests are coming to dot the landscape.

Fishing. Freshwater fish, including bream, carp, perch, pike, and roach, are available in the rivers of eastern England. Cod, haddock, whiting, herring, plaice, halibut, turbot, and sole are caught in the North and Irish seas. Several ports, including Lowestoft, Great Yarmouth, Grimsby, Bridlington, and Fleetwood, have freezing and processing plants nearby. Oyster farms are located along the creeks and estuaries in Essex, and rainbow-trout farming has become popular. Salmon fishing is prohibited in waters more than six miles from the coasts of England.

Industry. Iron ore is mined in Northamptonshire and Humberside. Cornwall is the only county in England that provides the nation with tin ore. Sand, gravel, and crushed rock, widely available, provide raw materials for the construction industry. Clay and salt are found in the northwestern England, and china clay is available in Cornwall.

More than two-thirds of those employed in England work in the service industries. London is a major financial, banking, and insurance centre. Cambridge, Ipswich, and Norwich are important service and high-tech centres. Nearly a quarter of England's workers are employed in manufacturing. Major industries located in the northern counties include food processing, brewing, and the manufacture of chemicals, textiles, computers, automobiles, aircraft, clothing, glass, and paper and paper products. Leading industries in southeastern England are pharmaceuticals, computers, microelectronics, aircraft parts, and automobiles.

England produces 90 percent of Britain's coal. More than four-fifths of its total energy is from coal-fired steam plants, less than one-fifth from nuclear power plants, and the remainder from oil and natural gas. The world's first large-scale nuclear power plant was opened in 1956, at Calder Hall, in Cumbria. Several nuclear power plants, including eight in southern England, are now in operation.

Energy production

Tourism plays a significant role in England's economy. The country's attractions appeal to a wide variety of interests, ranging from its rich architecture, archaeology, arts, and culture, to its horticulture and scenic landscape.

Transportation. England is well served by roads, railways, ports, and airports. Highways radiate from London in all directions, and the increase in traffic is visible in the congested highways. London, other large cities, and towns are linked by an efficient intercity network of trains. Several high-speed freight trains serve the major industrial centres. London's underground train system, the "tube," covers some 250 route miles. Inland waterways were de-

veloped during the 17th and 18th centuries, mainly to carry bulky raw materials such as coal, iron ore, and limestone between the industrial centres of Manchester, Leeds, Sheffield, Kingston upon Hull, Birmingham, and London. By the end of the 18th century, a "cross" system of canals connected the Thames, Humber, Mersey, and Severn estuaries. Most canals are now in disuse.

During the 1980s, Britain's trade with Europe has increased sharply, and the ports in southern and southeastern England now handle significantly higher traffic than the major ports of London, Liverpool, and Manchester. Leading ports that handle container traffic are Felixstowe, London, and Southampton; Dover, Grimsby, and Harwich chiefly handle roll-on traffic. Major airports around London are Heathrow, Gatwick, and Stansted, which together handle more than 40 million passengers annually. Airports at Birmingham, Manchester, Newcastle upon Tyne, and Luton also handle significant amounts of traffic.

ADMINISTRATION AND SOCIAL CONDITIONS

Government and politics. A specifically English role in contemporary government and politics is hard to identify, for these operate on a nationwide British basis. Historically, the English may be credited with the evolution of Parliament, which, in its medieval form, was related to the Anglo-Saxon practice of regular gatherings of notables; and the English may also be credited with the glory of the 1688 Revolution, which affirmed the rule of law, parliamentary control of taxation and of the army, freedom of speech, and religious toleration. Freedom of speech and opinion with proper opportunities for reasonable debate form part of the English tradition, but the development of party and parliamentary government in its modern forms took place after the Act of Union of 1707, when, in politics, the history of England became the history of Britain.

Local government

In government the English legacy remains conspicuous in local affairs, which are still largely administered on a county base that can be traced back to the Anglo-Saxon shires. The separation of county and town was the principle underlying the late Victorian reorganization of local government. The reforms of the 1970s established a two-tier system of counties and districts. In 1986 the Greater London Council and the metropolitan county councils of Greater Manchester, Merseyside, South Yorkshire, Tyne and Wear, West Midlands, and West Yorkshire were abolished. The functions of the Greater London Council were transferred to various authorities and boroughs of London, and those of the metropolitan county councils were transferred to the metropolitan districts. There is another local administrative unit, the parishes, which are parts of smaller boroughs and urban districts that existed prior to 1972. Of the some 10,000 parishes in England, four-fifths have their own councils. Although the functions of the county and district are distinct, the functions of the parish are concurrent with those of the districts.

Justice. The English have given the world, notably North America and much of the Commonwealth, the system of English law that has acquired a status and universality to match Roman law. This, too, had its origins in Anglo-Saxon times, and two of its hallmarks are its

Anglo-Saxon system of law

preference for customary law (the common law) rather than statute law and its system of application by locally appointed part-time magistrates, by locally chosen juries, and by the traveling judges going from one county town to another on circuit. Under the Normans, the Anglo-Saxon system was retained but formalized; for example, by the recording of case law to provide uniform precedents, which began in the 13th century. In modern times there has been a greater reliance upon the statute law contained in some 3,000 acts of Parliament, but there are more than 300,000 recorded cases to turn to for precedent. Other aspects of the English law are the fundamental assumption that an accused person is deemed innocent until proved guilty and the independence of the judiciary from intervention by crown or government in the judicial process.

The legal system is divided into civil and criminal courts. The House of Lords is the ultimate court of appeal for both civil and criminal cases brought through the High Court or the Court of Appeal. In 1971, the Crown Courts replaced the individual courts (quarter sessions and assizes), and it is now a single court that may sit anywhere in England, deal with any trial on indictment, and hear appeals and proceedings either on a sentence or on civil matters. At the base of the criminal court system, the magistrates' courts try more than 97 percent of the criminal cases.

Education. In England, the Department of Education and Science is responsible for all levels of education. Universities, however, are self-governing and depend on the central government only for financial grants. Education is compulsory between the ages of five and 16. About one-third of primary and secondary schools in England are administered by Anglican or Roman Catholic voluntary organizations. More than 90 percent of the secondary-school population (children aged 11 through 18) attend state-funded comprehensive schools, in which admission is not based on aptitude alone, and the remainder attend either grammar (founded on the principle of teaching grammar [meaning Latin] to boys) or secondary modern schools. Tertiary colleges offer a full range of vocational and academic courses to students aged 16 and older. Independent schools provide both primary and secondary education but charge tuition. In large cities, a number of independent schools are run by various ethnic and religious communities. So-called public schools, which actually are private, are often categorized as independent schools. Most public schools are residential, are privately financed, and provide education to children aged 11 through 19. Important boys' public schools include Eton (the oldest; established 1440–41), Harrow, Winchester, and Westminster; famous girls' public schools include Cheltenham, Roedean, and Wycombe Abbey. There are also private, mostly residential, preparatory (prep) schools, which prepare students aged 7 through 13 for the Common Entrance Examination required to enter senior secondary schools. At the completion of secondary education, students receive the General Certificate of Secondary Education.

Universities and colleges

More than a third of England's young adults receive some form of postsecondary education through colleges, polytechnics, and universities. The universities of Oxford and Cambridge date from the 12th and 13th centuries, and both have university presses that are among the oldest printing and publishing houses in the world. There are about 35 universities in England, some of which are referred to as "red brick" universities. These universities were founded in the late 19th or early 20th century in the industrial cities of Manchester, Liverpool, Leeds, Birmingham, Sheffield, and Bristol and were constructed of red brick, as contrasted with the stone construction of the buildings of Oxford and Cambridge. A continuing education program through the Open University (1969), in Milton Keynes, Buckinghamshire, offers education through correspondence courses and the electronic media.

Health and welfare. Improvements in health care are reflected by the increase in longevity for people in England. Life expectancy increased from 68 years in 1961 to 71.8 years in 1985 for males, and from 73.9 years to 77.7 years for females. Coronary heart disease and cancer are the major causes of death among men aged 50 and over and, in recent years, among women aged 40 and over. Although certain infectious diseases such as poliomyelitis and tuberculosis have virtually disappeared, in recent years the incidence of whooping cough and acute meningoccocal meningitis has increased among the children in England.

The National Health Service, an organ of the central government, provides comprehensive medical services for every resident of England. Doctors, dentists, opticians, and pharmacists work within the service as independent contractors. Social services are provided through local authority social service departments. The services are directed toward children and young people, low-income families, the unemployed, the disabled, the mentally ill, and the elderly. Several religious organizations provide help and advice as well. The National Insurance Scheme insures individuals against loss of income due to unemployment, maternity, and long-term illnesses. It provides retirement pensions, widows' and maternity benefits, child and guardian allowances, and benefits for job-related injuries or death.

National Insurance Scheme

CULTURAL LIFE

England's contribution to both British and world culture is too vast for anything but a cursory survey. In the contemporary cultural scene, England is not always distinguishable from Wales and Scotland or even Northern Ireland.

Literature. It is, arguably, in its literature that England has attained its most influential cultural expression. For more than a millennium, each stage in the development of the English language has produced its masterworks.

The heroic poem *Beowulf,* dating from the 9th or 10th century, preserves the earliest literary language of Britain, the Germanic Anglo-Saxon, known as Old English. Following the Norman Conquest of 1066, French influence shaped the vocabulary as well as the literary preoccupations of Middle English. Geoffrey Chaucer epitomized both the courtly philosophical concerns and the earthy vernacular of this period in his *Troilus and Criseyde* and *The Canterbury Tales,* respectively. The Elizabethan era of the late 16th century fostered the flowering of the European Renaissance in England and the golden age of English literature. The plays of William Shakespeare, while on their surface representing the culmination of Elizabethan English, achieve a depth of characterization and richness of invention that have fixed them in the dramatic repertoire of virtually every language. Political and religious conflicts of the 17th century inspired a wealth of poetry, ranging from the metaphysical introspections of John Donne to the visionary epics of John Milton.

The Elizabethan era

The dichotomy of Classicism and Romanticism, of reason and imagination, came to dominate the 18th century, with the Neoclassical satire and criticism of Alexander Pope, Jonathan Swift, and Samuel Johnson on the one hand, and the somewhat later Romantic self-expression of William Blake, William Wordsworth, Samuel Taylor Coleridge, and John Keats on the other. Also during this period, the novel emerged as a form capable of bringing everyday life into the province of literature. From this point, the distinctive regions of England began to exert a powerful influence on many writers—such as Wordsworth in the Lake District, the Brontë sisters on the Yorkshire moors, Thomas Hardy in Dorset, D.H. Lawrence in the Midlands coalfields, and Charles Dickens in London. In the modern period, English literature demonstrated a remarkable capacity to absorb and transmute alien elements, taking into the mainstream of its tradition poets as Irish as William Butler Yeats, as Welsh as Dylan Thomas, or as securely in the classic line as the U.S. expatriates T.S. Eliot and Henry James. Massive immigration from the far reaches of the Commonwealth has in the late 20th century further diversified England's literary landscape and has itself become the subject of numerous novels and plays. (For further discussion, see ENGLISH LITERATURE.)

Architecture. English architecture has varied significantly by location according to readily available building materials. The typical Cotswold village, for example, consists of structures of the local silvery limestone with slate roofs. A honey-coloured stone was much used in Oxford, and a rusty ironstone is typical in northern Oxfordshire and Northamptonshire, along the line of an ironstone belt. Half-timber framing and thatch roofing are characteristic of the river valleys, and excellent clay provides the warm red brick of southern England. In recent years the easy transport of cheap but alien materials is to be blamed for many jarring intrusions into the harmonious towns and villages originally built mainly of local materials.

Stylistically, English architecture has been much influenced from abroad, but foreign styles take on an English aspect. The Gothic architecture of France was transformed into a characteristically English style by the delicate use of stone to provide a framework for walls that were almost all glass, culminating in triumphs of the perpendicular style, such as King's College Chapel at Cambridge. The European Renaissance influenced the buildings of Christopher Wren, yet his many London churches seem essentially English. The magnificent country houses of the 18th century are not mere importations of a foreign fashion but fit their landscape; and many such landscapes were designed by the great English garden and park designers William Kent, Lancelot ("Capability") Brown, and Humphry Repton.

In the 20th century, many industrial slums have been condemned for demolition, but much contemporary building that is adequate for habitation or work is drearily uninspired. The reconstruction of the war-damaged city areas provided opportunities for notable new architecture, and some original design and construction was undertaken; for example, the Barbican scheme in a large bombed area north of St. Paul's, London, and the new National Theatre on the south bank of the Thames. Outside London, notable projects include the Coventry precinct and cathedral by Sir Basil Spence and a batch of new universities founded during the 1960s, among them those near Brighton, Canterbury, Colchester, Norwich, and York.

Modern building trends

Visual arts. *Sculpture.* Apart from traces of decoration on standing stones and the "transplanted" art of Roman occupation, the history of sculpture in England is rooted in the Christian church. Monumental crosses of carved stone, similar to the Celtic crosses of Ireland, represent the earliest sculpture of Anglo-Saxon Christians. The tradition of relief carving attained its highest expression in the stonework of the Gothic cathedrals, such as that at Wells (*c.* 1225–40). The influences of Renaissance and Baroque sculpture on the Continent were slow to reach England. What borrowings there were prior to the 18th century remained ill-conceived and crudely executed. From the 1730s, however, the presence of first-rate foreign artists, together with the flowering of archaeology and the resulting accessibility of antique art, brought a new refinement to English sculpture. The Roman influence that precipitated Neoclassicism gave way in England to the Greek with the arrival of the "rescued" Parthenon sculptures, now known as the Elgin Marbles, in the early 1800s. While the Romantic movement of the 19th century, which assailed the academic restraint of Neoclassicism in all the arts, invested continental sculpture with an increasing subjectivity, as well as a broader range of subject matter, the sculptors of England pursued a more conservative path. Many free-standing public monuments—the descendants of sepulchral effigies—date from this period. Not until the 20th century did English sculptors break free of traditional bounds and attain a deeply personal mode of expression. The sculptors Henry Moore and Barbara Hepworth both came from Yorkshire, and something of the quality of moorland stone can be seen in their work. Nearly a century after Moore's birth in 1898, his abstract reclining figures remained the dominant images of modern English sculpture.

Painting. Painting in England emerged under the auspices of the church. From the 8th to the 14th century, the illumination of Gospel manuscripts developed from essentially abstract decoration, derived from Celtic motifs, to self-contained pictorial illustration, more in keeping with the continental style. In the 15th century, Italian innovations in perspective and composition began to appear in English work. The advent of printing during this period, however, rendered the labour-intensive illumination increasingly rare. English painting remained largely unaffected by the concerns of the Renaissance, and it was not until the 1630s, when Charles I employed the Flemish Baroque painters Peter Paul Rubens and Anthony Van Dyck in his court, that a broader artistic current reached England's shores. Even so, provincial themes and the genres of portrait and landscape continued to preoccupy English painters for the next 150 years.

Manuscript illumination

The foundation of the Royal Academy of Arts in 1768 provided a focal point for the currents of Neoclassicism in English architecture, sculpture, and painting. Under the aegis of the academy, painters rendered historical and mythological subjects with a bold linear clarity. Just as the strictures of Neoclassicism developed partly in reaction against the excesses of the Baroque and Rococo, Romanticism emerged partly in defiance of academic formality. Classical antiquity, however, particularly in its ruined state, continued to provide themes and imagery. The works of the poet and painter William Blake epitomize the spiritual preoccupations of the period. Advances in science inspired a renewed artistic interest in the natural world. John Constable and J.M.W. Turner anticipated the French Impressionist movement by more than half a cen-

tury in their landscape paintings charged with light and atmosphere. The early Romantic fascination with biblical and medieval themes resurged in the mid-19th century among the so-called Pre-Raphaelite painters, who combined technical precision with explicit moral content.

Pre-Raphaelite painters

The emergence of the artist-craftsman, as exemplified by the Pre-Raphaelite Edward Burne-Jones and the designer and social theorist William Morris, brought new vigour to the decorative arts in England. Their successors exhibited a strong affinity for the continental Art Nouveau movement. In later 20th-century painting, England has produced such notable figures as the Pop artist R.B. Kitaj (born in the United States), the Op artist Bridget Riley, and Francis Bacon (born in Dublin of English parents).

Performing arts. *Theatre.* Theatre is probably the performing art for which England is best known. Theatrical performance as such emerged during the Middle Ages in the form of mumming plays, which borrowed elements from wandering entertainers, from traditional and ancient folk agricultural rituals, and from such dances as the Morris dance (with its set character parts). Under the influence of Christianity, mumming plays gradually were absorbed by mystery plays (centred on the Passion of Christ).

In the 16th century, when England's King Henry VIII rejected Rome and formed a national church, Latin theatrical traditions also were rejected; consequently, the Elizabethan and Jacobean ages forged a distinctive tradition and produced some extraordinary and highly influential playwrights, particularly Christopher Marlowe, Shakespeare, and Ben Jonson. A later influence on theatre in England was the rise in the 19th century of the actor-manager, the greatest being Henry Irving. That England remains one of the foremost contributors to world theatre can be seen in its lively theatrical institutions, such as the Royal Shakespeare Company (1864); in the ubiquity of its playwrights and actors; and in the great number of theatres that flourish in London's celebrated West End district. (For further discussion, see THEATRE, THE HISTORY OF WESTERN.)

Film. The making of motion pictures in Britain was first distinguished in the late 1940s and early '50s, when a series of social comedies made by Ealing Studios, such as *Kind Hearts and Coronets* and *Passport to Pimlico,* brought international acclaim to the British film industry. The nearness of film studios to the London stages allows directors and actors to pursue careers in both mediums to an extent unknown in the United States. (For further discussion, see MOTION PICTURES.)

Music. The beginnings of art music in England can be traced to plainchant. With the aid of monks and troubadours traveling throughout Europe, musical forms of many regions were freely intermingled and spread quickly. In the 16th and 17th centuries, England produced many notable composers, among them John Dowland, Thomas Morley, Thomas Tallis, and, perhaps greatest of all, William Byrd. The musical stature of the Baroque composers Henry Purcell and George Frideric Handel remains unquestioned. Music in England reached another peak in the late 19th century, when comic opera attained near perfection in the work of William Gilbert and Arthur Sullivan. Later significant composers include Edward Elgar, Gustav Holst, William Walton, and Benjamin Britten.

Opera is regularly performed by the Royal Opera at Covent Garden, London, by the English National Opera, and by other companies. A world-renowned opera festival is held annually at Glyndebourne, and music festivals of many other types thrive. England also has a number of orchestras, chamber groups, choruses, and cathedral choirs.

While more popular 20th-century music, such as the rock music of the Beatles and the Rolling Stones—as well as more contemporary music groups—may seem pervasive, it is English folk music, exemplified by ballads, sea chanteys, children's game songs, carols, and street cries, that truly has had the greater impact; its influence can be readily distinguished in the folk music, and even the hymnody, of the United States, Canada, and other former colonies.

20th-century music

Dance. Closely associated with song in folk tradition, folk dances have their origins in many of the same sources—mummers' dances, masques, and assorted ancient rituals of birth, courtship, war, death, and rebirth. In England remnants of early popular dances in the form of sword dances, Morris dances, and country dances remain popular participatory entertainment. From the 14th to the 17th century, performance-oriented dances, including court dances and dances developed for the stage, were much in evidence in more sophisticated circles of society. Though dancing masters and ballet as such were in existence from the 18th century, a native impulse toward the ballet only really began to take hold in England in the early 20th century, when Ninette de Valois and Lilian Baylis established the Vic-Wells Ballet (now the Royal Ballet) and Marie Rambert formed the Ballet Club (now Dance Rambert). These highly talented women fostered ballet and its offshoot, modern dance. With their leadership, England advanced to the forefront of dance in the 20th century, producing such internationally known artists as Frederick Ashton, Anton Dolin, Margot Fonteyn, Kenneth MacMillan, Alicia Markova, Bronisława Nijinska, and Antony Tudor.

Cultural institutions. All manner of general and esoteric societies, institutions, museums, and foundations can be found in England. One of its more prestigious learned societies is the Royal Society (1660), which awards fellowships, medals, and endowed lectureships based on scientific and technological achievements. The British Museum contains a wealth of archaeological and ethnographic specimens; its extensive library—containing ancient and medieval manuscripts and papyruses—was merged in 1973 with several other holdings to form the British Library, a centralized repository. The Zoological Society of London maintains the London Zoo and also conducts research, publishes journals, and supports a large zoological library. The Royal Botanic Gardens at Kew are significant both as a research institute and as one of England's many places of great natural beauty. There are also notable university libraries at the University of Cambridge and at the University of Oxford (the Bodleian Library).

Art galleries abound in England. The best-known are based in London and include the National Gallery, the Victoria and Albert Museum, the National Portrait Gallery, the Tate Gallery (with superb collections of John Constable and the Pre-Raphaelites), and the Wallace Collection.

Art galleries

Leisure. Though England has a lively cultural life, its characteristic pursuits are of a more popular kind. The exploitation of leisure is increasingly the concern of commerce: holiday camps, foreign holiday package tours, gambling of many kinds from bingo to horse-race betting, and the transformation of the traditional English pub by trendy interior decoration. The English weekend is the occasion for countryside trips and for outdoor activities from fishing to mountaineering. England gave to the world the sports of cricket, association football, and rugby football, but team and spectator sports tend to be giving way to more individualistic activities. Despite persistent commercial tempting to do something else, the English remain a stay-at-home people. Domestic comforts, epitomized in the cozy charm of cottages and gardens and the pervasive ritual of afternoon tea, continue to figure prominently in the character of English life. (W.H.Th./Ed.)

History

The history of England is given in the section above, *History of England and Great Britain.*

SCOTLAND

Scotland, the most northerly of the four parts of the United Kingdom, occupies about one-third of the island of Great Britain. It is bounded by England in the south and on the other three sides by sea: by the Atlantic Ocean on the west and north and by the North Sea on the east. Its mainland area is 28,269 square miles (73,217 square kilometres); including inhabited islands, it has an area of 30,418 square miles. The west coast is fringed by deep indentations (sea lochs or fjords) and by numerous islands, varying in size from mere rocks to the relatively large landmasses of Lewis and Harris, Skye, and Mull. The island clusters of the Orkneys and the Shetlands lie to the north. At its greatest length, measured from Cape Wrath to the Mull of Galloway, the mainland of Scotland extends to 274 miles (441 kilometres), while the maximum breadth, measured from Applecross, in the western Highland region, to Buchan Ness, in the eastern Grampian region, is 154 miles. But, because of the deep penetration of the sea in the sea lochs and firths (estuaries), most places are within 40 to 50 miles of the sea, and only 30 miles of land separate the Firth of Clyde and the Firth of Forth, the two great estuarine inlets on the west and east coasts, respectively.

The name Scotland (in Latin, Scotia) derives from the Scots, a Celtic people from Ireland who settled on the west coast in about the 5th century. The name Caledonia has often been applied to Scotland, especially in poetry. It is derived from the Roman name, Caledonii, of a tribe in the northern part of what is now Scotland. The kingdom of the Scots gradually gained control over neighbouring peoples until, by the 11th century, they ruled over roughly the country's modern mainland area. Medieval struggles for independence from England were successful, but in 1603 the king of Scots became king also of England, and in 1707 Scotland's parliament was joined to that of England. Thus Scotland no longer has a separate legislature or executive, nor diplomatic or consular representation abroad, and its economy is integrated into that of the rest of Britain. It does, however, have a separate administration, and certain important aspects of national life were preserved at the Union of 1707, notably its radically different legal and educational systems and its Presbyterian national church.

Above all, Scotland has retained much of its cultural identity. Superficially, the external perception of this may descend to an image of whiskey-swilling, tartan-clad Highlanders in mist-enshrouded castles, looking backward to bloody battles and romantic tales. But the tenacity of native culture has a deeper reality: in political and social attitudes distinct from those south of the border, in the strength of Scottish literature (still flourishing in three languages—English, Gaelic, and Scots), and in a musical and folktale tradition that survives to the present day. Though Scotland's population is only about five million, many millions abroad proudly claim Scottish descent and keep some of these traditions alive after many generations.

Physical and human geography

THE LAND

Relief. Scotland is traditionally divided into three geographic areas: the Highlands in the north, the Central Lowlands, and the Southern Uplands. A glance at a map will show that lowland areas are not confined to the central belt and indeed extend along the greater part of the eastern seaboard. The east coast also contrasts with the west in its smoother outline, thus creating an east–west distinction in topography as well as a north–south one.

The Great Glen

The Highlands are bisected by the fault line of the Great Glen, which is occupied by a series of lochs (lakes), the largest of which is Loch Ness, famous for its probably mythical monster. North of the Great Glen is an ancient plateau, which, through long erosion, has been cut into a series of peaks of fairly uniform height separated by glens (valleys) carved out by glaciers. The northwestern fringe of the mainland is particularly barren, the Lewisian gneisses having been worn down by severe glaciation to produce a hummocky landscape, dotted by small lochs and rocks protruding from thin, acidic soil. The landscape is varied by spectacular Torridonian sandstone mountains, weathered into sheer cliffs, rock terraces, and pinnacles. Southeast of the Great Glen are the Grampian Mountains, which were also formed by glaciation, though there are intrusions such as the granitic masses of the Cairngorm Mountains. The Grampians are on the whole less rocky and rugged than the mountains of the northwest, being more rounded and grassy with wider areas of plateau. But many have cliffs and pinnacles that provide challenges for the mountaineer, and the area contains Britain's highest mountains, including Ben Nevis (4,406 feet; 1,343 metres) at the southern end of the Great Glen. There are some flatter areas, the most striking being the Moor of Rannoch, a bleak expanse of bogs and granitic rocks, with narrow, deep lochs, such as Loch Rannoch and Loch Ericht. The southeastern margin of the Highlands is clearly marked by the Highland Boundary Fault, running northeast to southwest from Stonehaven, just south of Aberdeen, to Helensburgh on the River Clyde, and passing through Loch Lomond, Scotland's largest stretch of inland water. The southern boundary of the Central Lowlands is not such a continuous escarpment, but the fault beginning in the northeast with the Lammermuir and Moorfoot hills and extending to Glen App, in the southwest, is a distinct dividing line. In some ways the label Lowlands is a misnomer, for, although this part of Scotland is low by comparison with adjoining areas, it is by no means flat. The landscape includes hills such as the Sidlaws, the Ochils, the Campsies, and the Pentlands, composed of volcanic rocks rising as high as 1,898 feet (579 metres). The Southern Uplands are not so high as the Highlands. Glaciation has resulted in narrow, flat valleys separating rolling mountains. To the east of Nithsdale the hills are rounded, gently sloping, and grass-covered, providing excellent grazing for sheep, and they open out along the valley of the lower Tweed into the rich farming land of the Merse. To the west of Nithsdale the landscape is rougher, with granitic intrusions around Loch Doon, and the soil is more peaty and wet. The high moorlands and hills, of which Merrick (2,766 feet) is the highest, are also suitable for sheep farming. The uplands slope toward the coastal plains along the Solway Firth in the south and to The Machars and the Mull of Galloway farther west.

Soil and drainage. With Scotland's diversity in geologic structure, relief, and weather, the character of the soil varies greatly. In the northwest, the Hebrides, the Shetlands, and other areas, where the geologic substratum is ancient rock, resistant to weather, the soil is poor and cultivation is possible only at river mouths, glens, and coastal strips. On the west coast of some Hebridean islands, however, there are stretches of calcareous sand (the machair) suitable for farming. On moors and hills there is widespread peat. Where there is good soil for arable farming, as in the Orkneys, the eastern Highland region, the northeastern coastal plain, and the Lowlands, it has largely been derived from old red sandstone and younger rocks. Upward and eastward tilting of the Highlands some 50 million years ago formed a watershed near the west coast. Eastward drainage is dominant, but deeply glaciated rock basins in the northern Highlands have led to the presence of numerous large lochs. There are fewer lochs in the Grampian Mountains, although the large lochs of Ericht, Rannoch, and Tay are in the area. Well-graded rivers such as the Dee, the Don, and the Spey meander eastward and northeastward to the North Sea. The Tay and Forth emerge from the southern Grampians to flow out of the eastern Lowlands in two large estuaries. The Clyde and the Tweed both rise in the Southern Uplands, the one flowing to the west into the Firth of Clyde and the other to the east into the North Sea, while the Nith, the Annan, and a few other rivers run south into the Sol-

Peat

way Firth. Lochs are numerous in the Highlands, ranging from moraine-dammed lochans (pools) in mountain corries (cirques) to large and deep lochs filling rock basins. In the Lowlands and Southern Uplands, lochs are shallower and less numerous.

Climate. Scotland has a temperate oceanic climate, milder than might be expected from its latitude. Despite its small area, there are considerable variations. Rainfall is greatest in the mountainous areas of the west, as prevailing winds blow from the southwest and come laden with moisture from the Atlantic. East winds are common only in winter and spring, when cold, dry continental air masses envelop the east coast. Hence, the west tends to be milder in winter, with less frost and with snow seldom lying long at lower altitudes, but it is damper and cloudier than the east in summer. Tiree, in the Inner Hebrides off the west coast, has a mean temperature in winter of 41° F (5° C) in the coldest month (as high as southeastern England), whereas Dundee, on the east coast, has 37° F (2.8° C). Dundee's mean temperature in the warmest month is 59° F (15° C) and Tiree's 57° F (13.9° C). There is a smaller range of temperatures over the year in Scotland than in southern England. Rainfall varies remarkably. Some two-thirds of Scotland exceeds 40 inches (1,000 millimetres) annually, the average for Britain, with 142 inches in the Ben Nevis area and somewhat more near Loch Quoich farther to the northwest. In the flat Outer Hebrides, rainfall is lower, as are annual averages in the east, the Moray Firth having less than 25 inches and Dundee less than 32 inches. A significant amount of snow falls above 1,500 feet in the Highlands in winter.

Rainfall

Plant and animal life. Lower ground, up to about 1,500 feet, was once covered with natural forests, which have been cleared in the course of centuries and replaced in some areas by introduced trees, plants, and crops. Survivals of the original forest are found sporadically throughout the Highlands, for example in the pinewoods of Rothiemurchus in the Spey Valley. Grass and heather cover most of the Grampians and Southern Uplands, where the soil is not so wet and dank as in the northwestern Highlands. On peaty soil, shrubs such as bearberry, crowberry, and blaeberry (bilberry) grow, as does bog cotton. Alpine and Arctic species flourish on the highest slopes and plateaus of the Grampians, including saxifrages, creeping azalea, and dwarf willows. Ben Lawers is noted for the wealth of its mountain flora.

For its size, Scotland is rich in animal life. Herds of red deer graze in the corries and remote glens, and their population is estimated at about 300,000. Although formerly a woodland species, they are now found mainly on higher ground, but roe deer still inhabit the woods, along with sika and fallow deer, both introduced species, in some areas. Foxes and badgers are widespread, and the number of wildcats is thought to be increasing. Rabbits were earlier decimated by the disease of myxomatosis but are now recovering to earlier numbers. Pine marten, otters, and mountain and brown hares are among other wild mammals. A few ospreys nest in Scotland, and golden eagles, buzzards, peregrine falcons, and kestrels are the most notable of resident birds of prey. The red grouse, the Scottish subspecies of the willow grouse, has long been hunted for sport. Other species of grouse include the ptarmigan, found only at higher altitudes, and the large capercaillie, reintroduced into Scottish pinewoods. Large numbers of seabirds, such as gannets, fulmars, guillemots, and various types of gull, breed on cliffs and on the isolated rocks known as stacks around the magnificent coasts. Almost half the world's Atlantic, or gray, seals breed in Scottish waters, especially around the Northern and Western Isles, as do numerous common seals; dolphins and porpoises are regularly seen and whales occasionally, especially on the west coast.

Settlement patterns. In early times, mountains, rivers, and seas divided the people into self-sufficient communities, which developed a strong sense of identity. This sense has been eroded by social mobility, modern transport, broadcasting, and other standardizing influences, and by a general move from a rural to an urban way of life. Yet vestiges of regional consciousness linger. The Shetland islanders speak of Scotland with detachment. The Galloway area in the southwest, cut off by hills from the rest of the country, has a vigorous regional patriotism. The Gaelic-speaking people of the Hebrides and western Highlands find their language a bond of community. The northeast has its own local traditions, embodied especially in a still vigorous Scots dialect, and Borderers celebrate their local festivals with fervour. The most thickly populated rural areas are those with the best farming land, such as in Lothian region and in the northeast. The Highlands once nourished a large population, but continuous emigration since the 18th century has caused it to dwindle. Now settlements in the Highlands are mostly crofting townships; that is, small farms of a few acres grouped together in an irregular manner. The old pattern of crofting was one of communities practicing a kind of cooperative farming, with strips of common land allotted annually to individuals. Examples of the old system survive, but now crofters have their own arable land fenced in, while they share the common grazing land. In Lothian region and other areas of high farming, the communal farm has long been replaced by single farms with steadings (farmsteads) and workers' houses. Scotland is noticeably lacking in those old villages that evolved in England from medieval hamlets of joint tenants. Some planned villages were built by enterprising landowners in the 18th century. Burghs, often little bigger than villages, were mostly set up as trading centres, ports, or river crossings or to command entrances to mountain passes. Around the east and northeast coast, there are many surviving small towns that were once obliged to be self-contained in consumer industries and burghal institutions because of poor land transport. Growth of industry and transport has produced urbanization. Edinburgh, Dundee, and Aberdeen are centres of administration, commerce, and industry for their areas, but only central Clydeside, including Glasgow with its satellite towns, is large enough to deserve the official title of conurbation.

Crofting

THE PEOPLE

For many centuries there was continual strife between the Celtic Scots of the Highlands and the western islands and the Anglo-Saxons of the Lowlands. Only in the 20th century has the mixture been widely seen as a basis for a rich unified Scottish culture; the people of Shetland and Orkney have tended to remain apart from both of these elements and to look to Scandinavia as the mirror of their Norse heritage. Important immigrant groups have arrived, most notably Irish labour; there have also been significant groups of Jews, Lithuanians, Italians, and, after World War II, Poles and others, as well as a more recent influx of Asians, especially from Pakistan.

Scotland is remarkably free from racial and religious strife. The Church of Scotland is the established religion and largest communion, though membership has been steadily declining. It is presbyterian in structure and evangelical in doctrine. It is controlled by a hierarchy of church courts, from the kirk session (governing the affairs of a congregation), through the presbytery (covering a group of parishes) and the synod, to the General Assembly, at which clergy and lay representatives meet annually in Edinburgh. The Roman Catholic church is organized into two archdioceses and six dioceses. Roman Catholics have their own schools, built and staffed from public funds on the same terms as the state schools. The Scottish Episcopal Church is also significant, and there are congregations of other denominations, such as the Free Church of Scotland, Baptists, Congregationalists, Methodists, and Unitarians. In addition, immigrants have brought their diverse religions; Glasgow, for instance, now has several synagogues and mosques and a Buddhist centre.

Church of Scotland

Language. Scotland's linguistic heritage is complex. Though the vast majority now speak English, two other languages, Gaelic and Scots, still have wide influence. Gaelic, the Celtic language brought from Ireland by the Scots, is now spoken by a small minority mainly in the Western Isles and western Highlands, with pockets elsewhere, especially in Glasgow. Although it now faces a strong possibility of extinction, interest in Gaelic has

Gaelic and Scots

increased in recent years, and its literature flourishes as never before. Scots was originally a form of Old English that diverged from southern forms of the language in the Middle Ages, becoming a separate national tongue by the 15th century; political and other factors, notably union with England, caused English gradually to be adopted as the official and standard language, but Scots survives in the dialects of the Lowland areas, in a vigorous tradition of poetry and drama and in aspects of the English spoken by most Scots. Both Gaelic and Scots are recorded and supported by major works of scholarship, the *Linguistic Survey of Scotland* (1975–86); *The Scottish National Dictionary* (1931–76); *A Dictionary of the Older Scottish Tongue* (1931); and *The Historical Dictionary of Scottish Gaelic,* begun in 1966.

Demographic trends. While the land area of Scotland makes up about a third of that of the United Kingdom, its population constitutes only 9 percent, with the greatest concentration in the central belt. Since the mid-1960s there has been a significant shift in age structure, with an increase in the older age groups (65 and over) and an even more marked decrease in the young (aged 15 and under). North Sea oil has brought many people to the northeast and north, not only from other parts of Scotland and the United Kingdom but also from other countries, notably the United States. Large-scale emigration of Scots to countries such as Canada, the United States, and Australia has dwindled.

THE ECONOMY

Scotland's economy has in recent times shared in acute form the problems besetting many European countries, brought about by rapid changes that include the move away from heavy industries. Unemployment is a serious problem, especially in those areas where major industries have declined, and it is consistently higher than in the rest of the United Kingdom. Successive governments have made efforts to improve these conditions by a variety of measures. Chief among these is the Scottish Development Agency, which was set up in 1975 to encourage industry. The Highlands and Islands Development Board had been established 10 years earlier to carry out similar functions in the more remote regions.

Resources. Until recently, Scotland's chief mineral resource was coal. The industry reached its peak production of 43 million tons in 1913 but has since declined drastically. In particular, deep mining has become largely uneconomical, and by the late 1980s numbers employed had dwindled to a few thousand. Other minerals that have been worked intermittently include gold, silver, chromite, diatomite, and dolomite, but none has been successfully exploited. Though peat is available to a depth of two feet or more over some 1.7 million acres (688,000 hectares), its economic value is limited. It is still burned for fuel in the Highlands, but the time and labour involved in cutting and drying in an uncertain climate have led to decreasing use.

North Sea oil

During the 1970s a new Scottish resource, North Sea oil, was developed. The oil fields lie mostly in Scottish waters, but the British government holds their ownership and receives all the revenue yield. The oil has been located and extracted by large companies, most with the aid of U.S. technology. Aberdeen is the centre of the oil industry, and the Shetlands have also benefited from discoveries in adjacent waters. Tens of thousands of jobs were created in Scotland by onshore oil-related enterprises, such as building oil production platforms and servicing North Sea operators, though the new-found prosperity fell back when oil revenues were severely reduced in 1986. Natural gas from North Sea wells has replaced manufactured gas in Scotland.

Water is a valuable resource, especially for generating electricity. The North of Scotland Hydro-Electric Board (NSHEB) was set up in 1943 to build dams and power stations. It operates several hydroelectric stations and has pumped storage schemes by which electricity generated in off-peak periods may be used to pump water to a higher dam, from which it descends at peak periods to operate the turbogenerators. The NSHEB has also coal- and oil-fired stations. The South of Scotland Electricity Board (SSEB) has reduced its use of coal as a source of primary fuel, and nuclear generation, with the commissioning of the nuclear station at Torness, east of Edinburgh, now accounts for the major portion of Scotland's output of electricity. At Dounreay, in the Highland region, the United Kingdom Atomic Energy Authority has been carrying out experiments with a fast breeder-reactor since 1959.

Agriculture, forestry, and fishing. As an economic resource wild animals, birds, and river fishes are of minor importance, though deer stalking, grouse shooting, and fishing provide employment in parts of the Highlands in which other activities are hardly possible. Venison is exported to the European mainland, and deer farming produces an increasing amount of meat.

Fishing. The seafood industry continues to play a vital role in Scotland's economy. More than two-thirds of the total British fish and shellfish catch is now landed into Scottish ports. Peterhead ranks as the European Communities' top whitefish-landing port, and the Grampian region is one of the United Kingdom's main centres of fish processing. Haddock, cod, herring, sole, and mackerel are the main species landed. Nephrops (langoustine) is the most important shellfish, though scallop, queen scallop, lobster, and several crab varieties are also important. There is some commercial salmon fishing on the west coast, and in recent years fish farming has been developed, especially of salmon and shellfish along the coast and trout in the inland lochs.

Agriculture. In terms of productivity, no industry has made greater progress than agriculture in postwar years. Owing to mechanization, the labour force has fallen from about 88,000 in 1951 to barely a third of that number, though some casual and part-time workers are also employed. Most of Scotland is hilly or marginal land, with hill-sheep farming predominating, particularly in the Southern Uplands and in the Highlands; in the northeast livestock raising is dominant. In the southwest dairy farming suits the wetter, milder climate and has a convenient market in the central Clydeside conurbation. Arable cropping is mainly found along the eastern seaboard. Barley and wheat are the main cereals; the acreage devoted to potatoes, though substantial, has declined. New crops, such as rapeseed, have increased considerably, while oat production has fallen and been replaced by barley as the main cereal for livestock feed. Raspberry growing is mainly concentrated in the Tayside region. Tomatoes are still grown in greenhouses in the Clyde Valley, but the business has declined. The output of turnips and hay for livestock feeding has fallen, being replaced by an increase in grass silage. In animal production, the most striking feature has been the rise in the number of cattle and, to a lesser extent, sheep; there has also been expansion in pig and poultry production. Crofting is a special section of the agricultural scene that has to be supplemented by other work, such as forestry, road work, catering for tourists, and weaving. Though there are thousands of crofts in the northern area, many of them are no longer cultivated.

Principal crops

Forestry. Forestry is an expanding industry, which has helped retain the population in rural areas. It is managed by the Forestry Commission, a public body, and by private landowners, including forestry companies. Although the Forestry Commission plants throughout the country, it plays a particularly important role in Highland development. The trees it grows are mainly conifers, including Sitka spruce, Norway spruce, Scotch pine, European larch, and Douglas fir.

Industry. In its industrial heyday Scotland's prosperity was based on such heavy industries as coal, steel, shipbuilding, and engineering. In more recent times these have been the industries most exposed to foreign competition and to changes of demand. The task of correcting Scotland's industrial balance by reducing dependence on heavy industries and replacing them with high-technology enterprises and those making consumer goods has been slow and laborious, but considerable progress has been made in diversifying the structure of industry and in modernizing it. As with coal, the history of steel and shipbuilding is one of a reduction in the number of plants and employ-

ees. The sale of the nationalized British Shipbuilders to the private sector resulted in a continued reduction in the number of major shipyards in Scotland. Throughout the 1980s the special facilities built to provide rigs and platforms for exploiting the North Sea oil and gas reserves experienced fluctuating demand, and some of them closed. Severe cuts in capacity were also made in the steel industry as the state-owned British Steel Corporation strove to meet the government's financial targets as a prelude to privatization. Though not matching the older manufactures in terms of employment, the computer, office equipment, and electronic industries have expanded, notably in the Fife, Lothian, and Strathclyde regions. Much of this investment has come from overseas, particularly the United States.

New industries

Printing and brewing formerly were well-established industries in Edinburgh and Glasgow but are now in decline. Distilleries in the Highlands and the northeast produce the whiskey for which Scotland is internationally famous. Despite crushing taxation on home consumption, whiskey sales have continued to increase, and its appeal in foreign countries remains high. Separate records of Scotland's exports are not kept, but it is known that whiskey's contribution to the British overseas trade is substantial. In the 1980s its total was exceeded by office machinery and data-processing equipment, and petroleum products are also important. Mechanical, electrical, and electronic engineering industries export much of their output, and the textile industries of the Borders region and the Harris tweed in the Hebrides also have a considerable, though reduced, export business. Since the mid-1960s there has been a marked shift in employment from manufacturing to service industries, including tourism. The majority of tourists come from other parts of Scotland or the United Kingdom, but considerably more than a million annually come from abroad.

Finance. Scotland had eight joint-stock banks until the 1950s, but, as a result of mergers, the number was reduced to three: the Bank of Scotland, the Royal Bank of Scotland, and the Clydesdale Bank, which retain the right to issue their own notes. Since the mid-1960s, there has been a substantial expansion of financial and business services, with Edinburgh becoming second only to London in this field. The banking sector has grown and has now expanded into North America and Europe. It has also pioneered new applications of technology. Merchant banking facilities are more widely available, and the services historically associated with Scotland, such as the management of unit and investment trusts and life funds, have expanded. One-third of Britain's investment trusts are managed in Edinburgh, Glasgow, and Dundee. They have large investments in North America and specialized knowledge of conditions there. Unit trusts are represented in Edinburgh, where some leading British insurance companies also have their headquarters.

Transportation. Public transport was formerly largely state-owned, but there has been a considerable move toward the private sector. Deregulation of bus services in the 1980s led to greater competition, and the Scottish Transport Group, formed in 1968 to control bus and steamer services on the west coast, was largely privatized. Proliferation of automobiles has made it difficult for omnibus companies to maintain profitable services in rural areas, where they are being either subsidized by local authorities and the government or withdrawn. Ship services from mainland ports to island towns have been curtailed and replaced by car ferries using short crossings; such ferries operate from several west coast towns to the Hebrides and other islands and from north and east coast ports to the Orkneys and Shetlands.

Recent decades have seen considerable improvement in the Scottish road and bridge network, with some main routes being brought up to motorway standard and many single-lane roads in the Highlands widened. Improvements in the east and north were speeded up to cope with increased traffic generated by North Sea oil production, and bridges have been built over the Cromarty and Moray firths.

Railway services have been severely reduced since 1948, when more than 3,000 miles of track were open to passenger and freight traffic. Many branch lines and stations have been closed, and the route mileage has shrunk to less than two-thirds of the former total. Diesel engines have replaced steam locomotives. Suburban lines from Glasgow on both sides of the Clyde and to Airdrie in the east have been electrified, as has the main line from London (Euston) to Glasgow, and further electrification is planned.

Scottish ports handle many more imports than exports, a large proportion of which is sent abroad via English ports. Glasgow, the largest port, is under the administration of the Clyde Port Authority. The Forth ports, including Grangemouth and Leith, are grouped under the Forth Ports Authority, while Dundee and Aberdeen are independent. There are also important oil ports in Shetland (Sullom Voe), Orkney (Flotta), and on the east coast. Greenock and Grangemouth are equipped for container traffic, and extensive improvement schemes have been carried out at Leith and other ports. Coastal trade has dwindled because of the competition of motor transport, and inland waterways have never been a commercial success.

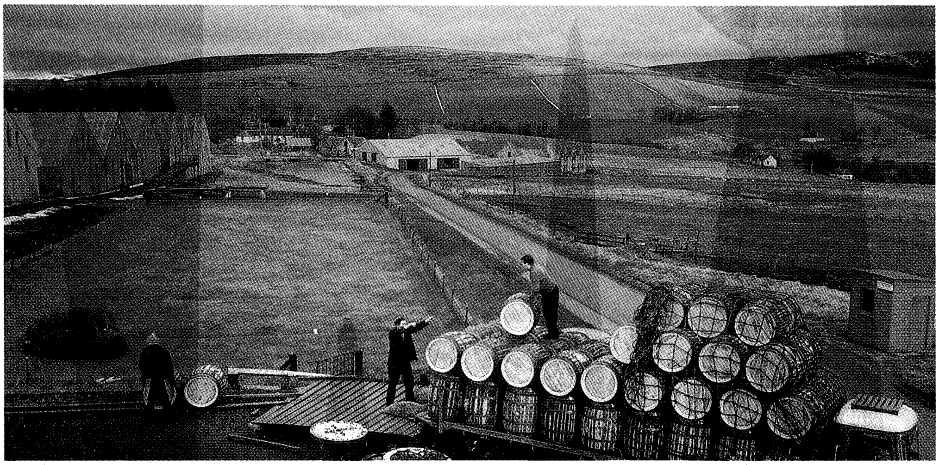
Major ports

There has been a marked increase in air travel, with some direct services to Europe, including a large number of charter flights. Scotland has major airports at Glasgow, Edinburgh, Aberdeen, and Prestwick on the west coast, the latter of which, being remarkably fog-free, is used for transatlantic flights.

ADMINISTRATION AND SOCIAL CONDITIONS

Government. Scotland is represented by 72 MPs in the United Kingdom House of Commons, and all Scottish peers are now entitled to sit in the House of Lords. The secretary of state for Scotland is responsible to Parliament for departments under his jurisdiction. These consist of the departments of Home and Health, Education, Agri-

© Brian Seed from TSW—CLICK/Chicago



Glenlivet whiskey distillery at Minmore, Scot.

culture and Fisheries, Industry, and Development. In his multifarious functions the secretary of state is assisted by a team of four ministers, one of whom is in the House of Lords. In the House of Commons, bills relating solely to Scotland are referred to the Scottish Grand Committee, which consists of all Scottish MPs. The Grand Committee has a general debate on a bill and usually sends it for detailed examination to the Scottish Standing Committee, a smaller body of Scottish MPs. The Select Committee on Scottish Affairs was discontinued by the government in 1988 because there were insufficient Scottish MPs to ensure a majority on the committee. Opposition parties set up an unofficial, extra-parliamentary committee to take its place.

The major British parties have separate organizations in Scotland, hold annual conferences, and take notice of Scottish problems and grievances. They hardly have distinctive policies for Scotland but rather adapt general proposals to Scottish purposes. Devolution, or self-government, became a lively issue during the late 20th century, though the question had agitated the political scene in earlier years. The Scottish National Party advocates complete independence, but since the 1970s devolution in the form of an assembly in Edinburgh has been a key political issue for all parties. A referendum in 1979 on the creation of such an assembly to govern domestic affairs but without full financial control failed to gain the support that the government demanded. The nonparty Campaign for a Scottish Assembly, however, has kept the issue alive.

Move for self-government

The local government structure of Scotland was reorganized in 1975. The number of local authorities was reduced and a two-tier system created. In place of the rather complex pattern of cities, burghs, and counties, the country was divided into nine regions: Borders, Central, Dumfries and Galloway, Fife, Grampian, Highland, Lothian, Strathclyde, and Tayside. Three island authorities—Orkney Islands, Shetland Islands, and Western Isles—have separate status. Power is distributed among the regions and the 53 districts into which they are divided. The regional authorities are responsible for social work, transport, education, police, and fire services, while the districts deal with such matters as housing, sanitation, libraries, and recreation facilities. Planning is shared by district and regional councils. Elections to these authorities are held every four years; most local contests are fought on party lines. Labour controls Strathclyde, the largest region, and is the predominant party in local as in parliamentary politics. Regional and district councils are financed partly by government grants and partly by local taxation. This was formerly levied by a system of rates on householders, commercial properties, and industrial plants, but in 1989 the household rates were replaced by a poll tax paid at a flat rate by every adult resident. Government grants, subject to annual negotiations, cover a proportion of the spending of local authorities.

Justice. In law Scotland has preserved its own system and courts. The lord advocate and the solicitor general for Scotland are the ministers responsible for justice; they advise the government on legal affairs and help to draft legislation. The country is divided into six sheriffdoms, each with a sheriff principal (chief judge) and a varying number of sheriffs. The most serious offenses triable by jury are reserved for the High Court of Justiciary, the supreme court for criminal cases. The judges are the same as those of the Court of Session, the supreme court for civil cases. An appeal may be directed to the House of Lords from the Court of Session, but not from the High Court of Justiciary. The Court of Session, consisting of the lord president, the lord justice clerk, and 22 other judges, sits in Edinburgh and is divided into Inner and Outer houses. The Outer House judges hear cases at first instance. The Inner House, of which there are two divisions, each of four judges, hears appeals from the Outer House and from inferior courts. The sheriff courts have a wide jurisdiction in civil cases, but certain actions, such as challenging government decisions, are reserved for the Court of Session. They also deal with most criminal offenses, with serious cases tried by jury. The police investigate cases of crime discovered by or made known to them, but the decision whether or not to prosecute is made by the lord advocate in the High Court and by procurator fiscals in the sheriff courts. District courts, presided over by lay judges, deal with minor criminal offenses. In 1971 a system of children's hearings was set up to deal with children in difficulties, whether from lack of parental control, criminal behaviour, or other cause.

High Court of Justiciary

Police forces. Since the secretary of state has a general responsibility for law and order, he shares control of the police forces with local authorities. The police committees provide the buildings and equipment needed by the forces. The secretary of state, assisted by the chief inspector of constabulary, is concerned with efficiency and discipline; he approves the appointment of chief constables and can make regulations on conditions of service.

Military services and defense. Scotland is the site of several key defense installations, including several belonging to the North Atlantic Treaty Organization (NATO) and a naval base for American nuclear submarines on the Holy Loch in the Firth of Clyde. The Royal Navy has a base at Rosyth on the Firth of Forth, and the Royal Air Force has stations at Kinloss, in Grampian region, and at Leuchars, in Fife region. Scottish infantry regiments of the army are still distinguished by their tartans, kilts for the Highland and trousers for the Lowland regiments. Recent amalgamations have reduced the number to seven, of which the Royal Scots has the distinction of being the oldest infantry regiment in the British army.

The Royal Scots

Education. In Scotland, education is supervised by the Scottish Education Department and administered by the education committees of the regional authorities. Unlike the English system, fee-charging, independent "public schools" play only a minor role in Scottish education. The number of state nursery schools, though increasing, remains insufficient; private schools and play groups help to fill the gap. School is compulsory to age 16. Teachers receive their professional training in colleges of education.

Since 1960 the number of universities has increased from four to eight. Of the original four, St. Andrews, Glasgow, and Aberdeen were all founded in the 15th century and Edinburgh in 1583. In the early 1960s, the University of Strathclyde in Glasgow and the Heriot-Watt University in Edinburgh, formerly technological colleges, were upgraded to universities, retaining their scientific and technological emphasis. The University of Dundee was separated from the University of St. Andrews, which had some departments in Dundee. The University of Stirling, the only completely new foundation, was opened in 1967. The new universities do not teach law and divinity but place most emphasis on science and technology and have close links with science-based industries in their neighbourhoods.

Health and welfare. Health and welfare services and housing are the joint responsibility of government departments and local authorities. The National Health Service is administered in Scotland by several Health Boards and the Common Services Agency; social work is administered mainly by local authorities. Scotland has a poor record in terms of certain diseases, with one of the highest incidences in Europe of heart disease and lung cancer.

CULTURAL LIFE

In spite of the threat of dominance by its more powerful partner, Scottish culture has remained remarkably vigorous in the 20th century. Its strength springs in part from the diverse strands which make up its background, in particular the sharp contacts with the European mainstream cultures. It has also been enriched by contacts with the European mainstream brought about by the mobility of the Scottish people from the Middle Ages on. All of the arts receive support from the Scottish Arts Council, which has a large measure of autonomy from the Arts Council of Great Britain. Scottish writers have the choice of three languages—English, Scots, and Gaelic. Hugh MacDiarmid, the poet, nationalist, and Marxist, gained an international reputation for his Scottish poetry, and others, such as Robert Garioch, followed his lead. Gaelic poets such as Sorley Maclean and Derick Thompson are highly esteemed, as is Iain Crichton Smith, also known for his novels in English. Painting and sculpture flourish, as

evidenced not only in official exhibitions but also in the burgeoning of many small galleries. In music the Scottish National Orchestra, Scottish Opera, and Scottish Ballet have achieved international standing. The annual Edinburgh International Festival, with its Fringe, has become one of the world's largest cultural events. Scotland has an unparalleled wealth of surviving traditional music, ranging from the work songs of the Hebrides to the ballads of the northeast. There has also been renewed interest in Scotland's traditional instruments, the bagpipe, the fiddle, and the clarsach (the small Celtic harp). All aspects of traditional culture are researched, archived, and taught in the School of Scottish Studies in the University of Edinburgh.

Edinburgh also houses cultural institutions such as the National Library of Scotland, which has a statutory right to receive copies of all books published in the United Kingdom and Ireland. The National Gallery of Scotland has paintings by many famous European artists in addition to works by Allan Ramsay, Sir Henry Raeburn, and other Scottish painters. The Scottish National Portrait Gallery portrays the principal personages in Scotland's history, and the Scottish National Gallery of Modern Art has works by contemporary European painters and sculptors, as well as those of native artists. The National Museums of Scotland contain archaeological and later evidence for the development of the material and domestic aspects of Scottish society and have extensive collections in their departments of art and archaeology, natural history, technology, and geology. These galleries and museums are the responsibility of the secretary of state and are maintained by public funds.

The press. Edinburgh was once one of the centres of the United Kingdom's publishing industry, but Scottish publishing declined drastically, especially in the postwar years, with the move of many major houses to London. Since the early 1970s, however, there has been a revival. Like the British press as a whole, the Scottish press has expanded as technological advance has cut costs. The *Scotsman,* published in Edinburgh, and the *Glasgow Herald* rank highest of the daily newspapers in antiquity and influence, though both are owned by multinational corporations with headquarters outside Scotland. *Scottish Field* and *Scots Magazine* are two well-established monthly magazines covering traditional, leisure, and historical interests.

Scottish newspapers

The British Broadcasting Corporation (BBC) has Scottish news and other programs, including some in Gaelic. Most broadcasting in that language is on radio. Radio Scotland has largely locally produced programs. There are three independent television companies and several independent radio companies.

Sports and recreation. Sports plays an important part in the life of Scotland. Association football (soccer) still has a wide following and is dominated by the Glasgow clubs, Rangers and Celtic. Rugby Union football (amateur) is played especially by private schools and by their former pupils, but in the towns of the Borders region it draws players and spectators from a wider social range. In the Highlands, shinty, a hockeylike game, is popular, Curling is another traditional sport, though temperatures are seldom low enough for it to be other than an indoor sport played on man-made ice. Golf, which originated in Scotland, is accessible not merely to the affluent through private clubs as in many countries but to most Scots through widespread public facilities. The Old Course at St. Andrews in Fife is the most famous of many excellent seaside courses. For hill-walking, rock-climbing, sailing, and canoeing, Scotland has outstanding natural advantages. Skiing facilities have been developed in the Cairngorms and other areas. Hunting, shooting, and fishing are traditionally the sports of the rich, but the last is popular with all classes. Other outdoor sports, such as tossing the caber (a heavy pole) and throwing the hammer are integral to the Highland games, at which pipe bands and Highland dancers (usually solo) also perform. Many Scots find these games and other traditions, such as Burns suppers (honouring Robert Burns) and eating haggis (a delicacy consisting of offal boiled in a sheep's stomach), a self-conscious parade of legendary characteristics that have little to do with ordinary Scottish life—a show put on, like national costumes, to gratify the expectations of tourists (encouraged by the royal family's annual appearance at the Braemar Gathering near Balmoral Castle). Scottish country dancing, however, is a pastime whose popularity has spread far beyond Scotland. (Ma.J.M./I.C.M.)

History

ANCIENT TIMES

Evidence of human settlement in the area later known as Scotland dates from the 3rd millennium BC. The earliest people, Mesolithic (Middle Stone Age) hunters and fishermen who probably reached Scotland via an ancient landbridge from the continent, were to be found on the west coast, near Oban, and as far south as Kirkcudbright, where their settlements are marked by large deposits of discarded mollusk shells. Remains suggest that settlers at the Forth estuary, in the area of modern Stirling, obtained meat from stranded whales. By early in the 2nd millennium BC, Neolithic (New Stone Age) farmers had begun cultivating cereals and keeping cattle and sheep. They made settlements on the west coast and as far north as Shetland. Many built collective chamber tombs, the example at Maeshowe in Orkney being the finest in Britain. A settlement of such people at Skara Brae in Orkney consists of a cluster of seven self-contained huts connected by covered galleries or alleys. The "Beaker folk," so called from the shape of their drinking vessels, came to eastern Scotland from northern Europe, probably beginning about 1800 BC. They buried their dead in individual graves and were pioneers in bronze working. The most impressive monuments of Bronze Age Scotland are the stone circles, presumably for religious ceremonies, such as those at Callanish in Lewis and Brodgar in Orkney, the latter being more than 300 feet (91 metres) in diameter.

From about 700 BC onward there was a distinct final period in Scottish prehistory. This period is the subject of current archaeological controversy, with somewhat less stress than in the past being placed on the importance of the introduction of iron fabrication or on the impact of large new groups of iron-using settlers. One key occurrence in the middle of the 1st millennium was the change from a relatively warm and dry climate to one that was cooler and wetter. In terms of technology, this period was marked by the appearance of hill forts, defensive structures having stone ramparts with an internal frame of timber; a good example is at Abernethy near the Tay. Some of these forts have been dated to the 7th and 6th centuries BC, which might suggest that they were adopted by already established tribes rather than introduced by incomers. Massive decorated bronze armlets with Celtic ornamentation, found in northeastern Scotland and dated to the period AD 50–150, suggest that chieftains from outside may have come to these tribes at this period, displaced from farther south by fresh settlers from the Continent and then by the advent of the Romans in AD 43. From 100 BC the "brochs" appeared in the extreme north of Scotland and the northern isles. These were high, round towers, which at Mousa in Shetland stand almost 50 feet (15 metres) in height. The broch dwellers may have carried on intermittent warfare with the fort builders of farther south. On the other hand, the two types of structures may not represent two wholly distinct cultures, and the two peoples may have together constituted the ancestors of the people later known as the Picts.

The "brochs"

The houses of this people were circular, sometimes standing alone, sometimes in groups of 15 or more, as at Hayhope Knowe in the Cheviot Hills on the border between modern Scotland and England. Some single steadings, set in bogs or on lakesides, are called crannogs. Corn growing was probably of minor importance in the economy; the people were pastoralists and food gatherers. They were ruled by a warrior aristocracy whose bronze and iron parade equipment has, in a few instances, survived.

Roman penetration and the Dark Age peoples. Gnaeus Julius Agricola, the Roman governor of Britain (AD 77–84), was the first Roman general to operate extensively in Scotland. He defeated the natives at Mons Graupius, possibly in Banffshire, probably in AD 84. In the following

year he was recalled, and his policy of containing the hostile tribes within the Highland zone, which he had marked by building a legionary fortress at Inchtuthil in Strathmore, was not continued. His tactics were logical, if Scotland was to be subdued, but probably required the commitment of more troops than the overall strategy of the Roman Empire could afford. The only other period in which a forward policy was attempted was between about 144 and about 190, when a turf wall, the Antonine Wall (named after the emperor Antoninus Pius), was manned between the Forth and the Clyde.

The still-impressive stone structure known as Hadrian's Wall had been built between the Tyne and Solway Firth in the years 122–128, and it was to be the permanent northern frontier of Roman Britain. After a northern rising, the emperor Severus supervised the restoring of the Hadrianic line in the years 209–211, and thereafter southeastern Scotland seems to have enjoyed almost a century of peace. In the 4th century there were successive raids from north of the Wall and periodic withdrawals of Roman troops to the Continent. Despite increasing use of native buffer-states in front of the Wall, the Romans found their frontier indefensible by the end of the 4th century.

At Housesteads, at about the midpoint of Hadrian's Wall, archaeologists have uncovered a market where northern natives exchanged cattle and hides for Roman products: in this way some Roman wares, and possibly more general cultural influences, found their way north, but the scale of this commerce was probably small. Roman civilization, typified by the towns and villas, or country houses, of southern Britain, was unknown in Scotland, which as a whole was never dominated by the Romans, or even strongly influenced by them.

The peoples of early Scotland

From about AD 400 there was a long period for which written evidence is scanty. Four peoples—the Picts, the Scots, the Britons, and the Angles—were eventually to merge and thus form the kingdom of Scots.

The Picts occupied Scotland north of the Forth. Their identity has been much debated, but they possessed a distinctive culture, seen particularly in their carved symbol stones. Their original language, presumably non-Indo-European, has disappeared; some Picts probably spoke a Brythonic Celtic language. Pictish unity may have been impaired by their apparent tradition of matrilineal succession to the throne.

The Scots, from Dalriada in northern Ireland, colonized the Argyll area, probably in the late 5th century. Their continuing connection with Ireland was a source of strength to them, and Scots and Irish Gaelic (Goidelic Celtic languages) did not become distinct from one another until the late Middle Ages. Scottish Dalriada soon extended its cultural as well as its military sway east and south, though one of its greatest kings, Aidan, was, in 603, defeated by the Angles at Degsastan near the later Scottish border.

The Britons, speaking a Brythonic Celtic language, colonized Scotland from farther south, probably from the first century BC onward. They lost control of southeastern Scotland to the Angles in the early 7th century AD. The British heroic poem *Gododdin* describes a stage in this process. The British kingdom of Strathclyde in southwestern Scotland remained, with its capital at Dumbarton.

The Angles were Teutonic-speaking invaders from across the North Sea. Settling from the 5th century, they had by the early 7th century created the kingdom of Northumbria, stretching from the Humber to the Forth. A decisive check to their northward advance was administered in 685 by the Picts at the Battle of Nechtansmere in Angus.

Christianity. Christianity was introduced to Scotland in late Roman times, and traditions of St. Ninian's evangelizing in the southwest have survived. He is a shadowy figure, and it is doubtful that his work extended very far north.

Christianity was firmly established throughout Scotland by the Celtic clergy, coming with the Scots settlers from Ireland, and possibly giving the Scots a decisive cultural advantage in the early unification of kingdoms. The Celtic church lacked a territorial organization of parishes and dioceses and a division between secular and regular clergy: its communities of missionary monks were ideal agents of conversion. The best-known figure, possibly the greatest, is St. Columba, who founded his monastery at Iona, an island of the Inner Hebrides, in 565; his life was written by Adamnan, abbot of Iona, within a century of his death. Columba is believed to have been influential in converting the Picts, and he did much to support the Scots king Aidan politically.

St. Columba

St. Aidan brought the Celtic church to Northumbria in the 630s, establishing his monastery at Lindisfarne. At the Synod of Whitby in 664 the king of Northumbria had to decide between the Celtic and the Roman styles of Christianity: he chose the latter. There had been differences over such observances as the dating of Easter, but there was no question of the Celtic monks' being regarded as schismatics. The *Ecclesiastical History of the English People* by Bede, a monk of Jarrow in Northumbria (died 735), is a first-rate source for the history of Dark Age Scotland and shows remarkable sympathy with the Celtic clergy, though Bede was a Roman monk.

In the early 8th century the church among the Picts and Scots accepted Roman usages on such questions as Easter. Nevertheless, the church in Scotland remained Celtic in many ways until the 11th century. Still dominated by its communities of clergy (who were called Célidé or Culdees), it clearly corresponded well to the tribal nature of society.

The Norse influence. Viking raids on the coasts of Britain began at the end of the 8th century, Lindisfarne and Iona being pillaged in the 790s. By the mid-9th century, Norse settlement of the western and northern isles and of Caithness and Sutherland had begun, probably owing largely to overpopulation on the west coast of Norway. During the 10th century, Orkney and Shetland were ruled by Norse earls nominally subject to Norway. In 1098 Magnus II Barefoot, king of Norway, successfully asserted his authority in the northern and western isles and made an agreement with the king of Scots on their respective spheres of influence. A mid-12th-century earl of Orkney, Ragnvald, built the great cathedral at Kirkwall in honour of his martyred uncle St. Magnus.

The Norse legacy to Scotland was long-lasting. In the mid-12th century there was a rising against the Norse in the west under a native leader, Somerled, who drove them from the greater part of mainland Argyll. A Norwegian expedition of 1263 under King Haakon IV failed to maintain the Norse presence in the Hebrides, and three years later they were ceded to Scotland by the Treaty of Perth. In 1468–69 the northern isles of Orkney and Shetland were pawned to Scotland as part of a marriage settlement with the crown of Denmark-Norway. A Scandinavian language, the Norn, was spoken in these Viking possessions, and some Norse linguistic influence is discernible in Shetland to the present day.

THE UNIFICATION OF THE KINGDOM

In 843 Kenneth I MacAlpin, king of Scots, also became king of the Picts and crushed resistance to his assuming the throne. Kenneth may have had a claim on the Pictish throne through the matrilineal law of succession; probably the Picts, too, had been weakened by Norse attacks. The Norse threat helped to weld together the new kingdom of Alba and to cause its heartlands to be located in eastern Scotland, the former Pictland, with Dunkeld becoming its religious capital. But within Alba it was the Scots who established a cultural and linguistic supremacy, no doubt merely confirming a tendency seen before 843.

Ascendancy of the Scots

As the English kingdom was consolidated, its kings, in the face of Norse attacks, found it useful to have an understanding with Alba. In 945 Edmund of England is said to have leased to Malcolm I of Alba the whole of Cumbria, probably an area including land on both sides of the western half of the later Anglo-Scottish border. In the late 10th century a similar arrangement seems to have been made for Lothian, the corresponding territory to the east. The Scots confirmed their hold on Lothian, from the Forth to the Tweed, when, about 1016, Malcolm II defeated a Northumbrian army at Carham. About the same time, Malcolm II placed his grandson Duncan I upon the throne of the British kingdom of Strathclyde. Duncan succeeded Malcolm in 1034 and brought Strathclyde into the kingdom of Scots. During the next two centuries the

Scots kings pushed their effective power north and west—William I was successful in the north and Alexander II in the west—until mainland Scotland became one political unit. Less discernible but as important was the way the various peoples grew together, though significant linguistic and other differences remained.

According to the Celtic system of succession, known as tanistry, a king could be succeeded by any male member of the *derbfine,* a family group of four generations: members of collateral branches seem to have been preferred to descendants, and the successor, or tanist, might be named in his predecessor's lifetime. This system, in practice, led to many successions by the killing of one's predecessor. Thus Duncan I was killed by his cousin Macbeth in 1040, and Macbeth was killed by Malcolm III Canmore, Duncan I's son, in 1057. Shakespeare freely adapted the story of Macbeth, who historically seems to have been a successful king and who may have gone on pilgrimage to Rome.

Up to the 11th century the unification was the work of a Scots Gaelic-speaking dynasty, and there is place-name evidence of the penetration of Gaelic south of the Forth. But from then on, the Teutonic English speech that had come to Scotland from the kingdom of Northumbria began to attain mastery, and Gaelic began its slow retreat north and west. This is not obscured by the fact that, from the 12th century onward, Anglo-Norman was for a time the speech of the leaders of society in England and Scotland alike. By the later Middle Ages, the language known to modern scholars as Old English had evolved into two separate languages, now called Middle English and Middle Scots, the latter with the court of the Stewart (Stuart) kings of Scots as its focus. After 1603, the increasing political and cultural assimilation of Scotland to England checked the further development of Scots as a separate language.

The persistence of distinctively Celtic institutions in post-12th-century Scotland is a more complex question, as will be seen from the way in which primogeniture replaced tanistry as the system of royal succession. It can be argued, however, that a Celtic stress on the family bond in society persisted throughout the Middle Ages and beyond—and not only in the Highlands, with its clan organization of society.

The royal succession

The development of the monarchy. Malcolm III Canmore (1058–93) came to the throne by disposing of his rivals and thereafter sought, in five unsuccessful raids, to extend his kingdom into northern England. Whereas his first wife, Ingibjorg, was the daughter of a Norse earl of Orkney, his second, Margaret, came from the Saxon royal house of England. With Margaret and her sons, Scotland entered a phase of being particularly receptive to cultural influence from the south. Margaret was a great patroness of the church but without altering its organization as her sons were to do.

On the death of Malcolm III on his last English raid, sustained attempts were made to prevent the application of the southern custom of succession by primogeniture. Both Malcolm's brother and Malcolm's son by his first marriage held the throne for short periods: but it was the three sons of Malcolm and Margaret who eventually established themselves—Edgar (1097–1107), Alexander I (1107–24), and David I (1124–53). Such was the force of Celtic reaction against southern influence that Edgar and Alexander I could be said to owe their thrones solely to English aid and were feudally subject to the English king. The descendants of Malcolm III's first marriage continued to trouble the ruling dynasty until the early 13th century, but the descendants of his second retained the throne. It happened that, until the late 13th century, the heir to the throne by primogeniture was always the obvious candidate. It is noteworthy that in charters of about 1145, David I's son Henry (who was to die before his father) is described as *rex designatus,* very much like the tanist of the Celtic system. It is thus very hard to date precisely the acceptance of southern custom as exemplified by primogeniture.

David I (1124–53). David I was by marriage a leading landowner in England and was well known at the English court. He was, nevertheless, an independent monarch, making Scotland strong by drawing on English cultural and organizational influences. Under him and his successors many Anglo-Norman families came to Scotland, and their members were rewarded with lands and offices. Among the most important were the Bruces in Annandale, the de Morvilles in Ayrshire and Lauderdale, and the Fitzalans, who became hereditary High Stewards and who, as the Stewart dynasty, were to inherit the throne, in Renfrewshire. (After the 16th century the Stewart dynasty was known by its French spelling, "Stuart.") Such men were often given large estates in outlying areas to bolster the king's authority where it was weak.

Anglo-Norman settlement in Scotland

The decentralized form of government and society that resulted was one of the many variants of what is known as feudalism, with tenants in chief holding lands, with jurisdiction over their inhabitants, from the king, in return for the performance of military and other services. An essentially new element in Scottish society was the written charter, setting out the rights and obligations involved in landholding. But the way in which the Anglo-Norman families, in their position as tenants in chief, were successfully grafted onto the existing society suggests that the Celtic and feudal social systems, although one stressed family bonds and the other legal contracts, were by no means mutually incompatible. The clan system of Highland Scotland became tinged with feudal influences, whereas Lowland Scottish feudalism retained a strong emphasis on the family.

David began to spread direct royal influence through the kingdom by the creation of the office of sheriff (*vicecomes*), a royal judge and administrator ruling an area of the kingdom from one of the royal castles. Centrally, a nucleus of government officials, such as the chancellor, the chamberlain, and the justiciar, was created by David and his successors; these officials, with other tenants in chief called to give advice, made up the royal court (Curia Regis). This body became formalized in various ways: by the mid-13th century it might meet as the king's council to discuss various types of business; and before the Wars of Independence (see below) the royal court in its capacity as the Supreme Court of Law was already being described as a Parliament. The almost total loss of all of the Scottish governmental records from before the early 14th century should not lead one to underestimate the efficiency of the Scottish kings' government in this period; historians have now done much to assemble the surviving royal documents from scattered sources.

Medieval economy and society. From David's time onward, the burghs, or incorporated towns, were created as centres of trade and small-scale manufacture in an overwhelmingly agrarian economy. At first, all burghs probably had equal rights. Later, however, royal burghs had, by their charters, the exclusive right of overseas trade, though tenants in chief could create burghs with local trade privileges. Burghs evolved their own law to govern trading transactions, and disputes could be referred to the Court of the Four Burghs (originally Berwick, Edinburgh, Roxburgh, and Stirling). Many of the original townspeople, or burgesses, were newcomers to Scotland. At Berwick, the great trading town of the 13th century, exporting the wool of the border monasteries, Flemish merchants had their own Red Hall, which they defended to the death against English attack in 1296. Besides commercial contacts with England, there is evidence of Scottish trading with the Low Countries and with Norway in the period before the Wars of Independence.

The church was decisively remodeled by David I and his successors. A clear division emerged between secular and regular clergy according to the normal western European pattern. A complete system of parishes and dioceses was established. But the system of "appropriating" the revenue of parish churches to central religious institutions meant that the top-heaviness in wealth and resources of the church in Scotland was a built-in feature of its existence until the Reformation. Kings and other great men vied in setting up monasteries. Alexander I had founded houses of Augustinian canons at Scone and Inchcolm, while among David's foundations were the Cistercian houses of Melrose and Newbattle and the Augustinian houses of Cambuskenneth and Holyrood. Augustinian canons

David's reorganization of the church

might also serve as the clergy of a cathedral, as they did at St. Andrews. Prominent foundations by the magnates included Walter Fitzalan's Cluniac house at Paisley and Hugh de Morville's Premonstratensian house at Dryburgh. Later royal foundations included the Benedictine house at Arbroath, established by William I.

From the standpoint of a later age, when the monasteries had lost their spiritual force, the piety of David I especially seemed a misapplication of royal resources. But the original monasteries, with their supply of trained manpower for royal service, their hospitality, and their learning, epitomized that stability which it was royal policy to achieve.

From at least 1072, the English church, particularly the archbishop of York, sought some control over the Scottish church; the Scottish church was weakened in the face of such a threat through having no metropolitan see. But, probably in 1192, the pope by the bull *Cum Universi* declared the Scottish church to be subject only to Rome; and in 1225 the bull *Quidam Vestrum* permitted the Scottish church, lacking a metropolitan see, to hold provincial councils by authority of Rome. Such councils, which might have served to check abuses, were, however, seldom held.

Cultural developments

It has been argued that the cultural developments encouraged by the church in pre-Reformation Scotland were not as great as might be expected, but this may be a false impression created because the manuscript evidence has failed to survive. The monasteries of Melrose and Holyrood had each a chronicle, and Adam of Dryburgh was an able theologian of the late 12th century. Surviving Romanesque churches show that Scotland partook of the common European architectural tradition of the time: good small examples are at Dalmeny, near Edinburgh, and at Leuchars, in Fife. Glasgow and Elgin cathedrals are noteworthy, and St. Andrews Cathedral is impressive even in its ruined state. There are also distinguished examples of castle architecture, such as Bothwell in Lanarkshire; and the castles of Argyll may reflect a distinctive mixture of influences, including Norse ones.

David I's successors. Malcolm IV (1153–65) was a fairly successful king, defeating Somerled when the latter, who had been triumphant over the Scandinavians in Argyll, turned against the kingdom of Scots. Malcolm's brother, William I the Lion (1165–1214), subdued much of the north and established royal castles there. After his capture on a raid into England, he was forced to become feudally subject to the English king by the Treaty of Falaise (1174); he was able, however, to buy back his kingdom's independence by the Quitclaim of Canterbury in 1189, though it should be emphasized that this document disposed of the Treaty of Falaise and not of the less precise claims of superiority over Scotland that English kings had put forward over the previous century. William's son, Alexander II (1214–49), subdued Argyll and was about to proceed against the Hebrides at the time of his death. His son, Alexander III (1249–86), brought these islands within the Scottish kingdom in 1266, adroitly fended off English claims to overlordship, and brought to Scotland the peace and prosperity typified by the commercial growth of Berwick. In the perspective of the subsequent Wars of Independence, it was inevitable that Scots should look back on his reign as a golden age.

THE WARS OF INDEPENDENCE

Competition for the throne. With the death, in 1286, of Alexander III and of his young granddaughter Margaret, the "Maid of Norway," four years later, almost two centuries of relatively amicable Anglo-Scottish relations came to an end. A complete uncertainty as to the proper succession to the throne provided Edward I of England and his successors with a chance to intervene in and then to assimilate Scotland. Though the two countries were feudal monarchies of a largely similar type, the English attempt was, in practice, too tactless to have any hope of success. Besides, the struggle for independence disclosed that a marked degree of national unity had arisen among the different peoples of Scotland. The Anglo-Scottish conflict thus begun gave Scotland a basic tendency—to seek self-sufficiency and at the same time to look to continental Europe for alliances and inspiration—that persisted at least until 1560.

The Maid of Norway

Before the death of the Maid of Norway, the Scottish interim government of "guardians" had agreed (by the Treaty of Birgham, 1290) that she should marry the heir of Edward I of England, though Scotland was to be preserved as a separate kingdom. After her death, 13 claimants for the Scottish crown emerged, most of them Scottish magnates. The Scots had initially no reason to suspect the motives of Edward I in undertaking to judge the various claims. It emerged, however, that Edward saw himself not as an outside arbitrator but as the feudal superior of the Scottish monarch and, therefore, able to dispose of Scotland as a fief. That Edward's interpretation was disingenuous is suggested by the fact that he had not invoked the old and vague English claims to superiority over Scotland while the Maid of Norway was still alive and had made a treaty with Scotland on a basis of equality, not as a feudal superior claiming rights of wardship and marriage over the Maid.

The claimants to the throne, who had much to lose by antagonizing Edward, generally agreed to acknowledge his superior lordship over Scotland. But a different answer to his claim to lordship was given by the "community of the realm" (the important laymen and churchmen of Scotland as a group), who declined to commit whoever was to be king of Scots on this issue and thus displayed a sophisticated sense of national unity.

Bruce and Balliol

Robert Bruce and John Balliol, descendants of a younger brother of Malcolm IV and William, emerged as the leading competitors, and in 1292 Edward I named the latter as king. When Edward sought to exert his overlordship by taking law cases on appeal from Scotland and by summoning Balliol to do military service for him in France, the Scots determined to resist. In 1295 they concluded an alliance with France, and in 1296 Edward's army marched north, sacking Berwick on its way.

Edward forced the submission of Balliol and of Scotland with ease. National resistance to English government of Scotland grew slowly thereafter and was led by William Wallace, a knight's son, in the absence of a leader from the magnates. Wallace defeated the English at Stirling Bridge in 1297 but lost at Falkirk the next year. He was executed in London in 1305, having shown that heroic leadership without social status was not enough. When Robert Bruce, grandson of the competitor, rose in revolt in 1306 and had himself crowned Robert I, he supplied the focus necessary for the considerable potential of national resistance.

Robert I the Bruce (1306–29). In several years of mixed fortunes thereafter, Robert had both the English and his opponents within Scotland to contend with. Edward I's death, in 1307, and the dissension in England under Edward II were assets that Bruce took full advantage of. He excelled as a statesman and as a military leader specializing in harrying tactics; it is ironic that he should be remembered best for the atypical set-piece battle that he incurred and won at Bannockburn in 1314. The Declaration of Arbroath of 1320 is perhaps more informative about his methods. Ostensibly a letter from the magnates of Scotland to the pope, pledging their support for King Robert, it seems in reality to have been framed by Bernard de Linton, Robert's chancellor. In committing Robert to see the independence struggle through, it likewise committed those who set their seals to it. Some of them were waverers in the national cause, whether or not Robert had proof of this at the time, and his hand was now strengthened against them.

Robert I secured from England a recognition of Scotland's independence by the Treaty of Northampton in 1328; the following year the pope granted to the independent kings of Scots the right to be anointed with holy oil, but that year also Robert died. By the appropriate standards of medieval kingship his success had been total; but by the nature of medieval kingship, his successor was left with the same struggle to wage all over again.

David II (1329–71). Robert I's son, David II, has perhaps received unfair treatment from historians through having been contrasted with his illustrious father. Just over five years of age at his accession, he was soon confronted

with a renewal of the Anglo-Scottish war, exacerbated by the ambitions of those Scots who had been deprived of their property by Robert I or otherwise disaffected. In the 1330s Edward Balliol, pursuing the claim to the throne of his father John, overran southern Scotland. In return for English help, he gave away to England southern lands and strongpoints not recaptured fully by the Scots for a century. After the Scottish defeat at Halidon Hill near Berwick in 1333, David was forced to flee to France in the following year. Berwick itself fell to the English and was never again in Scottish hands except in the period 1461–82.

Battle of Halidon Hill

The Scots gradually regained the initiative, and in 1341 David was able to return to Scotland. But in 1346 David II himself was captured at the Battle of Neville's Cross near Durham. He was released in 1357 for a ransom of 100,000 merks. This ransom, if paid (and three-quarters of it eventually was), would constitute a serious burden on Scotland, and there is evidence of Parliament's using this national emergency to establish some checks on the actions of the crown. In addition, the representatives of the royal burghs, which were important as an accessible source of finance, established a continuing right to sit in Parliament with the magnates and churchmen from the 1360s on, thus constituting the third of the "Three Estates."

Complex evidence relating to these transactions has been uniformly interpreted in a way discreditable to David. Another interpretation is possible. That he collected revenues more assiduously than he made ransom payments may indicate a reasoned attempt to strengthen the crown financially; and his negotiations, especially of 1363, whereby a member of the English royal house was to succeed him on the Scottish throne, may have been a diplomatic charade. Whatever his faults, David left Scotland with both its economy and its independence intact.

The long wars with England necessarily took their toll, retarding Scotland's economy and weakening the authority of its government. The buildings that have survived from this era are inferior to earlier work, much of which, of course, suffered damage at this time. War was increasingly expensive, and taxation was increased drastically to pay David II's ransom. But again, a rosier alternative picture can be painted, suggesting that the burgesses were able to meet the increased taxation because of increased prosperity through the still-continuing trade with England.

SCOTLAND IN THE 15TH CENTURY

The early Stewart kings. David was succeeded by Robert II (1371–90), previously the high steward and son of Robert I's daughter Marjory. The next king was Robert II's son John, restyled Robert III (1390–1406). It may be that the future Robert II's conduct was responsible for dissension in Scotland during David II's reign, particularly during his captivity in England. At any rate, neither Robert II nor his son Robert III were strong kings and some nobles regarded both as upstarts, and the latter as of doubtful legitimacy. There thus began a long period of monarchical weakness in Scotland, accentuated by a series of royal minorities in the 15th and 16th centuries. Historians have made much of the turbulence of these times, but there were comparable periods of governmental weakness in contemporary England and France; and "bonds of manrent" and other alliances made by the magnates with each other and with their social inferiors should be seen as much as attempts to secure political stability in their own localities as threats to the overall peace of the kingdom.

Period of weak monarchs

Robert III's younger brother, Robert Stewart, 1st Duke of Albany, more than once was given powers to rule in his brother's name, and Robert's son James may have been sent to France in 1406 in order to keep him out of Albany's clutches. But James was captured at sea by the English, and shortly afterward Robert III died. Following Albany's death in 1420, his son Murdac continued to misgovern the realm until 1424, when James I, then 29, was ransomed.

The Douglas family was becoming particularly powerful at this time. They had been rewarded with the gift of the royal forest of Selkirk and other lands in south and southwest Scotland for loyal service to Robert I. But the growing power of the Douglases in this vital border area posed by the end of the 14th century a growing threat to the crown. At the same time the Lords of the Isles had attained a stature in the western Highlands that overtopped that of the kings of Scots.

One notable event was the founding of Scotland's first university at St. Andrews. The Wars of Independence led Scottish students to go to Paris rather than to Oxford or Cambridge. But universities were the training grounds of the clergy, and when, in the period 1408–18, Scotland recognized the antipope Benedict XIII after he had been abandoned by France, it became expedient for Scotland to have its own university. The bulls of foundation from Benedict XIII reached St. Andrews in 1414.

James I (1406–37) was an active and able king, keen to make the crown wealthy and powerful again. Perhaps he was overeager to make up for time lost in his captivity, and thus he prompted the opposition that led to his death. The new posts of comptroller and treasurer were created to gather royal revenues more efficiently. Murdac, 2nd Duke of Albany, was executed in 1425, and other powerful men were overawed, even in the far north. The laws were to be revised, and in 1426 a court for civil cases was set up, presaging the later Court of Session.

The reign of James I

Possibly to balance the power of the magnates, it was enacted in 1426 that all tenants in chief should attend Parliament in person. More realistically, they were, from 1428, permitted to send representatives from each shire. Even this system did not operate until the late 16th century. If James had been inspired during his captivity by the English House of Commons, he was unable to transplant that institution to Scotland. The Scots Parliament, like that of many other European countries, remained throughout the medieval period the feudal court of the kings of Scots; lacking the distinctive development of the English Parliament, it did not differ essentially in kind from the feudal court of any great magnate. Despite, or perhaps because of, his innovative vigour, James made enemies for himself. His murder in 1437 was part of an attempt to seize the throne for Walter Stewart, Earl of Atholl, but the conspirators were executed and James's young son succeeded him.

James II (1437–60) was six years old at the time of his accession. His minority was marked by struggles between the Crichton and Livingston families. During this minority and that of James III, James Kennedy, bishop of St. Andrews, played a statesmanlike part in seeking to preserve peace. James II took a violent line against overambitious subjects. In 1452 he stabbed William Douglas, 8th Earl of Douglas, to death, and in 1455 James Douglas, 9th Earl of Douglas, was attainted. The main line of the Douglas family never regained its position, though a younger, or cadet, branch of the family, the earls of Angus, was important in the late 15th century. James II, like his father, thus sought boldly to reassert royal authority, and Scotland lost an able king when he was killed by the bursting of a cannon at the siege of Roxburgh Castle, one of the last Scottish strongpoints in English hands. Roxburgh was subsequently captured by the Scots. Among the cultural advances of the reign was the founding, in 1451, by Bishop William Turnbull of the University of Glasgow, Scotland's second university.

James III (1460–88), James's son, acceded at the age of eight. During his minority he was for a time the pawn of the Boyd family. The so-called Treaty of Westminster-Ardtornish of 1462 showed that John, Lord of the Isles, and the exiled Douglas were prepared to try to carve Scotland into two vassal states of England for themselves. The alliance came to nothing, but the Lords of the Isles were a threat to the territorial integrity of Scotland until their final forfeiture in 1493. On the other hand, the power vacuum left by their removal was responsible for much of the unrest in the western Highlands thereafter. It was in James III's reign that the territory of Scotland attained its fullest extent with the acquisition of Orkney and Shetland in 1468–69.

As James III came of age, he seems to have given grave offense to his nobles by shunning their company for that of artists. It has been suggested that his fine sensibility did

Character of James III

him credit, but this is probably an anachronistic view. So serious was James's lack of authority that Berwick fell in 1482, when the nobles, led by Archibald Douglas, 5th Earl of Angus, chose—rather than to defend the county against the English—to seize their opportunity to hang some of James's favourites. In 1488 James was murdered while fleeing from a battle against his opponents at Sauchieburn, though it seems that the death of the king was not intended, and he was succeeded without trouble by his son.

15th-century society. There is evidence of economic recovery in Scotland in this period, despite the continuing war and unrest. Castle building and the extending of monasteries and cathedrals were widespread; work was done on the royal residences at Linlithgow and Stirling. The building of collegiate churches and of fine burgh churches is additional evidence of prosperity. Royal burghs with their share in international trade and baronial burghs with their rights in their own locality were alike flourishing. The craftsmen threatened to rival the merchants in the running of burgh affairs, but an act of 1469 gave the merchants the majority on the town councils; this allowed self-perpetuating cliques to misapply the assets of the burghs, an abuse not remedied until the 19th century. Accompanying the prosperity general in Scotland at this time was a tendency to inflation, and a debasement of the coinage added to the troubles of James III's reign.

Cultural life

From the late 14th century onward, interesting Scottish writing, both in the vernacular and in Latin, has survived. John Barbour (1316?–95) wrote a verse life of Robert I in Scots. A Latin history of Scotland was compiled by John of Fordun and continued by Walter Bower, abbot of Inchcolm, in his *Scotichronicon.* Andrew of Wyntoun wrote a history of Scotland in Scots verse.

Little is left of the corpus of medieval writings in Scottish Gaelic. But the sophistication of the west Highland stone carvings of the later Middle Ages suggests that a strong literary culture, too, was associated with the courts of the Lords of the Isles and other chiefs. The *Book of Deer,* containing the Gospels, has in its margins an 11th-century Gaelic account of Columba's foundation of the monastery of Deer in Aberdeenshire, as well as a series of *notitiae,* or lists of church rights, which provide clues to the nature of Celtic society. The early-16th-century *Book of the Dean of Lismore* (the seat of the bishop of Argyll) contains more than 60 Gaelic poems. From the quality of the architecture that has survived from the 15th century, one can infer the existence of paintings and other objects, such as church furnishings, that have largely disappeared. An outstandingly intricate collegiate church is that at Roslin near Edinburgh, founded by Sir William Sinclair, 3rd Earl of Orkney, about 1450. There are fine burgh churches, such as St. John's in Perth and the Church of the Holy Rood in Stirling. Perhaps the outstanding piece of evidence of royal patronage of the arts is the altarpiece for James III's Trinity College Church in Edinburgh, which is almost certainly the work of the great Flemish painter Hugo van der Goes.

The church

In the 14th century the papacy had built up its claims to appoint to the higher offices in the church; in Scotland it had established a system of "provisions," or papal appointments, to vacant offices. This cut not merely across the rights of rulers who used the church to provide their loyal bureaucrats with a living and the rights of other local patrons; it also meant a drain to Rome of money in the form of the tax payable by a cleric "provided" to a vacant post by the pope. James I resisted these developments, and at the same time, in the Council of Basel (1431–49), the "conciliarists" were seeking to curb papal power in the church; a distinguished member of the Council of Basel was the Scot Thomas Livingston, one of the first St. Andrews graduates.

James also sought to revive the monastic ideal in its early purity and established a house of the strict Carthusians at Perth. A compromise between James I and the pope was probably pending when James was murdered, and his successors tended to let the popes collect their money as long as they "provided" to church offices along lines acceptable to the monarchy. In 1487 James III was granted the concession that the pope would delay promotions to the higher offices for eight months so that the king could propose his nominee.

St. Andrews was made the seat of an archbishopric in 1472, in itself a desirable step. But the first archbishop of St. Andrews secured the honour by supporting the papacy against the king, and there was, as a result, no welcome for the appointment in Scotland. Glasgow also became an archbishopric in 1492.

SCOTLAND IN THE 16TH AND EARLY 17TH CENTURIES

James IV (1488–1513) and James V (1513–42). James IV was well equipped for kingship, being physically impressive, cultured, generous, and active in politics and war alike. He eliminated a potential rival by carrying out the forfeiture of the last Lord of the Isles, in 1493, and dealt severely with unrest on the English border and elsewhere. James and Bishop William Elphinstone of Aberdeen founded King's College, Scotland's third university, in Aberdeen in 1495. This was the great age of Scottish poetry, and while one of the leading "makars," or poets, Robert Henryson (1430?–1506?), author of the *Testament of Cresseid,* was a burgh schoolmaster, the others were members of the court circle: Gavin Douglas (1474–1522), bishop of Dunkeld and kinsman to the earls of Angus, translated Virgil's *Aeneid* splendidly into Scots, and William Dunbar (1460?–1520), a technically brilliant poet, showed the versatility of which Scots was capable.

After initial disharmony with England, James concluded a "treaty of perpetual peace" with Henry VII in 1502 and married Margaret, Henry's daughter, in 1503. But Henry VIII of England became involved in the anti-French schemes of Pope Julius II, and in 1512 France and Scotland renewed their "auld alliance" as a counterbalance. In 1513 Henry VIII invaded France. James IV, consequently, invaded England; there he died, along with thousands of his army, in the rashly fought and calamitous Battle of Flodden.

Battle of Flodden

James's efficiency at home was thus offset by his excessive international ambitions. And both had cost money—for artillery; for a navy whose greatest ship, the *Great Michael,* cost £30,000; for embassies. The crown granted lands in feu-farm tenure, which gave heritable possession in return for a substantial down payment and an unchangeable annual rent thereafter. In the great European price rise of the 16th century, this policy in the long term weakened the crown.

James V (1513–42) was in his second year at his accession. The factional struggles of his minority were given shape by the division between those who adhered to Scotland's pro-French alignment and those who were determined that the price Scotland paid at Flodden should not be repeated. John Stewart, Duke of Albany, was regent until 1524, and favoured France; Archibald Douglas, 6th Earl of Angus, then maintained a pro-English policy until 1528 when James began his personal rule. James now found Scotland's support in international politics being sought on all sides. In the 1530s he obtained papal financial help in establishing a College of Justice, and he concluded two successive French marriages, each bringing a substantial dowry; his second wife, Mary, daughter of the Duke de Guise, became the mother of Mary, Queen of Scots. James's support for the papacy and France alienated some of his subjects, however, and his rule was not simply strict and financially vigorous but rather avaricious and vindictive. Lack of noble support seems to have caused the rout at Solway Moss in November 1542 of a force invading England. This and the deaths of his infant sons led to the death of James, probably from nervous prostration, in December, a week after the birth of his daughter, Mary.

Mary (1542–67) and the Scottish Reformation. The church in 16th-century Scotland may not have had more ignorant or immoral priests than in previous generations, but restiveness at their shortcomings was becoming more widespread. And the power structure of the church seemed to preclude the possibility of reform without revolution. The church made a poor showing at the parish level, since by 1560 the bulk of the revenues of nearly nine parishes in every 10 was appropriated to monasteries and other central institutions. The papacy, in return for receiving its

share of this wealth, abandoned spiritual direction of the Scottish church; from 1487, royal control over appointments to the higher church offices grew steadily. All this, at a time when the church's annual revenue—reckoned at £400,000 in 1560—was 10 times that of the crown, readily explains the attraction of church office for unspiritual career-seeking nobles. Church lands were feued to laymen, who also became collectors of church revenues and were given abbeys as benefices. Church property, particularly monastic property, was effectively being secularized, and if Protestantism offered to the nobles and lairds of Scotland a more spiritually alive church—and one with lay participation—it probably also appealed to them as a system under which they would not have to hand back what they had grabbed.

Particular laymen were as pious as ever, endowing collegiate churches as they had once endowed monasteries, and trenchant criticism of church abuses was expressed in the play *Ane Pleasant Satyre of the Thrie Estaitis* by Sir David Lyndsay (*c.* 1490–*c.* 1555). But reform from within was probably almost impossible. Archbishop John Hamilton, for instance, a would-be reformer who gave his name to a vernacular catechism (1552), belonged to the family who had most to lose if the careerists were curbed.

Mary, Queen of Scots

Mary (1542–67) began her reign as another Stewart child ruler in the hands of factions. The pro-French party upheld the old church, while the pro-English desired reform. By the Treaties of Greenwich (1543), Mary was to marry Edward, Henry VIII's heir. Cardinal David Beaton and Mary of Guise, the queen mother, had this policy rescinded, and the murder of Beaton (1546) and English punitive raids culminating in the Scottish defeat at Pinkie (1547) did not cause Scotland to love England more. France helped Scotland to expel the English, but only in return for such a hold over the country that by the time of young Mary's marriage to the dauphin in 1558 it was France that appeared to be about to absorb Scotland.

Anti-French feeling combined with Protestant preaching to bring about revolt. In 1559 the reformers took up arms to forestall Mary of Guise's action against them. Despite the preaching of John Knox and others and the plundering of the monasteries, the decisive issues were political and military: Queen Elizabeth of England sent troops to check French plans in Scotland. Mary of Guise died in June 1560, and by the Treaty of Edinburgh in July, both France and England undertook to withdraw their troops. With Scotland thus neutralized, England had the important advantage over France of relative nearness.

The Scots Parliament in August 1560 abolished papal authority and adopted a reformed Confession of Faith, but Mary, still in France, did not ratify this legislation. Still, the organization of local congregations, which had been going on for some years, continued, and the General Assembly emerged as the central legislative body for the church. In the *First Book of Discipline* (1560), John Knox and other ministers proposed for the church a striking social program, providing education and poor relief. But laymen had not despoiled the old church to enrich the new, and, as an interim settlement secured by Mary's government in 1562, the church and crown were together to share but one-third of the old church's revenue.

John Knox

Mary's husband died in 1560, and in 1561 she returned to Scotland. As a Roman Catholic in a Protestant land and as nearest heir, by descent from Henry VII's daughter, to Elizabeth of England, she had many enemies. Her personal reign was brief and dramatic—she married her cousin Darnley (1565); their son James was born (1566); Darnley was murdered (1567); Mary married the adventurer James Hepburn, 4th Earl of Bothwell; was imprisoned and forced to abdicate (1567); and subequently escaped and fled to England (1568). Her task as a ruler was hard, and the harder for her own errors of judgment, but she essayed it bravely and was a truly tragic rather than a pathetic figure.

James VI (1567–1625). James lived through the usual disrupted minority to become one of Scotland's most successful kings. In a civil war between his own and his mother's followers, laird (landed proprietor) and merchant support for James may have been decisive in his eventual victory. Elizabeth detained Mary in England and assisted James Douglas, 4th Earl of Morton, regent from 1572, to achieve stability in Scotland.

James's government ratified the reformed church settlement, and more permanent measures of church endowment were taken. The Concordat of Leith (1572) allowed the crown to appoint bishops with the church's approval. As in Mary's reign, the crown was intervening to prevent the wealth of the old church from being entirely laicized. And if the bishopric revenues were saved from going the same way as the monastic wealth, the crown expected a share in them for its services.

A new presbyterian party in the church, whose members wanted parity of all ministers and freedom from state control, rejected this compromise. Led by Andrew Melville, a rigid academic theorist, they demanded, in the *Second Book of Discipline* (1578), that the new church receive all the wealth of the old, that it be run by a hierarchy of courts, not one of bishops, and that the state leave the church alone but be prepared to take advice from it. Many historians have seen these demands, as James undoubtedly did, as an attempt to achieve full-blown theocracy. James was not strong enough for out-and-out resistance immediately, and he sometimes made concessions, as in the Golden Act of 1592, which gave parliamentary sanction to the system of presbyterian courts. But he gradually showed his determination to run the church his own way, through the agency of his bishops, who were brought into Parliament in 1600. From 1606 Melville was detained in London and later banished. By 1610 the civil and ecclesiastical status of the bishops was secure. The continued existence of church courts—kirk sessions, presbyteries, synods, and the General Assembly—show James's readiness for compromise; and he showed a wise cautiousness toward liturgical reform after encountering hostility over his Five Articles of Perth (1618), which imposed kneeling at communion, observance of holy days, confirmation, infant baptism, and other practices.

In the 1580s James, as he became personally responsible for royal policy, faced the need to control unruly subjects at home, nobles and kirkmen alike, and to win friends abroad. He concluded a league with England in 1586, and when Elizabeth executed his mother in the following year as a Roman Catholic threat to the English throne, he acquiesced in what he could not prevent. He thus inherited his mother's claim, and his efforts thereafter to keep in the good graces of Elizabeth and her minister William Cecil were successful. He succeeded peacefully to the English throne in 1603, though his two monarchies, despite his own personal inclinations, remained distinct from one another.

Accession of James VI to the English throne

His policy was one of overall insurance; he avoided giving offense to Catholic continental rulers, and, while he dealt effectively with lawbreakers on the border and elsewhere, he showed marked leniency to his Catholic nobles, even when the discovery of letters and blank documents (the "Spanish Blanks" affair, 1592) showed that several of them were in treasonable conspiracy with a foreign power. Neither a heroic king, like James IV, nor the pedantic and cowardly buffoon depicted in Sir Walter Scott's *The Fortunes of Nigel,* James VI was a supple and able politician. His theories of divine-right monarchy were a scholar-king's response to an age when the practice and theory of regicide were fashionable. Except perhaps at the very end of his life, James was too realistic to let his theories entirely govern his conduct.

James excelled in picking good servants from among the lairds and burgesses; they were his judges and privy councillors and sat on the Committee of Articles, with which he dominated Parliament. After 1603 they governed Scotland smoothly in his absence. From 1587 Parliament was made more representative by the admission of shire commissioners to speak for the lairds, thus realizing the program of James I. The privy council had judicial as well as legislative and administrative functions; there were, in addition, the Court of Session for civil cases (it had evolved from the council in the early 16th century and, as the College of Justice, had been endowed with church funds in the 1530s) and justice courts for criminal cases.

Local justice and administration continued, however, despite James VI's efforts, to be largely the prerogative of the landowners.

Scotland still had a subsistence economy, exporting raw materials and importing finished goods, including luxuries. But such luxury imports showed that the greater landowners and merchants were gaining in prosperity. Despite the absence of adequate endowment, the reformed church began to create a network of parish schools, and there was advance in the universities. Melville brought discipline and the latest scholarship to Glasgow and St. Andrews in turn, and there were new foundations at Edinburgh (the Town's College, 1582) and Aberdeen (Marischal College, 1593).

Scotland and England were drawing closer together, as the period of continual strife between them receded in time. Though the two national churches were not identical in structure, they shared a common desire to protect and preserve the Reformation. James VI's accession to the English throne in 1603 as James I encouraged further cultural and economic assimilation. It was far from guaranteeing further political assimilation, but a century of the barely workable personal union of the crowns was increasingly to sharpen for the Scots the dilemma of choosing between complete union and complete separation.

THE AGE OF REVOLUTION (1625–89)

Charles I (1625–49). James VI's son, Charles I, grew up in England, lacking any understanding of his Scottish subjects and their institutions. He soon fell foul of a nobility restless in a Scotland that lacked the natural focal point of a royal court. The king also caused widespread anger by high taxation, by the special demands made on Edinburgh to build a Parliament House and to provide a cathedral for the bishopric founded there in 1633, and by a Spanish and a French war that were intended to further English diplomacy but also disrupted Scottish trading ties. The aristocratic leaders of the opposition found ideal material on which to build clerical and popular support. Charles and his Scottish bishops were fond enough of ritual and splendour in church services to make plausible the (wholly incorrect) suggestion that they were ready for compromise with Rome. The new Book of Canons (1635–36) and Liturgy (1637) therefore offended by their content, as well as by being authorized by royal prerogative alone. The National Covenant (1638) astutely collected national support for the opposition's pledge to resist Charles's innovations. Condemnation of popery was written into it for the benefit of those who feared that Charles might be a crypto-Catholic; others, more sophisticated, welcomed its implicit condemnation of a royal arbitrariness with religion and private rights that was contrary to all Scottish precedent.

Religious and political opposition

The Covenanters humbled Charles in two almost bloodless campaigns, the Bishops' Wars (1639–40), and left him with no alternative to asking for money from an English Parliament in which his opponents were strongly represented. Charles had authorized a general assembly of the Scottish church (1638) and a Scottish Parliament (1639); the Covenanters packed these meetings, scrapped all the king's innovations, and abolished episcopacy. There was, therefore, by 1641 a revolutionary situation in both kingdoms, and in August 1642 war broke out between Charles and his English opponents. Both sides sought Scottish help, which was soon accorded to the English parliamentary opposition. By the Solemn League and Covenant (1643) the English promised, in return for military aid, to help preserve government by the Presbyterian church in Scotland and, so at least the Scots believed, to set it up in England. James Graham, 1st Marquess of Montrose, and others who then left the Covenanting side argued that by this second Covenant, and by certain constitutional constraints they had placed upon the crown, the Scots had gone unwarrantably far beyond the aims of the first Covenant. But those of the Scots who were prepared to make common cause with the English opposition, even if the English did have a more deep-seated quarrel with their king than the Scots, had reasoned justification; for it was realistic to expect that Charles, as soon as it proved possible, would withdraw concessions made to men whom he regarded as his enemies. Personal antipathies also helped to split the ranks of the original Covenanters—notably the antipathy between Montrose and Archibald Campbell, 1st Marquess of Argyll, sincerely devoted to the cause but equally devoted to the advancement of his family. Montrose's military efforts for Charles in Scotland were crushed in 1645, and by 1646 Charles had lost the war in England, too. When Charles surrendered to the Scottish army in England, the Scots failed to reach agreement with him and handed him over to the English. The Scottish contribution to the English war effort had been substantial, but not spectacular enough to leave a sense of obligation; and the English army under Oliver Cromwell, now eclipsing Parliament in English politics, preferred independency to presbyterianism in the church and did not propose to honour the Solemn League and Covenant. A conservative element among the Covenanters in 1647 reached a compromise, or "Engagement," with Charles by which they promised him help in return for the establishment of Presbyterianism in both kingdoms for three years and went to war on his behalf; their ill-planned campaign was crushed at Preston in 1648. The clerics, who had bitterly opposed this compromise, were now able, under the leadership of a few nobles such as Argyll, to purge the Scottish Parliament and army of all those tainted with collaboration with the king. The execution of Charles by the English in 1649 genuinely shocked most Scots, who were prepared to fight for his son, Charles II, once he had been constrained to accept the Covenants and once Montrose had been executed (1650). Cromwell's victory over the Scots at Dunbar (1650) gave more moderate Scots the ascendancy again, but this brought no better military result. Another, and decisive, defeat at Cromwell's hands came to a Scottish royalist army at Worcester in 1651.

Execution of Charles

Cromwell. Cromwell imposed on Scotland a full and incorporating parliamentary union with England (1652). This could not enjoy the popularity of a union by consent, maintained as it was by an army of occupation, but Cromwell's administration of Scotland was efficient, and his judges, some of them Englishmen, achieved an admired impartiality. Public order was well maintained, even in the Highlands after the collapse of royalist resistance in 1654. Cromwell did not overturn Presbyterianism but ensured toleration for others, save Roman Catholics and Episcopalians.

The Restoration monarchy. The restoration in 1660 of Charles II (1660–85) was welcomed by many moderate men of both his kingdoms. Charles had learned much from his father's fate and was prepared to forget many injuries, though his government executed some Scots, including the Marquess of Argyll.

In 1662 Charles formally restored church government by bishops, but they were to act in association with synods and presbyteries, much as under James VI's compromise. Charles seems not to have been moved by rancour toward the Covenanters, who had bullied him in the early 1650s, but merely by a desire to achieve the system that satisfied most people. Many laymen accepted his system, and few nobles opposed it. Approximately 270 ministers, however—just over a quarter of the total—were deprived of their parishes for noncompliance. The Pentland Rising (1666) was easily put down and was countered by an experimental period of tolerance by the government. Persons who still persisted in attending conventicles were strong only in the southwest and to some extent in Fife and among the small lairds and common people. These men adhered to the "Protester" position, regarding Scotland as still bound by the Covenants. In another trial of strength with the government, they were defeated at Bothwell Bridge (1679). The remnant of Cameronians (from Richard Cameron, a leading Covenanter) remained in being, meeting governmental violence with further violence, and in 1690 refused to join a Presbyterian but uncovenanted Church of Scotland. Their brave and fanatical "thrawnness" (recalcitrance) endeared them to later generations of Scots.

When Charles's brother succeeded as James VII of Scots and James II of Great Britain and Ireland (1685–88), most Scots showed that they were prepared to support him despite his Roman Catholicism. But he showed his

James VII

ineptitude by requesting Parliament to grant toleration to Catholics (1686); this stirred up unprecedented opposition to royal wishes in the Scottish Parliament. Nevertheless, although many exiled Scots were at the court of William of Orange in Holland, the collapse (1688–89) of James's regime in Scotland was entirely a result of the Revolution of 1688 in England and the landing there of William.

THE ERA OF UNION

The Revolution settlement. James VII having fled to France, a Convention of Estates (really the same assembly as Parliament but meeting less formally) gave the crown jointly to the Protestant William of Orange (William III of Great Britain, 1689–1702) and his wife Mary (II of Great Britain; 1689–94), James's daughter. William's first major decision was a moderate one: episcopacy was abolished in 1689 and presbyterianism reestablished the following year. A series of crises throughout William's reign, however, exposed his total lack of interest in Scotland and placed a strain on the system that had developed whereby the Scottish ministry took orders not only from the monarch but also from the English ministry.

The Act of Union and its results. William fought one war against France (1689–97) and on his death in 1702 bequeathed another (1701–13) to his successor, his wife's sister Anne (1702–14). These circumstances made a union of Scotland and England seem strategically as well as economically desirable. That union was achieved in 1707 is at first sight surprising, since intervening sessions of the Scots Parliament had been in a mood to break the English connection altogether. But by 1707 England's appreciation of its own strategic interests, and of the nuisance value of the Scots Parliament, was lively enough for it to offer statesmanlike concessions to Scotland and material inducements to Scots parliamentarians to accept union.

The union was an incorporating one—the Scots Parliament was ended and the Westminster Parliament increased by 45 commoners and 16 peers representing Scotland. Scotland benefited by gaining free trade with England and its colonies, by the grant of a money "Equivalent" of the share of the English national debt that Scotland would assume, and by the explicit safeguarding of its national church and legal system. After Queen Anne's death in 1714, when the Jacobites missed their best opportunity, the worst crises of the union were past.

Jacobitism: the Highlands. The Jacobites were seldom more than a nuisance in Britain. An expedition from France in 1708 and a West Highland rising with aid from Spain in 1719 were abortive; bad leadership in the rebellion in 1715 (known as "the Fifteen") of James VII's son, James Edward, the Old Pretender, and divided counsels in that of 1745 ("the Forty-five") led by the Old Pretender's son Charles Edward, the Young Pretender, crippled invasions originating in France which had in any case less than an even chance of success. The government was not always sufficiently prepared against invasions, but the generalship of John Campbell, 2nd Duke of Argyll, at Sheriffmuir in 1715 sufficed to check, and that of William Augustus, Duke of Cumberland, at Culloden in 1746 to deal the coup de grace to, a Jacobite army. The Jacobites never had full French naval and military assistance, and support in Scotland itself was limited; not many more Lowland Scots than Englishmen loved the Stuarts enough to die for them. Many politicians, especially before 1714, corresponded with the royal exiles simply as a matter of insurance against their return, and in the dying days of Stuart hopes there were fewer people than there have been since who were struck by the romantic aura surrounding Prince Charles Edward, the "bonny Prince Charlie." In the main the Stuarts had to rely on the clans of the Gaelic-speaking regions, and Highland support in itself alienated Lowlanders. Not all Highlanders were "out" in the Fifteen or the Forty-five; such clans as the Campbells and Munros, Macleods, and Macdonalds of Sleat were Hanoverian because Presbyterian, or through their chiefs' personal inclinations. Many clans were, however, Roman Catholic or Episcopalian and favoured a Catholic monarch; they were legitimists and reasonably so, since both James VII and his son James Edward, the Old Pretender, appreciated Highland problems. These were the problems of an infertile land, overpopulated with fighting men who owed personal allegiance to their chiefs and were partly dependent on plunder to maintain their standard of living. It is hard to see what in the end could have happened to this society, other than what did happen: a series of attempts by the chiefs in the late 18th, and particularly in the early 19th, century to emulate the new capitalist agriculture of the Lowlands, thus creating an impersonal cash relationship with their tenants and leaving those who were redundant in the new economy no alternative to moving south or overseas. But the catastrophe of the Fifteen and Forty-five made this process more rapid and more painful. This is the central fact of the situation, even though the atrocities of government soldiers and the repressiveness of government legislation did very much less than economic and social forces to usher in the new order.

The Fifteen and the Forty-five

The Scottish Enlightenment. No straightforward connection can be drawn between the union and the exceptional 18th-century flowering of intellectual life known as the "Scottish Enlightenment." Absence of civil strife, however, permitted the best minds to turn, if they chose, from politics and its 17th-century twin, religion; and few of the best minds from 1707 onward were in fact directly concerned with politics. Philosophy, in which 18th-century Scotland excelled, was a proper concern for a country where for generations minds had been sharpened by theological debate. Scottish culture remained distinctive, and distinctively European in orientation. The historian and philosopher David Hume sought to remove Scotticisms from his speech, and the architect Robert Adam gained extra experience as well as income from being able to design buildings in London as well as in Edinburgh. Nevertheless, Adam drew most of his stylistic inspiration from the classical architecture he had studied in Italy, and Hume, "le bon David," was an honoured member of continental polite and intellectual society. Hume's *The History of England* (1754–62) made his literary reputation in his lifetime; but it is his philosophical works, such as his *A Treatise of Human Nature* (1739–40), which have caused the continuous growth of his reputation since his death. Adam Smith, author of *The Wealth of Nations* (1776), was the philosopher of political economy. Henry Home, Lord Kames, may be singled out from a number of other significant figures to illustrate the versatility characteristic of the times. He was a judge, interested in legal theory and history; an agricultural reformer in theory and practice; a Commissioner of the Forfeited Estates (of the rebels of 1745); and a member of the Board of Trustees for Manufactures (which encouraged Scottish industries, notably linen). In poetry there was a reaction, possibly against union, and certainly against assimilation, with England; revived interest in Scots vernacular poetry of the past was the herald of a spate of new vernacular poetry, which culminated in the satires of Robert Fergusson and the lyrics of Robert Burns. Some of the greatest Gaelic poets, such as Alexander MacDonald, were also writing at this time.

David Hume

The Scots educational system, its foundations so securely laid throughout the previous century, made possible, though neither it nor any other single factor could be held to explain, this extraordinary cultural outpouring. The Scottish universities enjoyed their heyday, with Edinburgh notable for medicine and preeminent in most other subjects. Gradually the regents who taught students throughout their university course were replaced by professors specializing in single subjects. That students seldom troubled to graduate was little disadvantage in an age when appointments depended on patronage; and, not being bound by a rigid curriculum, they were able to indulge the Scot's traditionally wide intellectual curiosity by attending lectures in a variety of subjects. Scientific study was encouraged, and practical application of discoveries given due place. Francis Home, professor of Materia Medica at Edinburgh, studied bleaching processes and plant nutrition; and James Watt, instrument maker to the University of Glasgow for a time, was there encouraged to work on the steam engine, to which he was to make crucial improvements.

Scottish universities

19TH-CENTURY SCOTLAND

Agitation for constitutional change was considered treasonable by many during the years (1793–1815) when Britain was fighting revolutionary France. Several advocates of universal suffrage, including a young Glasgow lawyer, Thomas Muir of Huntershill, were sentenced to transportation in 1793. After repression had broken this first radical wave, postwar industrial depression produced another—the "Radical War" of 1820, an abortive rising of workers in the Glasgow area. Intellectual campaigning of a more moderate sort had greater short-term success. The *Edinburgh Review,* founded in 1802 by a group of young lawyers led by Francis Jeffrey and Henry Brougham, was influential in radical politics and in literature. Edinburgh life was particularly brilliant during the war years, when students unable to study abroad found the University of Edinburgh more attractive than ever. Outstanding in this period was Sir Walter Scott, although not until 1827 was he known to be the author of the Waverley novels. Scott's greatness as a novelist lay in the way he took Scottish society as a whole for his main character; and his best books are a lament for an era that he knew was dying, the organic society of preindustrial Scotland.

Novels of Sir Walter Scott

The Industrial Revolution. The Scottish Industrial Revolution was in full swing from the 1820s. Linked with this, in a way historians have not altogether disentangled, was a dramatic upsurge of population. There were perhaps one million people in Scotland in 1700. By 1800 there were more than 1.5 million and by 1900 nearly 4.5 million. The manufacturing towns showed spectacular increases. Hundreds of thousands of Irish emigrants went to Scotland in the 19th century, notably during the Irish potato famine of 1846–50. In some country regions there was a population decrease as people moved to the towns, to England, or overseas. Part of the overall increase was the result of improved medical care that had lessened the ravages of epidemic diseases by the mid-19th century. Much of the food for the increased population was supplied by progressive Scottish agriculture. Farming in the southeast was celebrated for its efficiency in the early 19th century, and the northeast became famous for its beef cattle and Ayrshire for its milking herds.

But the key advance was in heavy industry, which from about 1830 took the industrial primacy from textiles, at a time when industry as a whole had replaced agriculture as the nation's chief concern. Coal production rose as did that of iron, with James Beaumont Neilson's hot-blast process (1828) making Scottish ores cheaper to work. Major canals, such as the Forth and Clyde, completed in 1790, enjoyed a short boom before being rendered obsolescent by the railways, of which the Glasgow-to-Garnkirk (1831) was noteworthy for using steam locomotives (rather than horses) from the start. Above all, Scottish international trade was catered to, and Clydeside's reputation made, by the building of ships. Robert Napier was the greatest of many great Scots marine engineers.

Politics. An installment of parliamentary (1832) and burgh (1833) reform ended fictitious county votes and corrupt burgh caucuses but disillusioned the working classes by failing to give them the vote. As in England, they had to await the 1867 and subsequent Reform Acts. But the great bulk of the Scottish middle classes were delighted with the Whigs, who had brought the reforms. The Whig Party, or Liberal Party (as it became known in the 1860s), dominated Scottish mid-19th-century politics; and William Ewart Gladstone, of Scottish parentage, was the great Liberal hero, whose moral dynamism and fire far outweighed in Scottish eyes his High Church Episcopalianism.

Trade unions of skilled workers had had an uninterrupted existence since the early 19th century. By the 1880s unskilled workers were being organized. Various factors delayed the permanent organization of the miners until there emerged from their ranks a major leader, James Keir Hardie. Failing to engage the Liberals sufficiently in support of organized labour, he helped form the Scottish Labour Party in 1888. In 1893 he created the Independent Labour Party for Britain as a whole, and this body in 1900 federated with the trade unions for the purpose of running the Labour Party (given its present name in 1906). Scottish political opinion was with the left in the years before 1914, with Liberal fortunes reviving—partly due to the leadership (1899–1908) of Sir Henry Campbell-Bannerman, a Scot—and three Labour MPs elected.

The Highlands. By 1800 the Highlands has become overpopulated relative to the means of subsistence. Many lairds, seeking to support their tenantry through the kelp industry, were ruined when it collapsed in the decade 1815–25. Other landowners introduced sheep, sometimes violently removing their tenants in the "Highland Clearances," as agents of the Sutherland family did in Strathnaver, Sutherland, about 1810–20. The potato famine of the mid-1840s caused distress. By the 1880s Highland tenants, or "crofters," faced a new problem. Deer forests had replaced sheep runs as the most immediately profitable land use open to landowners; and high rents were asked for the land that was still worked as crofts, though common grazings might at the same time be taken away. Parliamentary agitation by the crofters, who voted for the first time in 1885, and by their Lowland sympathizers, as well as sporadic outbursts of violence beginning in 1882 (the "Crofters' War"), secured an act of 1886 that gave the crofters security of tenure and empowered a Crofters' Commission to fix fair rents. Unlike their Irish counterparts, the Highlanders sought not ownership of their land but the imposition of certain standards of conduct and responsibility upon their landlords. The crofting agitation of the 1880s was a key stage in the forging of a modern Scottish consciousness in that Highlanders and Lowlanders had been united in the struggle.

The "Crofters' War"

MODERN SCOTLAND

World War I and after. The war of 1914–18 had a great impact on Scottish society, with 74,000 lives lost and industry mobilized as never before in a coordinated national effort. Clyde shipbuilding and engineering were crucial, and Clydeside was the key munitions centre in Britain. This expansion of heavy industry, however, seemed in the 1920s to have been an overexpansion. The collapse of the wartime boom in 1920 began a period of economic depression in Britain, in which Scotland was one of the worst-affected regions.

Economic distress bred political radicalism. The Liberals were eclipsed, and in most seats the real contest was between the Unionists and Labour, which became Scotland's biggest single party for the first time in the election of 1922. Willie Gallacher, Scotland's only notable communist member of Parliament and an able political theorist strongly influenced by Lenin, was at the same time a radical belonging to a revered Scots tradition. The death (1930) of John Wheatley, who had been minister of health in the first Labour government (1924) and the author of an important housing act, deprived left-wingers in the Labour Party of a skilled leader, and counsels of moderation in the party prevented its taking any distinctive initiative on the economic crisis. Ramsay MacDonald, a Scot who had led two minority Labour governments, agreed to form a national government in 1931. The Labour Party refused to participate, disowned MacDonald, and was heavily defeated at the polls in Scotland as elsewhere.

Another political development that resulted partly from economic distress was the formation in 1934 of the Scottish National Party (SNP), a merger of two previous parties. It had some distinguished supporters, especially literary men, but it was suspected, sometimes unfairly, of political extremism and made little electoral impact before World War II. The national government of the 1930s was dominated by the conservatives. While opposed to an independent Scottish legislature, this government furthered the extension of the Scottish administrative system in 1939, installing it in St. Andrew's House in Edinburgh.

World War II and after. During World War II Scotland sustained some 34,000 deaths in action and 6,000 civilians killed, many in air attacks on Clydeside. The outstanding Scot on the home front was Tom Johnston, a Labour member of Parliament who acted as secretary of state for Scotland in the wartime national government. He was active in setting up the North of Scotland Hydro-Electric Board in 1943.

The postwar Labour governments contained no Scots of the calibre of Wheatley or Johnston. The SNP enjoyed a short-lived electoral success in the 1970s, when the flow of North Sea oil excited dreams of Scottish sovereignty. To stave off militant nationalism, the Labour government called a referendum to approve its devolution legislation, designed to grant Scotland its own assembly with limited legislative and executive powers. In the polling on March 1, 1979, most Scots who voted favoured the scheme, but because so few went to the polls, the total vote in favour did not amount to the required 40 percent of eligible voters. The Labour government fell, and in the general election the SNP lost nine of its 11 seats in Parliament.

Devolution

In the early 1980s the world recession, coinciding with a collapse in oil prices and a series of closures of large industrial plants, contributed to rising unemployment in Scotland as throughout the United Kingdom. In response, Scotland turned to new directions to revive its economy. Special government agencies were created to attract new investment, most notably U.S. electronics companies, with the result that Scotland has since become one of Europe's major electronics manufacturing centres. The decline in shipbuilding and the traditional heavy industries, however, has meant that Scotland has had to confront the associated social problems of worker retraining. Scotland's resource industries—farming, fishing, and forestry—continue to play an important role in its economy.

For later developments in the history of Scotland, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (J.M.S./Ed.)

WALES

Wales—or, to give the principality its Welsh name, Cymru—is a squat peninsula, with an area of 8,019 square miles (20,768 square kilometres), jutting westward from England into the Irish Sea. It is thus an integral part of the island of Britain, with whose development—and hence with that of the European continent beyond—its character and destiny have been intimately linked. In addition, the location of Wales on that extreme western fringe often characterized as Atlantic Europe has given the country a distinctive endowment in environment and landscape, and it has been responsible for a unique cultural, social, and economic development. The effects of this dual, and often uneasy, heritage continue to permeate much of its national life. Wales exhibits many of the problems encountered by the smaller nations of the world while lacking—as it has for the past seven centuries—their politically independent status.

On three sides the boundaries of Wales are natural, with the shores of the Dee estuary and Liverpool Bay to the north, the Irish Sea to the west, and the coastline of the Severn estuary and its development into the Bristol Channel on the south. The extent of the country from north to south is about 130 miles (210 kilometres); and its width varies, reaching 90 miles across in the north, narrowing to about 40 miles in the centre, and widening again to more than 100 miles across the southern portion. The eastern boundary, that with England, is an administrative but not political frontier created in 1536 as part of the Act of Union finally and effectively linking England and Wales. This boundary, being purely administrative in origin, does not in detail conform to any specific topographical characteristics or features. It runs in general from north to south through the region where the upland massif of Wales gives way to the lowlands of the western Midlands of England. Though not perhaps a natural frontier, it is one that separates distinctive regions, and over the centuries the borderland zone through which it runs has developed a marked character of its own.

Wales in 1974 was divided into eight administrative counties, which are subdivided into 37 districts. One of the districts, Ynys Môn-Isle of Anglesey, though physically an island off the northwest coast, is linked with the mainland by road and rail bridges. Cardiff was designated the country's capital city in 1955.

Physical and human geography

THE LAND

Relief. There are four major features in the relief of Wales, namely mountains (above 2,000 feet), dissected plateaus and hills (600 feet to 2,000 feet), coastal areas, and valley lowlands. The two mountain areas, Snowdonia in the northwest and the Brecon Beacons in the south, have been dominated by the effects of mountain glaciation. In Snowdonia, which contains the highest peak, 3,560 feet (1,085 metres), in the country, these effects have been accentuated by resistant rock, often volcanic in origin, and the magnificent scenery is much more stark and rugged than the softer outlines of the Beacons.

The two mountain areas are both joined and surrounded by a region of plateaus and hills called the Cambrian Mountains, which consist of a number of distinct upland massifs that are themselves partially fragmented by river action. These rolling, river-cut upland surfaces, with an accordance of summit levels and broken by valley lands that tongue in toward the centre, form a backbone dominating the entire relief pattern of Wales.

The upland core

The upland plateaus are themselves girdled on the seaward side by a series of coastal plateaus ranging from 100 feet to 700 feet in height and also well cut into by the erosive action of rivers. They are sometimes broken by the "steps" of fossil cliff lines, and where they do not end in spectacular sea-pounded cliffs they are replaced by coastal flats that are estuarine in origin.

The fourth relief element in the landscape, that of valley land, is made up of the larger river valleys originating in the central upland mass and broadening westward near the sea or eastward as they merge into the lowland plains of the English border.

Drainage and soils. Wales has an essentially radial drainage pattern, the main watershed running approximately from north to south along the backbone of the country. Fragmentation of drainage is therefore the characteristic feature, and, apart from the Severn and the Dee, river systems either reach the sea or debouch upon the English plain before they have an opportunity of becoming fully developed. Natural lake development in Wales is almost entirely glacial in origin and in no way controls the drainage pattern.

The country is dominated by older rocks, ranging from those of Precambrian time, the earliest geological era, to representatives of the Jurassic, some 208 million to 144 million years ago. Glaciation has blanketed most of the landscape with till (boulder clay), scraped up and carried along by the underside of the great former ice sheets, so that rarely can soils be directly related to parent rock. Brown earth and podzol soils predominate, all of them acid in nature and not fully developed.

Climate. Wales enjoys a maritime climate dominated by Atlantic air masses with a considerable variety of character and a relative unpredictability of occurrence. It is therefore weather, rather than climate, that influences the life of the country. Rainfall is frequent and often more than adequate, the annual average being 55 inches (1,385 millimetres), with four inches (88 millimetres) in April and six inches (142 millimetres) in January. Winter snowfall can be significant in the uplands, where some 10 days are recorded on average with snow or sleet falling. The annual mean temperature is 50° F (10° C) but ranging from 40° F (4° C) in January to 61° F (16° C) in July and August. These average figures fail to reflect the extreme variability of actual weather conditions and mask the considerable variations that result from the diverse physical environment.

Average temperatures

Plant and animal life. The combination of physical conditions and centuries of human activity has led to a vegetation coverage in which induced grassland and plantation woodland currently predominate. Grassland varies

from mountain grasses and heather to lowland agrostis–ryegrass pastures, while woodlands are either mixed parkland, boundary and private plantation, or Forestry Commission in origin.

The remoter parts of Wales shelter some mammals and birds extinct or rare elsewhere in Britain. The polecat is fairly common in central Wales though hardly known elsewhere; the pine marten occurs in a few places; a few pairs of red kites represent the sole British survivors, and the chough breeds inland as well as at some coastal sites. Various seabirds and shorebirds occur in large numbers as might be expected in a country with about 600 miles (965 kilometres) of varied coastline. There are three designated national parks in Wales (Snowdonia, Pembrokeshire Coast, Brecon Beacons) and five Areas of Outstanding Natural Beauty (Gower, Lleyn, Anglesey, Clwydian Range, and the Wye Valley).

Settlement patterns. *Traditional regions.* The French name for Wales is the "Pays de Galles" (the "Region of Wales"), and in many ways Wales as a whole can be regarded as a natural or geographical region, in the fullest sense of the word *pays.* More detailed examination, however, reveals the existence of six traditional or popular regions.

The heartland

The first is that of the upland and mountain "heartland," a region with peaks above 2,000 feet (610 metres) and annual rainfall of nearly 60 inches, coinciding partly with the counties of Powys, Clwyd, and Gwynedd. This region extends from the Brecon Beacons in the south to Snowdonia in the north and includes the two national parks based on these mountain areas. Sheep farming and stock raising are predominant. Despite depopulation the region has retained most of the traditional cultural characteristics of Wales, particularly that of language. The hills, uplands, moorlands, and rivers all provide attractions for tourists. Several reservoirs in the region supply water to South Wales and to Merseyside and the Midlands in England.

This core is girdled on the north and northwest by the coastal lowland of North Wales and the Isle of Anglesey, together with the Lleyn Peninsula, all forming part of the county of Gwynedd. Scenic attractions and accessibility from English population centres make the whole region a popular tourist area, but, in addition, the peninsula and the Isle of Anglesey have rich farming areas and light manufacturing industries. Holyhead, on Holy Island off the coast of Anglesey, has a deep-water port and is a busy transportation hub. There is a nuclear power station at Wylfa and an offshore oil terminal at Amlwch. Nearly two-thirds of the people in the region are Welsh-speaking.

To the west of the heartland, and coinciding with the county of Dyfed, lie the coastlands of Cardigan Bay, with pebble and sandy beaches and with cliffs and coves. This region is sparsely populated, with fishing, dairying, tourism, and traditional crafts being the mainstays of the economy. Aberystwyth, with its university college and the National Library of Wales, is an important town. This western-facing region, by its more isolated location, has also been able to preserve many essentially Welsh elements in its social life.

To the southwest, in the county of Dyfed, the heartland gives way to low-lying areas. This is a rich dairy farming and fishing region, with its own coastal tourist attractions, as well as oil terminals and refineries along the vast natural harbour of Milford Haven. The land rises eastward from St. David's Head, through moorlands and uplands, to 1,760 feet at Prescelly Hills.

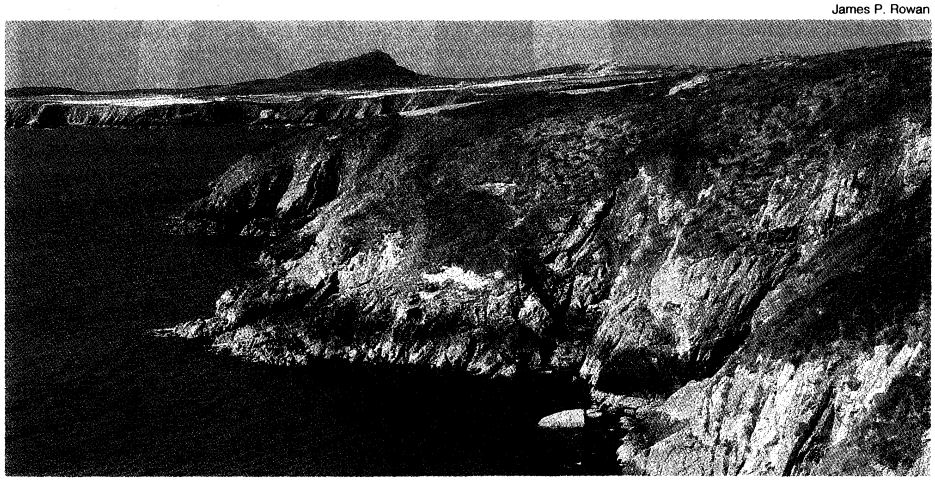
South of the heartland lies the South Wales coalfield, with the Vale of Glamorgan and the Gower Peninsula providing a narrow southern agricultural and coastal fringe. The coalfield was developed in the 19th century as one of the premier mining regions of Britain, and such urban settlements as the Rhondda, with bleak, tightly packed rows of terraced housing strung out along the narrow valleys, are perhaps among the most widely known characteristics of Wales. The region, badly hit by a series of economic depressions in the period between World Wars I and II, and more recently with depopulation, is now less of a coalfield proper and more of a diversified industrial complex. It is the most densely populated region in Wales and coincides with the counties of Gwent and West, Mid, and South Glamorgan. Cardiff is the major commercial and financial centre for Wales. Other important cities and towns include Swansea, Newport, Port Talbot, Neath, Barry, Penarth, and Caerphilly.

The borderland

Finally, to the east of the heartland lies the Welsh borderland, known historically as a marcher, or patrolled frontier region. This is characterized by an amalgam of two cultures, Welsh and English. The north is dominated by the North Wales coalfield and its associated industries, often linked with Liverpool and the Merseyside region; the south is oriented toward the Severn estuary, while the agricultural middle borderland has attempted, by diversifying its economy, to halt the depopulation from which it has long suffered.

Rural settlement. Contemporary Welsh rural settlement shows a dominant pattern of scattering, or dispersion, with nucleation, or a clustering of buildings, playing a secondary role. Isolated whitewashed stone cottages and farm buildings, dotting the cultivated areas, are a striking feature of the rural landscape and remain a strong element underlying the Welsh social fabric. The initial element involved in achieving this pattern was the Welsh tribal economy, of seminomadic pastoral origin, which, by medieval times, had become stabilized enough to produce a settlement pattern dispersed around blocks of suitable land, with very limited nucleation on some of the larger tribal domains. The individual monastic or cell habitations of the Celtic Saints, who carried a tradition of civilization and learning northward through the Atlantic fringe of Europe during the centuries following the collapse of the Roman Empire, increased this dispersion, though some of them, because of favourable site or position, became subsequent growth points. The Anglo-Norman manorial system was introduced into Wales after the conquest of

James P. Rowan



Coastline, Pembrokeshire Coast National Park, Wales.

1282, but, because of physical and political conditions, only in the eastern and southern peripheries of the country did the traditional nucleated village become significant. Post-medieval developments have increased scatter as a result of the farm settlement of former common grazings at the higher altitudes and also added to nucleation in the lowlands at route junctions and crossing places.

Urban settlement. Prior to the Norman Conquest, there were scarcely any traces of urban development in Wales, but the conquest introduced the bastide, or castle town, that still dominates the contemporary urban landscape, in number if not in size. Today these are the nonindustrial market and county towns of the country, commercial, administrative, and social in function. Their physical appearance, however, often betrays their military and colonial origins. Superimposed upon this urban pattern was that generated by the Industrial Revolution. In Wales this simply initiated the growth of unplanned, overcrowded urban settlements resulting from mining activities in the two coalfields. In South Wales it was also linked with development of a steel and tinplate industry and with ports designed for the export of coal. More recent developments have been those of packet stations, for traffic to Ireland, and of resort towns in some of the coastal areas. Some four-fifths of the population live in urban areas. There are two designated New Towns—Cwmbran in the southeast and Newtown, although its name is medieval, situated in the middle borderland. The New Towns were developed as part of a policy to stem depopulation.

The bastide

THE PEOPLE

Although there was some Paleolithic (Old Stone Age) occupation in coastal caves, dating back some 200,000 years, and although traces of Mesolithic inhabitants (8,000–3,000 BC) remain, it was not until the Neolithic and Bronze Age periods, from the 3rd millennium BC onward, that Wales received its basic, and still important, ethnic stock, namely, that of the dark, short Mediterranean type. This western- and coastal-based entry of peoples, part of movements along the whole of Atlantic Europe, was paralleled on the east by migrations of people who had stronger links with lowland Britain and thence with continental Europe. Throughout the whole history of its ethnic and linguistic development, Wales has shared in, and sometimes been a battleground for, these two cultural elements, the one Atlantic and the other continental in origin.

It was this latter element, in the form of an immediate pre-Roman invasion of a western-based Iron Age people, that first brought a Celtic language to Wales and that introduced a Nordic element into the population. The basic culture of these peoples survived the Roman occupation, and it continued to develop, strengthened and broadened by immigrations of similar culture from other parts of Britain in the immediate post-Roman era. It was the language of these people, a Brythonic branch of Celtic speech, that formed the basis of modern Welsh. Their heroic poetry, dating from the 6th century AD, forms the basis of one of the oldest literary traditions of Europe. Subsequent developments included a considerable absorption of Latin elements.

Apart from limited seaborne Dark Age Scandinavian influences upon some coastal fringes, eastern trends, namely, those of Anglo-Saxon and Anglo-Norman penetrations from the English border, have subsequently dominated the ethnic and linguistic story of the country. The effect of the Industrial Revolution in attracting immigrants from outside Wales to the developing coalfields was offset by its effect in attracting to those coalfields people from other parts of the principality. The end result was therefore not so much an influx but rather an internal redistribution of population.

As a result of this history of an indigenous native culture, based on the Welsh language, and the constant pressure from the east, there are two linguistic elements in Wales. Welsh, spoken by about one-fifth of the population, is dominant only in the north and west, the so-called Y Fro Gymraeg, or Welsh-speaking region, where more than 80 percent of the inhabitants speak the language. To the south and east is Anglo Wales, where the proportion of Welsh speakers falls to below 10 percent in the extreme southeast. Measures to sustain the language, which was spoken by half the population in 1900, have become one of the dominant cultural and political themes in Welsh life. In 1967 the Welsh Language Act gave equal validity to the language, and in 1988 a Welsh Language Board was established by the secretary of state.

The Welsh language today

In religion, Wales is predominantly Protestant and nonconformist, Calvinistic Methodism being the most widespread denomination, especially in the Welsh-speaking areas. However, secularization has greatly diminished its influence. The Church in Wales, with its own archbishop, represents the episcopal or Anglican church from which it was disestablished in 1920. The denomination is widely and evenly distributed, but Roman Catholicism forms a small but rapidly growing minority, mainly in the northeast.

The people living in Wales are unevenly distributed, the pattern of settlement being largely determined by five elements. The first of these is the mountain and upland core of the country, which is either very sparsely populated or uninhabited. Surrounding this core, and occasionally entering into it along the river valleys, is a girdle where the population is sparsely distributed. This girdle is, in turn, itself fringed by a peripheral zone, reaching on three sides to the coasts and on the fourth to the English border, where the population density reaches normal or dense levels for Britain. This concentric pattern, which is, of course, a reflection of the agricultural economy, has superimposed upon it two further elements, those of nonindustrial and industrial urban settlements. The former of these is represented by commercial, marketing, and residential towns, which are, in the main, confined to the peripheral zone. The latter consists of industrial settlements deriving from the coalfields of northeastern and of South Wales and their associated industrial developments. In these two areas only can the population distribution be regarded as dense urban. Approximately two-thirds of the total population of Wales is to be found in the South Wales industrial zone.

Density of industrial settlement

Owing to the decline of heavy industry and an increasingly capital-intensive agriculture, Wales during most of the 20th century was characterized by strong outward migration of population, mainly of the young, so that decline was exacerbated by natural decrease related to an aging population. But in the last quarter of the century this loss was checked by new industrial growth, although it still remained characteristic of the coalfields, especially the mining valleys of the south. The result is a rough balance between inward and outward migration. In-migration, however, has consisted primarily of rural retreaters from metropolitan England, a pattern that has set up considerable tensions in the Welsh-speaking areas.

THE ECONOMY

Because of the complete political and economic integration of Wales with the remainder of the United Kingdom, the national economy simply reflects the trends and patterns of the larger unit. Certain specific characteristics of the economy can, however, be recognized, of which the first is that the proportion of capital-intensive industries is relatively high. This means that, since World War II, the rise in the per capita product of the economy has exceeded the United Kingdom average; but, because so much of this is due to capital contribution, the trend is not reflected in the level of personal income, which remains below the average for Britain. The other two individual features are that the economy does not generate sufficient capital for its needs, necessitating a net inflow from other areas, while the most unsatisfactory characteristic of all is that unemployment figures for Wales remain consistently above the national British average.

Resources. Coal remains the one major mineral resource of Wales. Nonferrous ores occur in small quantities, which are not now economically viable; while iron-ore deposits, though important during the early development of the industrial regions, are now exhausted. The two coalfields are in the northeast and in South Wales; the field in the latter is by far the largest and includes some higher-

grade anthracite. Following nationalization in 1947, modernization and the closure of uneconomic pits, largely as a result of geologic problems, made coal mining a declining industry. One of the outstanding features of Wales in the second half of the 20th century was the relegation of coal mining to a minor role in the economy.

Apart from coal and, of course, agriculture, the only other natural resources of Wales are water and woodlands, with the Forestry Commission (a government department) owning and operating on a commercial basis estates of a few hundred thousand acres. Although there are three hydroelectric undertakings in Wales, water resources are mainly exploited by impounding for domestic and industrial purposes. About half of the output of waterworks in Wales serve areas in England.

Industry and agriculture. Despite the fact that coal is a major raw material in Wales, it is the energy-producing industries of gas and electricity, and also metal manufacturing (mainly of iron, steel, and tinplate), that lead in the table of total income produced per person employed. Mining, well down the list, is exceeded by agriculture, by general manufacturing, and by transport, distribution, and construction. In terms of proportions of the labour force employed, pride of place goes to services (including construction), followed by engineering, general manufacturing, distributive trades, metal manufacture, and mining. The various other economic activities are of less significance.

A problem of particular significance in Wales is that of restructuring industry to meet greatly changed economic conditions, especially the decline of the heavy "smokestack" industries. Improved accessibility has made both the southeast and the northeast attractive to new manufacturing activity. The southeast has the largest concentration of Japanese investment in Britain, mainly in electronics. Engineering and pharmaceutical industries have also developed. Government policy has helped; in 1968 the Royal Mint was moved to Llantrisant, in Mid Glamorgan county. With these developments, banking and financial services have expanded, although Wales has no independent financial system. But a (private) Bank of Wales has been established, and two government agencies, the Welsh Development Agency and Mid Wales Development, provide advice as well as financial incentives to industrialists considering moving into Wales.

Diversification of industry

Transportation. For physical, economic, and historical reasons, the main lines of movement in Wales have been lateral, from west to east—that is, along the respective northern and southern coastal belts and across the centre, where they follow the Severn valley into the English Midlands. Subsidiary lines have been from north to south, along the west coast and the English border. Cross-country links have always been poor, and even with increased road modernization they leave much to be desired. There is thus no internally integrated system of transportation, and movement into or out of the country is much easier than movement within it.

Within Wales there is a close and well-developed network of general or all-purpose roads, but those of motorway standard are confined to the links between South Wales and the English Midlands, on the one hand, and between South Wales and London and the Bristol area, on the other. Considerable upgrading to near motorway standards has been undertaken along the North Wales coast.

As part of the rationalization policy of British Rail, Wales experienced a very high rate of closure of branch lines during the late 1950s and in the 1960s. By the late 20th century the end pattern was similar to that of the roads, with through routes following the north and south coasts, each of them providing a link with sea routes to Ireland, the former via Holyhead and the latter via Fishguard. Other routes are minor. Wales has also a number of picturesque narrow-gauge railways that are now operated only during the tourist season, some of them run by private groups of enthusiasts.

Railway closures

Wales has no commercial inland waterways, but there are three categories of ports, all of them, except the Anglesey port of Holyhead, on the south and southwest coasts. In the first category are Holyhead and Fishguard, the two packet stations for the Irish traffic; while the second consists of the South Wales ports whose former coal-exporting functions have been replaced by those of importing iron ore, oil, and general cargo. The third category consists of one port—Milford Haven, which has become one of the major oil-importing and oil-refining centres in western Europe.

Cardiff (Wales) Airport handles domestic flights to other parts of the United Kingdom and international flights to several other countries. Smaller airports are situated at Swansea and at Hawarden.

ADMINISTRATION AND SOCIAL CONDITIONS

Government. The Principality of Wales is governed from Whitehall, the name by which the United Kingdom's political and administrative centre in London is known, but devolution of administration has become an accepted policy of all the major British political parties. The post of secretary of state, with cabinet rank, was created in 1964, and there has been considerable devolution of power to the Welsh Office in Cardiff. In a referendum in 1979, however, Wales overwhelmingly rejected a proposal for a Welsh Assembly, although such a body would have had no legislative powers.

In April 1974, as part of an administrative reorganization throughout the United Kingdom, the administrative structure of Wales was simplified. Many of the former 13 counties, four county boroughs, and numerous smaller units were too limited in extent, population, and resources to be economically feasible under modern conditions. Wales was therefore divided into eight new counties: Clwyd (incorporating the former Flintshire and a large part of Denbighshire), Dyfed (incorporating the former Cardiganshire, Carmarthenshire, and Pembrokeshire), Gwent (incorporating the former county borough of Newport and parts of Monmouthshire and Breconshire); Gwynedd (incorporating the former Anglesey, Merionethshire, and Caernarvonshire and part of Denbighshire), South Glamorgan (incorporating the southern part of the former Glamorgan, part of Monmouthshire, and the county borough of Cardiff, which retained its status as a city and capital of Wales), Mid Glamorgan (incorporating the former county borough of Merthyr Tydfil, the northern part of Glamorgan, and parts of Breconshire and Monmouthshire), West Glamorgan (incorporating the former county borough of Swansea and the western part of Glamorgan), and Powys (incorporating the former Montgomeryshire and Radnorshire and part of Breconshire). The eight counties are subdivided into a total of 37 districts. The functions of these counties and districts are those common to the United Kingdom.

In politics, Wales has always had a strong leaning toward radicalism. During the 19th and early 20th centuries, this was evidenced by the country's overall electoral adherence to the Liberal Party, but the industrial regions soon transferred their allegiance to Socialism. Keir Hardie, for example, the British Labour leader and the first independent Labour member of Parliament, represented the constituency of Merthyr Tydfil, in the heart of the South Wales coalfield. The strong political awareness of the population is evidenced by the fact that the percentage turnout at the polls is consistently higher on average in Wales than it is for Britain as a whole.

The radical tradition in politics

The Welsh Nationalist Party (Plaid Cymru), founded in 1925, is committed to achieving, by democratic and constitutional methods, a measure of political independence that envisages Wales as an independent "region" within the European Economic Community. The party first won a parliamentary seat in a by-election in 1966 and, since that time, has won additional seats in Parliament. On the fringe of this nationalist, separatist movement, and completely disassociated from Plaid Cymru, there are more extremist movements. These have as their aim the realization of political independence based upon cultural and linguistic differentiation. Some of the movements, notably the Welsh Language Society, have been prepared to use various methods of civil disobedience to further these ends.

Police services in Wales are the responsibility of the Home Office in Whitehall. They are operated and adminis-

tered through local police forces or constabularies. Unlike Scotland, Wales has no separate system of justice, and the country has no independent defense arrangements. Three army regiments, however, the Welsh Guards, the Royal Welch Fusiliers, and the Royal Regiment of Wales, have particular associations with the nation.

Education. With its rich cultural heritage, Wales has maintained a tradition of quality education at all levels. Responsibility for the administration of primary and secondary education lies with the secretary of state for Wales and the education department of the Welsh Office, operating through the county Local Education Authorities and the Welsh Joint Education Committee, a body with certain statutory duties and powers. As the name implies, all Local Education Authorities in Wales contribute to and share in the work of this national organization. Higher education in Wales, apart from the University of Wales, also comes under the Welsh Office. The university is financed largely by the Treasury in London operating through the Universities Funding Council.

Health and welfare. As with primary and secondary education, responsibility for health and social services rests with the secretary of state for Wales. The Welsh Office, working in partnership with the local authorities, is responsible for the planning and execution of the housing policy of the government of the day. These responsibilities embrace the provision of new housing as well as the improvement of existing housing conditions, an activity that is of particular importance in the older industrial areas of Wales.

CULTURAL LIFE

Wales has been united politically, administratively, and economically with England ever since the Act of Union in 1536; yet during that period, Wales has been able to preserve, maintain, and develop a cultural identity that, while sharing in that of its larger neighbour, is nevertheless in some important aspects quite different from it. It is the interplay between these two elements, sometimes united, sometimes independent, and sometimes in conflict, that characterizes contemporary cultural life in Wales.

The Welsh cultural identity

The traditional culture was oral and aural in its manifestation and nonmaterial, even unworldly, in its philosophy. It developed into a uniquely blended culture, in which great stress was laid on the spoken and the written word in poetry and in prose; in which vocal music, particularly choral singing, played a great part; and one whose rural, domestic, and noncommercial standards were strengthened and made more austere by the spread of a puritan, religious nonconformity, itself intensified by its associated 19th-century religious revival movements. This ascetic and often erudite culture, essentially of the spirit, sensitive but not sensual, and with no traditional visual expression, now finds itself struggling to maintain its identity in the face of a more material culture that is propagated by ease of communication, including the effects of the mass media, not merely from England but from the whole of the English-speaking world.

The most obvious manifestation of this native culture is, of course, the Welsh language. There has been a substantial educational program to sustain the language, including a nursery school movement, both Welsh-language and bilingual primary and secondary schools, and university degree schemes through the medium of Welsh, as well as facilities for teaching the language to adults.

Government assistance for the arts is operated through the Welsh Arts Council. Funds for literature, art, music, and drama are allocated by the council. Literary activities involve the financial support of publications in both Welsh and English and the granting of bursaries to writers to enable them to undertake specific projects. Tours of Wales by British and foreign orchestras are arranged through the council, and support is given to amateur music-making societies and to a number of music festivals, particularly in respect to commissioning new works. The Welsh National Opera Company, one of the leading companies in Britain, also operates under the aegis of the council. Support is given to art exhibitions, to design, and to a number of Welsh- and English-language theatre companies and theatres. The council also sponsors the formation of regional arts associations, which command financial support from the relevant local authorities.

Cultural institutions

The National Library of Wales, like the British Museum, is a copyright library, receiving, by law, copies of virtually all books published in the United Kingdom. Situated at Aberystwyth and receiving a government grant, it is not merely the national reference library but also a repository of documents and manuscripts relating to Wales from earliest times to the present day.

Another government-aided institution is the National Museum of Wales, situated in Cardiff, with the Welsh Folk Museum in the castle and grounds of nearby St. Fagans. The museum proper embraces the antiquities and the natural history of Wales with, in addition, a very full art collection, while examples of Welsh material culture, such as rural buildings, have been transported to, and reerected in, the grounds of the Folk Museum.

The University of Wales is a federal university with six constituent institutions, located at Aberystwyth, Bangor, Cardiff, Lampeter, and Swansea, with the Welsh National School of Medicine at Cardiff. The originally independent College of Librarianship Wales became part of the University College of Wales at Aberystwyth in 1989. The Polytechnic of Wales is located near Pontypridd.

The above institutions are all on permanent sites, but the Royal National Eisteddfod of Wales, a competitive festival and perhaps the most individual of Welsh cultural institutions, is peripatetic in that it is held annually but alternately in North and South Wales. The actual venue results from the invitation of a particular town or area, and the festival is housed in a transportable hall or pavilion. It is always sited on open ground that, for the duration of the Eisteddfod, becomes a campus on which most of the cultural, and some of the commercial, organizations of Wales also hold exhibitions illustrating their activities. The Eisteddfod itself is held for one week in August and consists of competitions in all aspects of music, literature, drama, and art, together with a series of dramatic performances and concerts, all held through the medium of the Welsh language. During the period of the Eisteddfod, and associated with it, the bardic circle, or Gorsedd, holds its ceremonies. For this week, the Eisteddfod field is the gathering ground for the nation, but local Eisteddfodau are held in towns and villages throughout the year.

The Welsh region of the BBC, which has always had considerable independence, provides service in both English and Welsh. HTV Ltd., a commercial company, covers Wales and the western part of England. A Welsh-language television channel, usually known as S4C (Sianel Pedwar Cymru, or Channel Four Wales), began broadcasting in 1982. (J.G.Th./H.Ca.)

History

WALES BEFORE THE NORMAN CONQUEST

The prehistory of Wales. Meaningful study of prehistoric Wales has to be pursued against the broader background of British prehistory, for the material remains of the period 3500–1000 BC, especially funerary monuments, provide regional manifestations of features characteristic of Britain as a whole. The Celtic origins of Britain, probably to be sought in a gradual process within the last millennium BC, are a matter of continuing scholarly debate. Traditional archaeological and linguistic interpretation emphasizes an influx, from the late Bronze Age onward, of Celtic-speaking peoples, though not perhaps in vast numbers, and a dynamic relationship between continental and insular communities. Modern views emphasize that the ethnogenesis of the Celts must be seen as a complex process of social change and not entirely the result of migrations. As regards their social structure, the metalwork associated with feasting and military prowess, such as that found at Llyn Fawr and Llyn Cerrig Bach, coupled with the broad distribution of fortified sites, typifies the highly stratified but politically fragmented and warlike society which prevailed in Wales down to the Roman period.

Roman Wales (1st–4th century). Wales in the Roman period shared broadly the experience of other parts of

highland Britain, but modern archaeological study has tended to moderate the traditional contrast drawn between military and civil zones. Mediterranean culture is best exemplified in southern Wales, where there were important Roman towns at Caerwent and Carmarthen and villas at a number of other sites. Remains elsewhere consist mainly of the roads and forts of a phase of military occupation that lasted to about AD 200. But at Segontium (Caernarfon) there was a continuous well-ordered settlement to about AD 400, and it is likely that civil influences were exerted much more widely than was once thought. Linguistic study suggests that the native language, known to scholars as Brittonic or Brythonic, was infused with Latin terms, though distinction needs to be made between borrowings of the period of Roman rule and the scholarly borrowings of subsequent periods. Early Welsh consciousness of a Roman heritage may owe a great deal to the Latinity sustained in later centuries by the Christian church.

The founding of the kingdoms. The origin of early Welsh political organization must be sought in the period following the cessation of Roman rule in about AD 400. Native leaders, unable to sustain Roman methods of governance, initiated the processes that were to lead to the founding of a number of kingdoms. The *Historia Brittonum,* an antiquarian compilation dating from the early 9th century, explains the origin of the kingdom of Gwynedd by relating a tradition that Cunedda Wledig migrated from northern Britain to northwestern Wales to expel the Irish who had occupied the area. This may be an example of the origin stories that were current in early medieval Europe, and the *Historia* also contains an early reference to the Welsh claim to Trojan origin, which was to prove an enduring theme in Welsh historical consciousness. Tradition attributes the names of the various parts of Gwynedd, such as Dunoding and Rhufoniog, to a division of the kingdom said to have been made among Cunedda's sons after his death, but these may be the names of territories that were gradually incorporated into the kingdom during a long period of growth. Cunedda's descendants were to rule as kings; a 7th-century representative of the dynasty is commemorated upon an inscribed stone in Anglesey as Catamanus Rex (Cadfan the King).

Cunedda Wledig

In southwestern Wales the Irish presence led to the founding of the Irish kingdom of Dyfed, and some Irish influence was felt further afield in the neighbouring lands of Ceredigion, Ystrad Tywi, and Brycheiniog. In the southeast Glywysig and Gwent emerged, to be united, though impermanently, to form Morgannwg. In north-central Wales the kingdom of Powys, originally centred at Pengwern (a place not identified with certainty), was established, and it embraced at least part of the Roman province of the Cornovii, centred at Wroxeter in Shropshire.

Early Christianity. There are indications of a Romano-British Christian church in southeastern Wales, but Christian influence may also have penetrated much deeper into Wales in the Roman period. Inscribed stones, though themselves belonging to the 5th or 6th century, carry terms such as *sacerdos* (probably meaning bishop) and *presbyter* (priest), which may reflect a well-established Christian church of early origin. Stones with Irish (or Ogham) inscriptions and Christian symbols in southwestern Wales suggest that the immigrants, if not already Christian upon arrival, were Christianized soon afterward. The extent of the continuity of the early Romano-British church, however, has become the subject of scholarly debate. From the mid-20th century onward historians have argued in favour of discontinuity by stressing the importance of what they considered to be new Gallo-Roman influences exerted on western Britain in the 5th and 6th centuries. According to this view Celtic "saints" coming from the Continent reestablished Christianity by their missionary activities along the western seaways. More recently, however, the view held by an earlier generation of historians that the Christian church had a continuous existence from the Romano-British period has regained support. Scholars defending this position argue that much of the evidence for the missionary activity of Celtic saints was derived from the saints' lives and church dedications of a later date. Basing their arguments on the *De excidio et conquestu Britanniae,* a work by the 6th-century British monk Gildas that suggests a long-established Christian tradition of Romano-British derivation, they postulate that the trade and cultural contacts along the western seaways may have served not to introduce Christianity or to revitalize a lingering faith but to bring to an existing church a form of monasticism that had proved to be an important influence in the development of the Gallic church.

Continuity of the Romano-British church

Support for the argument of a "Celtic church" rests upon the church's monastic character. A major church (*clas,* plural, *clasau*) was headed by an abbot and bishop who was responsible for daughter houses. The *clas* was not a cloistered community, and its head was responsible for the ordination of priests and the pastoral care of the laity in neighbouring areas. But the hereditary succession to office and to ecclesiastical property that developed with time was among traditional practices, and even though at places like Llanbadarn Fawr there were ecclesiastical families who maintained a Christian learning of a high order, these practices were considered to be contrary to the teaching of the reformed church.

Political development. The settlement of Anglo-Saxon peoples along the Welsh borderland separated the Brythonic peoples of Wales from those of northern and southwestern Britain. Whereas to the English they were "Welsh" (foreigners), they identified themselves as "Cymry" (compatriots). Offa's Dyke, the great linear earthwork built in the times of King Offa (d. 796) of Mercia, represents the demarcation line of English penetration into Wales.

Attempts at unity

Attempts during the next two centuries to bring the Welsh kingdoms west of the dike into a political unity proved to be only partially successful and impermanent. Rhodri Mawr ("the Great"; d. 878), the king of Gwynedd who provided stern resistance to the Viking attacks, brought Powys within his dominion and then briefly extended his sway over two areas in the southwest (lying north and east of Dyfed), namely Ceredigion and Ystrad Tywi, which had previously been united to form the kingdom of Seisyllwg. The period following Rhodri's death proved to be of far-reaching significance. The outlying kingdoms of Wales—Dyfed, Brycheiniog, Glywysing, and Gwent—being subjected to pressure by Rhodri's sons or by Mercia, turned to the kingdom of Wessex and by a formal commendation entered into that allegiance, ultimately expressed in homage and fealty, which each of the kings of Wales owed, individually and directly, to the English monarchy. Anarawd (d. 916), a son of Rhodri, subsequently submitted to Alfred (d. 899) and completed the formal subjection of the Welsh kingdoms to the English sovereign. Rhodri's grandson, Hywel ap Cadell (Hywel Dda, "the Good"; d. 950), starting from a patrimony in Seisyllwg, secured Dyfed by marriage, thereby creating the kingdom of Deheubarth. Eventually Gwynedd and Powys also came under his rule. Hywel, possibly inspired by admiration for the Wessex court but more probably constrained by the power of King Athelstan (d. 939), accepted the status of a *sub-regulus,* or under-king, of the king of Wessex. But whatever its compulsions, Hywel's policy provoked a reaction (expressed in the poem *Armes Prydein*) that envisaged the formation of a great alliance to withstand the Anglo-Saxon suzerain.

Before the close of the 10th century Maredudd ap Owain (d. 999), a grandson of Hywel Dda, brought the northern and western kingdoms once more into a transitory unity. But his death opened a period of prolonged turmoil in which internal conflicts were complicated and intensified by Anglo-Saxon and Norse intervention. The established dynasties were challenged by men who asserted themselves within the kingdoms and exercised ephemeral supremacies. Of these, the most successful was Gruffudd ap Llywelyn (d. 1063), who brought Gwynedd, then Deheubarth, and finally (though briefly) the whole of Wales under his dominion. The devastation wrought upon the English borderland, still not erased at the time of the making of Domesday Book (1086), was probably in large measure due to him. His death in 1063 meant that the most powerful ruler of independent Wales was destroyed only a few years before the Norman forces came to the Anglo-Welsh frontier.

Temporary unity under Gruffudd

Early Welsh society. The endeavours of the dynasties in the 9th and 10th centuries, though only partially successful with regard to the problem of Welsh unification, had important and lasting consequences. Scholarly activity such as that represented in the *Historia Brittonum* and in annals and genealogies, material relating both to northern Britain and to Wales, may well reflect the attempt of the descendants of Rhodri Mawr to consolidate their position and enhance their prestige. With regard to creative literature, it is likely that the origins of some texts preserved in medieval manuscripts, including some material in triad form (triple groupings of legal, literary, historical, and other materials), may be traced back to this period. The earliest Welsh law texts, though they date from the 13th century onward, attribute the original codification of law to Hywel Dda; and it is possible that a significant development in Welsh jurisprudence took place under the aegis of that ruler. These texts, along with other materials, reveal a society of relatively settled kingdoms ruled by kings (*brenhinoedd,* singular, *brenin*) who were endowed with an extensive range of powers, notably the public enforcement of legal obligations.

The kingdoms were normally divided for purposes of royal administration into *cantrefs.* These in turn consisted of groups of *maenors* occupied by the bond or free elements of which Welsh society was composed. The bond population, which was probably larger than once thought and which was concentrated in fairly compact *maenors* in lowland areas that were favourable to an agrarian economy, was organized on conventional manorial principles. In the economy of the upland areas the emphasis was upon a pastoral economy practiced by free communities, which were accorded more extensive *maenors.* As a result of changes that quickened considerably in the 12th century, the *maenor* organization of Welsh society was superseded by new forms designed to ensure a more intensive exploitation of the soil. A smaller unit, the *tref,* or township, then replaced the *maenor.* In the sphere of royal administration the *cantref,* by a process probably already well advanced on the eve of the first Norman invasions, was largely replaced by a small unit, the *commote,* which was to remain, under Welsh and alien lords, the basic unit of administration and jurisdiction throughout the medieval period.

WALES IN THE MIDDLE AGES

Norman infiltration. The Norman Conquest of England saw the establishment upon the Welsh border of the three earldoms of Chester, Shrewsbury, and Hereford, and from each of these strongpoints advances were made into Wales. Norman progress in southern Wales in the reign of William I (1066–87) was limited to the colonization of Gwent in the southeast. Domesday Book contains evidence suggesting that King William and Rhys ap Tewdwr, king of Deheubarth (d. 1093), made a compact that recognized the Welsh ruler's authority in his own kingdom and perhaps also his influence in those other areas of southern Wales outside Deheubarth, particularly Morgannwg and Brycheiniog, that still lay outside Norman control. Meanwhile, from Chester and Shrewsbury, the Normans had penetrated more deeply into Wales, so that at Domesday, though the area colonized was limited, Norman lordship had been asserted over numerous *cantrefs* and *commotes* that had previously formed portions of the kingdoms of Gwynedd and Powys. The political situations in the northern and southern parts of the country were reversed during a period of renewed conflict in the reign of William II (1087–1100). Soon after Rhys ap Tewdwr's death in 1093 while opposing the Norman advance into Brycheiniog, the Normans invaded virtually the whole of southern Wales. Advances from several bases along the Welsh border enabled Norman lords to establish the major lordships of Cardigan, Pembroke, Brecon, and Glamorgan. This advance constituted the decisive stage in the creation of the March of Wales; in this land, consisting of lordships, Norman lords and their successors exercised rights founded on the powers previously enjoyed by the Welsh kings but greatly expanded so as to give the lords, under "the custom of the March," extensive powers in their lordships and a large measure of autonomy in their relations with the king of England.

Creation of the March

Gwynedd, Powys, and Deheubarth. The crucial years after 1093 saw also the initiation in northern Wales of a period of conflict by which the area was gradually recovered from Norman rule and the kingdoms of Gwynedd and Powys reconstituted as major political entities. Gwynedd, first under Gruffudd ap Cynan (d. 1137) and then under his son Owain Gwynedd (d. 1170), gained a firm governance that enabled the younger ruler, controlling a kingdom extending from the Dyfi to the Dee, to withstand foreign pressure, which was particularly severe during the reign of Henry II (1154–89). In Powys the rule of Madog ap Maredudd (d. 1160) likewise proved to be a period of stability and of expansion eastward beyond Offa's Dyke into lands that had been subjected to alien settlement in both the Anglo-Saxon and Norman periods. In southwestern Wales, too, representatives of the dynasty of Deheubarth for more than 30 years waged a campaign that finally enabled Rhys ap Gruffudd (d. 1197), a grandson of Rhys ap Tewdwr, to win from Henry II a recognition of his position. Rhys ruled a land that was not as extensive as the ancient kingdom, for Norman control of the lordship of Pembroke and of other lordships along the southern coastline was conceded, but it nevertheless constituted a considerable dominion.

The three kingdoms of Gwynedd, Powys, and Deheubarth formed by the third quarter of the 12th century a well-defined sphere of Welsh political influence (Wallia, or Pura Wallia) in contradistinction to the sphere of Norman influence (Marchia Wallie). Throughout the remainder of the period of Welsh independence there remained a memory that Wales, outside the March, had consisted historically of three kingdoms ruled from the three principal seats of Aberffraw in Gwynedd, Mathrafal in Powys, and Dinefwr in Deheubarth. The rulers of these three kingdoms formulated a concept of Welsh kingship in which indigenous elements were blended with the new influences at work in the feudal monarchies. Each ruler, still known as a king (*rex, brenin*) but later to be styled prince (*princeps, tywysog*) or lord (*dominus, arglwydd*), governed an autonomous territory for which he did homage and fealty to the king of England.

Welsh kings outside the March

Political stability enabled these territories to recover from the depredations of the Norman period, and the rulers sought to increase the resources of their demesne lands both by exploiting the labour services of bondmen and by providing some bondmen with more favourable tenurial conditions as an incentive to the colonization of marginal lands. With regard to lands held by freemen, a trend toward more intensive agricultural exploitation and a more precise definition of fiscal obligations may explain the description in late medieval land surveys of territorial assets vested in lineages that often traced their descent from a 12th-century ancestor. The endowment of some privileged proprietors with extensive estates facilitated, despite continued adherence to partible succession, the growth of a class of landowners who were linked with the rulers by ties of service and provided the personnel of their administration.

A renewed cultural vitality is noticeable in the Latin scholarship of this period and in a flowering of the literary tradition, exemplified in prose and in eulogistic poetry. *The History of Gruffydd ap Cynan,* probably written in the reign of his son Owain Gwynedd, provides a classic statement of the political and cultural values of independent Wales. Emphasizing the stability and prosperity of an ordered society, it provides an indigenous counterpoint to the more critical view of Welsh society embodied most notably, despite his subsequent identification with the cause of an independent Welsh church, in the works of the Welsh historian Giraldus Cambrensis.

In ecclesiastical affairs, the early Norman period saw the inauguration of a process by which the *clas* organization was replaced by arrangements consonant with the practice of the reformed church. The four territorial dioceses of Bangor, St. David's, Llandaff, and St. Asaph were created, and a parochial organization was gradually established. The church structure was a creation of the Normans, and

the bishops appointed to Welsh sees owed a profession of obedience to Canterbury. Even so, Bernard, bishop of St. David's in 1115–48, claimed the status of an archbishop and, in furthering his campaign, appealed to the historical legacy of an early independent Welsh church. His bid was revived at the end of the century by Giraldus Cambrensis. But no less significant than Giraldus' endeavour was the resistance of the clergy of Bangor, who, acting under the protection of Owain Gwynedd at a time of national resistance toward the end of his reign, steadfastly refused to meet the demands of Thomas Becket, archbishop of Canterbury, that the newly elected bishop should swear fealty to Canterbury. The lay powers found adherents in the Cistercian Order; houses such as Margam and Tintern, situated in the March, had close associations with their marcher patrons. The offshoots of the Cistercian monastery of Whitland, notably Strata Florida and Aberconway, were handsomely endowed by the Welsh rulers, who in return were supported in their political endeavours.

Llywelyn ap Iorwerth. In each of the three kingdoms of Gwynedd, Powys, and Deheubarth, the death of its powerful ruler was followed by a contested succession. In Powys and Deheubarth the unity of the kingdom was never restored; but with the emergence to power in the late 12th century of Llywelyn ap Iorwerth (d. 1240), a grandson of Owain Gwynedd, Gwynedd was united once more under the strong hand of a single ruler. Llywelyn's aggression against neighbouring territories incurred resistance, which King John turned to his advantage in a campaign in 1211 whereby the prince of Gwynedd was subjected to humiliating terms. But availing himself of a general Welsh reaction to John's measures for the permanent subjugation of the country, Llywelyn directed a sustained campaign in which his former adversaries participated. Llywelyn achieved a dominant position among the princes, which, while the contest with John persisted, augured the forging of a Welsh polity by bonds of homage and fealty to himself. But, though he remained a powerful influence over the other Welsh princes and thereby minimized the crown's involvement in the affairs of Wales, Llywelyn was unable to secure a formal royal recognition of the territorial and conceptual achievements of the period of conflict. Llywelyn's aspirations for a wider Welsh principality based upon the supremacy of Gwynedd then centred upon David ap Llywelyn, his son by Joan, daughter of King John. David was designated as Llywelyn's heir in preference to his elder but bastard son, Gruffudd, and the Welsh dynasty looked to the English monarchy to ensure an unchallenged succession. In the event, the crown was able to use the dissension between the two sons and the disparate ambitions of the other Welsh princes to restrict David's power to Gwynedd alone. During the war of 1244–46 David contended for a broader influence, but his promising endeavour was cut short by his early death in 1246, without heir.

United Gwynedd

In the following year his nephews Owain and Llywelyn, two of the four sons of Gruffudd, entered into a treaty obligation by which the crown decreed the partition of a truncated Gwynedd into two parts, with the prospect of further division to provide for the younger brothers. But between 1255 and 1258 Llywelyn ap Gruffudd (d. 1282), one of the four brothers, asserted his supremacy first in Gwynedd and then farther afield. In this he was helped by the preoccupation of the English crown with the baronial conflict that led to the Provisions of Oxford in 1258. The Prince secured a hegemony that was formally acknowledged by Henry III in 1267 by the Treaty of Montgomery, in which Llywelyn's style, "prince of Wales," first assumed in 1258, and his right to the homage and fealty of the Welsh lords of Wales were recognized. Llywelyn had thereby brought into being a Principality of Wales composed of the lands that had formed the 12th-century kingdoms of Gwynedd, Powys, and Deheubarth as well as parts of the March. Historically, this meant the reversal of a situation, for which there were several centuries of precedent, whereby the increasingly fragmented territories under Welsh rule had been fiefs held directly from the king of England. The opportunity to consolidate the governance of the principality proved to be brief.

War with Edward I

Friction between Llywelyn and Edward I led in 1277 to a war in which the prince, isolated by the withdrawal of his vassals' fealty and confronted with the great resources and superior organization of England, was forced to accept terms that restricted his power to Gwynedd west of the Conway. By 1282 a deterioration in relations between Edward and a number of Welsh princes resulted in renewed conflict. Although Llywelyn may not have been the instigator of the rebellion, he placed himself at its head. In his negotiations with Archbishop Pecham late in 1282 he forcefully expressed the aspirations that had inspired his great endeavour to secure the internal unity of Wales and to stabilize its relationship with England. Shortly afterward, on December 11, Llywelyn was slain in combat, and the resistance, though sustained by his brother David ap Gruffydd (d. 1283) for several months, finally collapsed in the summer of 1283.

The Edwardian settlement. Edward I provided for the security of his conquests by means of a program of castle building, initiated after the war of 1277 and subsequently extended to include the great structures of Conway, Caernarfon, Harlech, and, later, Beaumaris. Each castle sheltered a borough where English colonists were settled. The king's arrangements for the governance of Llywelyn's former lands in northwestern Wales were embodied in the Statute of Wales (1284). Three counties—Anglesey, Caernarfon, and Merioneth—were created and placed under the custody of a justice of North Wales. In northeastern Wales a fourth county, Flint, was attached to the earldom of Chester. In southwestern Wales the counties of Cardigan and Carmarthen, under the custody of the justice of West Wales, were formed out of lands over which royal power had been gradually extended by a process completed upon the failure, in 1287, of the revolt of Rhys ap Maredudd, the last of the princes of the dynasty of Deheubarth. Structurally, the shires that formed the Principality of Wales were similar to those of England, and certain common-law procedures were introduced into their courts, but the shires remained outside the jurisdiction of the central courts of Westminster and they did not elect representatives to Parliament. The March of Wales was extended through the creation by royal charters, out of parts of Gwynedd and Powys, of the lordships of Denbigh, Ruthin, Bromfield and Yale, and Chirk. In his relations with two of the major barons of the older March, Gilbert de Clare of Glamorgan and Humphrey de Bohun of Brecon, Edward showed a determination to assert the sovereignty of the crown over the March and to eradicate abuses of the Custom of the March such as the claim, defiantly expressed by Gilbert, to the right to wage war in the March. But neither Edward nor his successors attempted any far-reaching changes in the organization of the March, and political fractionization persisted over the next two centuries.

The March of Wales

Rebellion and annexation. Both the crown and the marcher lords employed in the administration of their lands Welshmen drawn from an administrative class that had been fostered by the princes themselves. Those of the principality revealed a particular loyalty to Edward II in the political crises of his reign, and their continued attachment to his cause even after his deposition created a tense situation in 1327. During the 14th century there were occasional variances, but the identity of interest established between the crown and the leading Welshmen proved durable. Even so, the community endured both the economic difficulties encountered over wide areas of Europe at this time and the specifically Welsh problems created by the fact that an important phase in the transition from early medieval social arrangements coincided with the pressures exerted by an alien and fiscally extractive administration. At the very end of the century the deposition of Richard II, who had influential Welshmen among his partisans, released from allegiance to the monarchy a group that, associated with Owain Glyndwr (Owen Glendower), raised a great rebellion which drew its strength from the community as a whole. In the period 1400–07 the royal government lost control of the greater part of Wales, and in some areas the insurrection remained unextinguished several years later.

The rebellion, however, quickened certain processes that were to lead ultimately to the enfranchisement provided by Tudor legislation. In northern Wales particularly, the availability of civil actions by English law led to an early but unrequited demand for English land law. After the rebellion the disabilities incurred by reason of Welsh nationality were underlined. Although often expressed in literature in militant terms and, during the years of dynastic conflict, manipulated by the protagonists of York and Lancaster, the aspirations of the community were focused in a demand for English denizenship. First individual petitioners looked for enfranchisement, and then whole communities in northern Wales secured from Henry VII, by negotiation and payment, charters conferring upon them English land law and other advantages. A realization by the crown of its inability to reverse a decline in the financial yield of its Welsh lands, an experience shared by the marcher lords, contributed to Henry VII's policy.

WALES FROM THE 16TH TO THE 20TH CENTURY

Union with England. In 1536 Henry VIII's government enacted a measure that made important changes in the government of Wales. Whereas the Statute of Wales (1284) had annexed Wales to the crown of England, the new act declared the king's wish to incorporate Wales within the realm. One of its main effects was to secure "the shiring of the Marches," bringing the numerous marcher lordships within a comprehensive system of counties. For the first time in its history Wales was to have uniformity in the administration of justice. Welshmen were to enjoy the same political status as Englishmen, and the common law of England, rather than Welsh law, was to be used in the courts. Wales also secured parliamentary representation by the election of members for shires and boroughs. The implementation of the act was set aside until more detailed provision was made by a second act in 1543. Statutory recognition was now given to the Council of Wales and the Marches, which exercised a jurisdiction over both Wales and four border counties of England. But the council fell into abeyance during the Civil Wars and was finally abolished after the revolution of 1688.

Acts of 1536 and 1543

In 1543 the Courts of Great Sessions were also created, modeled on the practice already used in the three counties that, since 1284, had formed the principality of North Wales (Anglesey, Caernarfon, and Merioneth), but with 12 counties now grouped into four judicial circuits and the 13th, Monmouthshire, linked with the Oxford circuit. The Great Sessions remained the higher courts of Wales until 1830 when, despite considerable opposition, they were abolished. Finally the Courts of Quarter Sessions were instituted in the manner in which they were already held in England, with the administration of the law vested in justices of the peace. Besides their judicial functions, the justices undertook a wide range of administrative duties, which they continued to fulfill until, with the reform of local government by the Local Government Act of 1888, the county councils were established.

The Reformation. Enacted in the wake of Henry VIII's break with Rome, the union legislation, stressing the need for uniformity with the realm, required those who participated in administration under the crown to use the English language. The need to secure the Protestant faith, however, was to lead to an acknowledgement that the Welsh language had to be used in public worship. William Salesbury and Richard Davies were among a group of distinguished scholars, motivated both by Protestant conviction and passionate concern for the nation's cultural heritage, who realized that the provision of the Scriptures and the Book of Common Prayer in Welsh was essential for the promotion of the faith and the vitality of the language. A petition to the Privy Council led to an act of Parliament in 1563 that required the translation of the Bible and the Book of Common Prayer into Welsh by 1567. Translations of the New Testament and the Book of Common Prayer were indeed published in 1567. The New Testament included an introductory essay by Davies that interpreted the establishment of the Protestant faith as a restoration of the true religion, which had flourished in Wales before the corrupt faith of Rome had been imposed upon its inhabitants. The demands of the Elizabethan government and the aspirations of the Welsh Protestant humanists were met in full when William Morgan's translation of the entire Bible appeared in 1588. Alone among the Celtic nations in securing the Scriptures in the vernacular within half a century of the Reformation, the Welsh people had scored a success of profound significance for the future of the language and the sense of nationhood. Scholarly devotion to the language, also shown by Catholic exiles such as Gruffydd Robert, was accompanied by new interest in Welsh antiquities, and the work of the 16th-century historians Humphrey Llwyd, David Powel, George Owen, and their successors conserved the heritage of the Middle Ages and laid the foundations of modern historical scholarship.

Morgan's translation of the Bible

Social change. Wales in the mid-16th century probably had a population not much above 250,000, though it was by then growing once more after a period of prolonged stagnation. Towns, though often prosperous, remained small, and Wales possessed no major urban and commercial focus. Although industrial enterprises had an effect upon the economy of certain localities and some Welshmen were enriched by entrepreneurial ventures outside Wales, income was largely derived, directly and indirectly, from pastoral and arable farming. During the 16th and 17th centuries gentry estates were enlarged and consolidated, and the holdings of innumerable proprietors were absorbed into the larger estates. Consequent changes in tenurial status, the growth in population, and inflation created intense problems that were only partially relieved by the enclosure of waste areas and the cultivation of marginal lands. While more and more land was concentrated in fewer hands, smaller proprietors who retained their stake in the soil were often forced to divide their holdings and convert summer dwellings, hitherto used by shepherds in summer months, into permanent homesteads. Many were forced off the land altogether, and it was not until the late 18th century that industry became a major outlet for a rural population which the land could not sustain.

Effects on small landowners

Social trends and the interplay of indigenous and foreign influences were reflected in domestic architecture. The timber-framed hall house, already characteristic of the eastern borderland and of the northern parts of Wales in the late Middle Ages, continued to represent a strong vernacular tradition. But the varying scale and refinement of the houses told of a growing disparity in wealth. In some areas, notably in Glamorgan and Monmouthshire, where a tradition in masonry houses had long existed, vernacular characteristics were increasingly set aside in favour of a new type of Renaissance house. Some houses were built on a scale that indicated the emergence of a class of great landowners who were to stand apart from Welsh society at large on account not only of their wealth but also of their intermarriage with English and Scottish families.

Politics and religion, 1640–1800. On the eve of the Civil War in 1642 there was much sympathy for the royalist cause in Wales. But the parliamentarians also found adherents among some landowners, such as Robert Devereux, 3rd Earl of Essex, and Thomas Myddelton, as well as among individuals committed to the Puritan cause, such as the writer Morgan Llwyd and the zealous soldier John Jones of Maesygarnedd. It was mainly in the border counties and in Pembrokeshire, however, that Puritan influence and commercial contacts served to win support for the parliamentary cause. The imposition of parliamentary power on Wales and the sequestration of royalists' property incurred resentment, and Puritan missionaries found themselves labouring in what they believed to be a dark corner of the land. The Act for the Propagation of the Gospel in Wales (1650) set up a coercive authority encompassing both political and religious life, but state intervention remained largely unproductive.

Nonetheless, the Interregnum saw the formation of Dissenting congregations, which were to lay the foundations for some of the abiding influences of modern Welsh life. The most radical were the Quakers who, making particular headway in Montgomeryshire and Merioneth, penetrated not only Anglicized border territory but also the heart of the Welsh-speaking areas. Incurring the animosity of

churchmen and other Dissenters alike, they were repressed with a severity experienced only by Roman Catholics and forced into emigration to Pennsylvania, in large numbers. On the other hand, small gathered churches of Congregationalists and Baptists, whose theology was Calvinist and whose belief and personal conduct were governed by a strict code expounded in their church covenant, established the Dissenting tradition within rural communities and small towns.

In the 18th century Methodism became a new and potent influence. Launched by a revival movement of great intensity in the years after 1735, Methodism was sustained within the established church by means of local societies and a central association. The combined influences of the old Dissent and the new Methodism, however, eventually transformed the religious adherence of the Welsh people at the expense of the established church. Although served by innumerable men of learning and devotion, among them Griffith Jones, whose circulating schools contributed immeasurably to the growth in literacy, the church was racked by poverty and inadequate leadership. Thus the Methodist secession from the Anglican church made the ultimate triumph of Nonconformity inevitable.

Methodism and Dissent were not the only influences at work in 18th-century Wales. The resilience of a native culture no longer able to depend upon traditional sources of patronage showed itself in a patriotic fervour to preserve a cultural heritage threatened by progressive Anglicization. Although its proponents drew upon Welsh scholarly achievements, notably those of Edward Lhuyd, Wales had no academic institutions capable of appraising critically the work of romantic antiquarians who looked back to Celtic myth and British druidism. Yet despite its shortcomings, the 18th-century cultural movement was an important expression of a preindustrial society's resourcefulness in protecting its heritage. One of its key figures was Edward Williams (Iolo Morganwg), whose endeavours encompassed a vast range of literary and historical studies and who also represented the political radicalism inspired by the French Revolution. Radical convictions were held only by a small minority, some of them eccentrics and others distinguished expatriates, but their endeavours marked a significant stage in the emergence of a distinctively Welsh political consciousness.

Movement to preserve native culture

The growth of industrial society. By 1800 Wales was rapidly ceasing to be a land whose people were almost entirely dependent upon a rural economy. Industrial development, already present in certain localities, now took place on a larger scale. There was considerable development in the coalfield of northeastern Wales; in the southwest, in Swansea, copper smelting, in particular, served to make the town an important metallurgical centre, and for a period it also could count fine porcelain among its range of manufactures. The main industrial expansion occurred, however, with the growth of substantial ironworks on the northern rim of the South Wales coalfield in Glamorgan and Monmouthshire. Not hitherto served by any major urban centres, the area now became densely populated, largely as a result of immigration from other areas of Wales. A natural increase in population accelerated considerably by the late 18th century, with an estimated population for Wales of 450,000 in 1750 rising to 587,245 by 1801. The population continued to grow, again mainly by natural increase, to 1,163,139 by 1851 and to 2,012,875 by 1901. It was only in the following decade that immigration into Wales occurred on a massive scale. In this period 126,529 persons, the majority of whom came from outside Wales, migrated into Glamorgan and Monmouthshire alone, further enlarging the population of the counties that had benefited most from the internal migration of the earlier decades.

Industrial growth made it possible for large numbers of people whom the rural economy could not sustain to find a livelihood within Wales, and the industrial communities of Glamorgan and Monmouthshire contained a substantial Welsh-speaking element throughout the 19th century. Merthyr Tydfil grew rapidly to become the main urban centre of a new industrial society, but it sadly lacked the facilities normally associated with settlements of a more gradual growth. Conditions in the mining areas were harsh; workers were subjected to long hours and low wages, their children were often forced to work at the mines from a young age, and their families lived in wretched and overcrowded circumstances. It was in Merthyr Tydfil that industrial and social unrest, first expressed in wage-related disputes and sporadic rioting in several areas, erupted into a serious rising in 1831. In the following years the workers' main channel for expressing their aspirations was Chartism, which found its most forceful manifestation in the insurrection at Newport in 1839. Rural Wales, too, was subject to social unrest, and between 1839 and 1843 the Rebecca Riots, ostensibly directed against the exaction of road tolls, gave expression to the underlying difficulties of the tenant farmers of southwestern Wales.

Industrial and social unrest

Improved economic conditions from the middle years of the century onward ushered in a period of comparative quietude in industrial Wales. A new phase in industrial growth, brought about by the exploitation of the steam-coal reserves of Glamorgan and Monmouthshire during the last decades of the century, created new valley communities that drew immigrants both from rural and industrial settlements in Wales and from elsewhere. The growth in coal exports led to the building of docks on the coast between Newport and Swansea, notably at Cardiff, while increasing dependence on imported ore led to the relocation of the growing steel industry to areas in close proximity to the coast.

Political radicalism. Wales only gradually embraced the radicalism that came to be regarded as its traditional political allegiance. The political exigencies of the years of the French wars of the Napoleonic period forced the Methodists, in particular, into a passivity that was underlined by a sterner interpretation of Calvinist theology. Political passions were aroused, however, by the moral indictment of the Welsh nation, and especially its Nonconformity, carried in the Report of the Commissioners for Education in 1847. The Welsh, especially the women, were portrayed as depraved and immoral, backward and ignorant. Even the more able ones among them were thought to be impeded by their theological wrangling and often by their lack of English. The growth of a Welsh-language periodical press from its largely denominational origins into a distinctively radical force proved an important influence in rural and industrial Wales alike. By the later years of the 19th century, following the franchise reforms of 1867 and 1884, the hegemony of Welsh Liberal Nonconformity was well established. The passing of legislation specifically concerned with Wales, such as the Welsh Intermediate Education Act (1889) and the Church Disestablishment Act (1914), was a parliamentary success matched in cultural life by the founding of three university colleges and the federal University of Wales and the securing of a royal charter for the establishment of the National Library of Wales and the National Museum of Wales. The attempt by Welsh Liberal associations to secure a representative assembly reached its peak with the Home Rule movement of 1886–96, but it was wrecked by dissension within the associations.

Advances in education

The 20th century. By 1900 there were signs that the Liberal-Nonconformist supremacy would be gradually undermined. Traditional beliefs were challenged, and the experience of World War I created new tensions. The massive flow of workers into the steel and coal areas, largely from outside Wales, affected the composition, and hence the language, of the industrial communities, and immigration coincided with a new era of industrial unrest and political militancy. The miners' efforts at combination led in 1898 to the founding of the South Wales Miners' Federation; the coal owners strengthened their position by forming powerful combines. Despite fierce resistance, the miners won their campaigns for an 8-hour day and a minimum wage. Within the federation a new militancy, expressed in the policy document entitled the *Miners' Next Step* (1912), espoused an industrial unionism with syndicalist tendencies. These influences, though potent in the Rhondda Valley, did not pervade the coal industry, nor did they shape the steelworkers' and tinplate workers' unions. After the war syndicalist influence was subsumed

in orthodox communism, or, more generally, in democratic socialism. By 1922 the Liberal Party in South Wales had lost its hold upon the industrial communities to the Labour Party, whose influence was felt both at Westminster and in local government. In the northeast and in the slate-quarrying communities of northwestern Wales, which also experienced prolonged and enervating industrial disputes, an allegiance to radical Liberalism gradually evolved into adherence to Labour.

Economic depression between the world wars, made particularly acute by the collapse of the export market upon which the Welsh economy depended, brought massive unemployment. Wales lost some 430,000 people by emigration to England and overseas. The war years of 1939–45 brought industrial recovery, and upon the cessation of hostilities strenuous efforts were made to modernize the basic industries of steel and coal and to achieve a diversification of industry. Even so, coal production continued to fall, and in the industrial areas the exodus of workers and their families continued. The rural economy benefited from government subsidies that facilitated investments in land improvements and new buildings as well as from advisory and marketing agencies. Nonetheless, even in the rural areas the working population continued to fall.

Labour remained the main political allegiance of the Welsh people. After World War II gains were registered in rural Wales, until in 1966 as many as 32 of the 36 Welsh seats were held by Labour. This position was not maintained, however. Plaid Cymru, founded in 1925 as the Welsh Nationalist Party, achieved a measure of influence that was not reflected in successes at parliamentary elections until 1966. Although success at Carmarthen in that year was later forfeited, three seats were subsequently held in northwestern Wales. The Conservative Party also made gains at the expense of Labour, several members being returned in constituencies where boundaries were redrawn to take account of the growth of suburban communities. In the 1987 general election, for example, Wales returned 24 Labour members, eight Conservatives, three Plaid Cymru members, and three Liberals.

Plaid Cymru

In 1964 the Labour government honoured a pledge to appoint a secretary of state for Wales with departmental responsibility, and subsequent Labour and Conservative administrations promoted an extensive transfer of functions to the Welsh Office. Demands for an elected assembly with powers devolved from Westminster, similar to those voiced in Scotland, led to the appointment of a Royal Commission on the Constitution that in 1973 recommended devolution for Wales. An act giving Wales a measure of devolution was passed at Westminster in 1978, but at a referendum held in 1979 at the behest of opponents of the measure devolution was rejected by a large majority.

Since 1945 Welsh identity has been reflected in the growth of a plethora of institutions and agencies such as the Welsh Office, the Welsh Arts Council, the Welsh Books Council, and the Welsh Development Agency. These years have seen active endeavours at national and local level to reinvigorate Welsh cultural life and to conserve the literary, artistic, and material heritage. The Welsh Language Society, founded in 1962, brought a measure of militancy to the cause of preserving the language and was among the influences that spurred a more positive response to the problems of its continuing decline, including the Welsh Language Act of 1967. A remarkable growth in Welsh-language education and extensive use of Welsh by the media were important developments. Nonetheless, a new phase of immigration into rural Wales in the 1980s and a sharper awareness of the extreme vulnerability of the economy to the free play of market forces made sustained and cohesive Welsh endeavour at conserving its heritage more imperative than at any other time in the nation's history.

For later developments in the history of Wales, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (J.B.Sm.)

NORTHERN IRELAND

The geographic position of Northern Ireland holds the key to much of its unique social, economic, and political development—a process beset by deep-rooted religious and political antagonisms. Periodically, these antagonisms have flared into civil disorder, most severely in 1921–22, but again from 1968.

The country—created a self-governing state within the United Kingdom by the Government of Ireland Act of 1920—lies in the northeast of the island of Ireland, itself located on that western continental periphery often characterized as Atlantic Europe. Northern Ireland is often referred to as the province of Ulster (and its inhabitants as Ulstermen), though it includes only six of the nine counties which made up that historic Irish entity. It occupies an area of 5,452 square miles (14,120 square kilometres), about a sixth of the whole of the island, and is separated from Scotland, another constituent country of the United Kingdom, by the narrow North Channel, which is at one point only 13 miles (21 kilometres) wide. Historically, this channel has been a link rather than a barrier, and from the earliest times it witnessed a constant coming and going of peoples. This interchange gave the northern part of the island a distinctive regional character, which was confirmed during the Industrial Revolution when the province emerged as a linen-manufacturing region and which later was strengthened when a shipbuilding and engineering industry, based on imported raw materials, made Belfast (now the capital) a major city. The intrusive influences that brought about this cultural and economic transformation were attenuated in the west and south, and the political border with the republic of Ireland is more of a compromise conveniently based on long-existing county borders than a clear-cut regional boundary.

The political separation of Northern Ireland has merely confirmed its economic position as a region existing within the framework of the United Kingdom as a whole, serving as an extension of the larger unit's industrial resources. Although the cultural links of some of the people of Northern Ireland with the republic of Ireland are strong—and much in evidence politically since 1968—it seems as though these ties are likely to suffer increasing erosion as state legislation follows economic trends and progressively caters to educational and social needs along lines similar to those introduced in England, Scotland, and Wales.

Physical and human geography

THE LAND

Relief. Northern Ireland can be thought of topographically as a saucer centred on Lough (lake) Neagh, the upturned rim of which forms the province's highlands. Five of the six former counties—Antrim, Down, Armagh, Tyrone, and Londonderry—met at the lake, and each had a highland region on the saucer's rim. To the north and east the Mountains of Antrim (physiographically a plateau) tilt upward toward the coast. They reach a height of 1,817 feet (554 metres) at Trostan, with the plateau terminating in an impressive cliff coastline of basalts and chalk that is broken by a series of the glaciated valleys known as glens, which face Scotland and are rather isolated from the remainder of the province. The rounded landscape of drumlins—smooth, elongated mounds left by the melting ice of the final glaciation—in the southeast is punctuated by Slieve Croob (1,745 feet [532 metres]) and culminates in the Mourne Mountains, which rise to Slieve Donard (2,789 feet [850 metres]) within two miles of the sea. This impressive landscape of granite peaks is bounded by Carlingford Lough to the south.

Topography

The scenery to the south of Lough Neagh is gentler, but the land rises to 1,886 feet in Slieve Gullion near the border with Ireland. West of Lough Neagh the land rises gently to the more rounded Sperrin Mountains; Sawel

(2,224 feet [678 metres]) is the highest of several hills over 2,000 feet. The far southwest, the former county Fermanagh, is focused geographically on its own lake basin, occupied by Lough Erne, in a drumlin-strewn area ringed by hills more than 1,000 feet high.

Drainage and soils. Much of the Northern Ireland landscape is gentle, and in most low-lying areas it is covered with swarms of drumlins that have played havoc with the local drainage and are interspersed with marshy hollows. Glaciation also gave the land its main valleys, the River Bann—draining Lough Neagh to the Atlantic Ocean in the north—the Blackwater in the southwest, and the River Lagan in the east. All of these valleys have been important routeways but none more important than the Lagan, penetrating from Belfast Lough to the very heart of Ulster.

Soils are varied. Although much glacially transported material covers the areas below 700 feet, the nature of the soil is predominantly influenced by the nature of the underlying parent rock. Brown earth soils, forming arable loams, are extensive and are derived from the ancient Silurian rocks of the southeast and from the basalts of the northeast. There are peaty gleys and podzols in the Sperrins, and the impeded drainage of much of the southwest gives rise to acid brown soil. Peat soils are common, particularly in the hollows lying between the drumlins, and hill peat is widespread over the province. Although it is of no great commercial value, peat traditionally has been a source of fuel for the peasant farmer and is still cut extensively.

Maritime influences

Climate. Northern Ireland's climate is temperate and maritime; most of its weather comes from the southwest in a series of lows bringing the rain and clouds that often lend character to the landscape. Because the country is near the central track of such lows, it often experiences high winds, and in the north and on the east coast, particularly, severe westerly gales are common. Above the 800-foot level, distorted trees and windbreaks testify to the severity of the weather. Rainfall decreases from west to east, although the hills accentuate the amount to some 80 inches (2,000 millimetres) in parts of the west, and there is as little as 32.5 inches (825 millimetres) at Lough Neagh and the extreme southeast. A relatively dry spring gives way to a wet summer and a wetter winter. Daily conditions generally are very changeable, but there are no extremes of heat and cold. The country is exposed to the ameliorating effects of the North Atlantic Current, with little fluctuation in the mean annual temperature. January temperatures vary from 38° F (3.3° C) on the north coast to 35° F (1.7° C) in the east; in July, temperatures of 65° F (18.3° C) are common. In late spring and early summer, the eastern part has slightly lower temperatures, accompanied by a coastal fog called the "haar." These mild and humid climatic conditions have, in sum, made Northern Ireland a green country in all seasons.

Plant and animal life. The general features of the plant life of the province are similar to those in the northwest of Britain. The human imprint is heavy on the landscape and is particularly evident in the absence of trees. Most of the land has been plowed and drained and cultivated for centuries. Above the limit of cultivation, rough pastures have been grazed extensively and beyond them lies a zone of mountain vegetation. Only about 5 percent of the land is now under forest, and most of this has been planted by the state. Young trees in these plantations are economically unimportant, but locally they help to diversify the landscape. Deforestation

The fauna of Northern Ireland is not very different from that of Great Britain; there are, however, fewer species of mammals and birds. Only two mammals—the Irish stoat and the Irish hare—and three species of birds are exclusively Irish. The province is rich in fish, particularly pike, perch, trout, and salmon; the first is the only fish introduced in historic times.

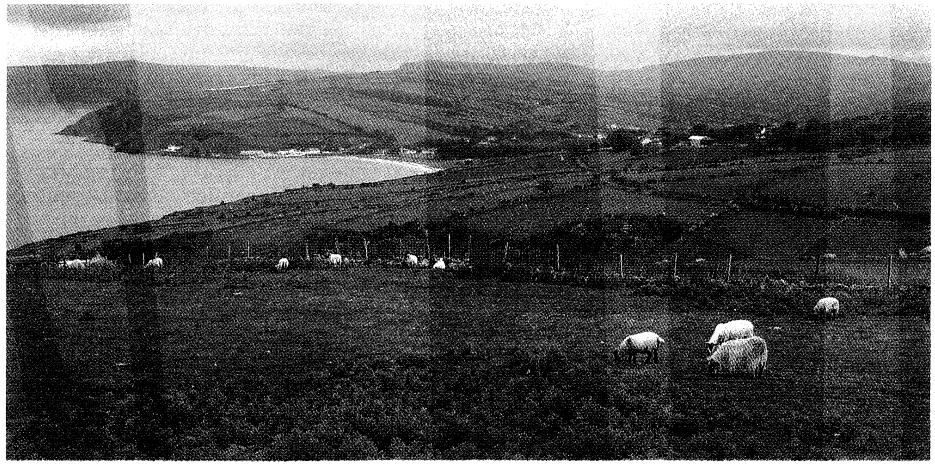
Both plant and animal life benefit from an increasing concern with conservation. There are some 40 nature reserves and several bird sanctuaries controlled by the Ulster Trust for Nature Conservation and by the Department of the Environment.

Settlement patterns. The traditional regions of Northern Ireland correspond closely to the main topographical elements, although they are also the outcome of the cultural evolution of the province. In the north and east the influence of the Scots and English has been paramount. West of Lough Neagh and in the fastness of the Mourne Mountains and of Slieve Gullion or in the more distant Lough Erne region indigenous elements have maintained a distinctiveness that is still apparent. Such relatively isolated pockets as the glens of the northeast coast and Kilkeel, on the southeast coast, have retained a local consciousness that gives colour and interest to the human geography of the province.

The predominant impression of Northern Ireland's landscape is of scattered and isolated farms. Occasional relics of tiny hamlets, or clachans, show that peasant crofts once were huddled together in kinship groups who worked their strips in an open-field system. Between the end of the 18th and the middle of the 19th century most of the land was enclosed and the scattered strips consolidated, partly as a policy of the landlords but finally because of the effect on the population of the potato famines of the 1840s. The end result was the orderly, small square fields that dominate the contemporary landscape. Some landlords rearranged their tenants' land in narrow ribbons, from valley bottom to mountain pasture, giving a characteristic ladder of fields with the farms strung along the road on the valley side. Drumlins also have had an effect on siting; houses are found away from the peaty bottom lands but below the windswept skyline. Most farmhouses are small and a few are still thatched. The occasional larger farm often has a Georgian house—simple and dignified and a reflection of the age of consolidation.

Small market towns rather than villages are frequent. Built by the English and Scots planters or by the landlords of the 18th century, they have a foreign touch of orderli- Market towns

Milt and Joan Mann—CAMERAMANN INTERNATIONAL



Sheep grazing on the Antrim coast, N.Ire.

ness and urbanity. Many are grouped around a "diamond" (meeting place), or square, which is used as a marketplace. Some of them acquired a mill in the 19th century, but in few cases has this changed the essentially rural context.

Few of the market centres have grown into substantial towns. In the western half of the province, regional services and administration have enlarged Omagh and Enniskillen. Some towns have grown with the introduction of industry, particularly Dungannon, specializing in fabrics, and Carrickfergus, noted for rayon. Armagh is an ecclesiastical centre with two cathedrals, while Lisburn, Lurgan, and Portadown, all in the Lagan Valley, form an extension of the Belfast industrial complex, their size based on the textile industry. Bangor is a resort and a residential outlier of Belfast. Londonderry, a centre for shirtmaking, was the centre of the Lough Foyle lowlands until the hinterland that it served was split by the partition of Ireland, but it remains the main focus of the west. The size of Belfast, at the head of Belfast Lough on the northeast coast, underlines its dominance of the region, as well as its significance as an industrial centre and major port. Although its dependence on shipbuilding, linen, and engineering has declined in recent years, shipping remains a major employer and the aircraft industry has gained in importance. The city is also the centre of government, finance, education, and culture. Reflecting the city's 19th-century origin, most of the streets are inextricably and bleakly mixed with mills and factories, while the reclaimed land at the head of Belfast Lough is given over entirely to industry.

THE PEOPLE

The cultural differences that underlie many of Northern Ireland's contemporary social problems have a long and troubled history. The province has had lasting links with parts of western Scotland, strengthened by constant population movements. After the Tudor invasions and particularly after the forced settlements, or plantations, of the early 17th century, the English and Scottish elements were further differentiated from the native Irish by their Protestant faith. Two distinct and often antagonistic elements—the indigenous Roman Catholic Irish and the intrusive Protestant English and Scots—date from that period, and they have played a significant role in molding the province's development. The intrusive element dominated former County Antrim and northern Down, controlled the Lagan corridor toward Armagh, and also formed powerful minorities elsewhere.

This situation contributed to the decline of spoken Gaelic, and it is reflected in the contemporary distribution of religions. Gaelic is now only occasionally heard at a traditional entertainment. The accents given to English, however, are regionally distinctive. The northeastern dialect, dominating the former counties of Antrim and Londonderry and parts of Down, is an offshoot of central Scots dialect and reflects the latter in almost all of its features. The remainder of the province, including the Lagan valley, has accents derived from England, more particularly from Cheshire, Merseyside, Greater Manchester, and southern Lancashire, as well as the "west country" counties of Gloucestershire, Avon, Somerset, and Devon. The towns show more of a mixture and an overlay of standard English.

Religious groups

More than one-fourth of the population is Roman Catholic, and a little less than one-fourth is Presbyterian, with about one-fifth Episcopalian (Church of Ireland); Methodists and members of other sects make up the remainder. The distribution of Roman Catholics and Protestants is, however, very uneven. In country districts, the latter are in a majority in the north and east. Elsewhere, they are in a minority, though fairly highly localized. Most towns have a Protestant majority: this is the case in Belfast, where Roman Catholics compose less than 30 percent of the population. Towns remote from Belfast—Newry and Londonderry—are more than 50 percent Roman Catholic. In the towns there is a high degree of segregation of the sects, and mixing is minimal. Industrial western Belfast is split into two sectors along two axial roads. The Falls Road is as exclusively Roman Catholic as the Shankill Road is Protestant. In many streets adjoining the boundary line, segregation is effectively 100 percent. East Belfast has an exclusively Roman Catholic core, but segregation is less apparent in the middle-class suburbs. Segregation increases as socioeconomic status decreases. Civil disturbances are almost always confined to locations in which segregation is highest. The proportions of the sects are changing slightly because of a differential in birth rate. In Belfast, for example, fertility rates in Roman Catholic districts are much higher than in Protestant areas. There has been a relative decline of Protestants in the 20th century, though they have remained more or less stable in absolute numbers, and a relative and absolute increase of Roman Catholics.

Northern Ireland has the smallest population of the countries of the United Kingdom. The population is fairly static, as migration roughly balances natural increase. The birth rate is much higher than in Great Britain. Consequently Northern Ireland has a young population, with relatively fewer older people. The population is unevenly distributed, sparse over the uplands, heavy in the valleys, and greater in the east than in the west. Since the 1950s, Belfast, like other British cities, has lost population, especially from its centre, and has shared in the decline of mainland cities.

THE ECONOMY

Economically, Northern Ireland is an integral part of the United Kingdom. Its trade is dominated by imports from and exports to other parts of the United Kingdom, though some of the latter are reexported directly from England or are the bases of manufactured goods that are subsequently exported. Trade with the republic of Ireland accounts for a small percentage of the total, and there are lesser links with parts of the Commonwealth and with other European countries.

Northern Ireland is an industrialized country, and the mass of its exports is made up of manufactured goods. This situation, however, has placed a heavy reliance on the import of raw materials, and the Belfast region might well be thought of as an extension of the industrial regions of northwestern England and of the Clyde. Its own mineral resources are extremely meagre: coal must be imported from Great Britain, although local chalk, clays, limestones, and gravels are used to produce lime, bricks, and cement. Northern Ireland's power resources, too, depend on imported coal and oil.

Agriculture and fishing. Northern Ireland possesses considerable agricultural resources, which it has developed as a major part of its economy. Fundamentally, it is still a country of farmers; but in the 20th century this situation has begun to change. The number of farms has decreased, giving a substantial increase in the average size. Consolidation has meant a better livelihood for fewer farmers.

Consolidation of farms

Almost all farms now have electricity, and there has been a great increase in the number of tractors and combine harvesters in use. The frequent rain, the high humidity, and the prospect of wet harvests discourage arable farming, but local conditions produce good grass and rich pasture; nearly all grassland is plowed and there is very little "rough grazing." Mixed farming was traditionally universal, but there has been a considerable movement toward specialization; nearly half the farms concentrate on sheep and beef, and one-fifth specialize in dairying. Principal crops include potatoes, barley, wheat, and oats; turnips are grown to feed livestock. The production of grass seed and seed potatoes for export is also important. To the south of Lough Neagh lies a rich orchard country, and apple growing and market gardening are constant features of the landscape. Most of the agricultural land is held by the occupiers in fee simple, but there persists the peculiar feature of conacre, a system of short (11-month) lets, which still accounts for a portion of the agricultural land. Two-thirds of the farmers are "working owners."

Sea fishing is more or less confined to the northern Irish Sea and is limited to trawlers that operate primarily from the ports of Kilkeel, Ardglass, and Portavogie. Prawns, cod, whiting, and herring are among the main catches. There has been increasing development of marine farming, particularly for oysters. Inland, salmon and eel fishing is

traditional, the latter concentrated where the Bann leaves Lough Neagh.

Industry and services. Although farming dominates the landscape of the country, it provides a livelihood for only about one in 10 of Northern Ireland's people. Real wealth formerly lay in manufacturing—engineering, shipbuilding, the production of linen and clothing, and food and tobacco processing. Linen manufacturing in particular, based on local flax and plentiful soft water, was a feature of many small towns as well as being very important in Belfast. Much of the heavy industry, based on imported raw materials, was concentrated in Belfast, where shipbuilding especially became a mainstay of employment. But however successful these activities were until the earlier part of the 20th century, the industrial structure of Northern Ireland has proved very vulnerable and fallen prey to recessions and to high unemployment. From 1950 to 1980 the government actively encouraged the regeneration of industry and particularly its diversification by, for example, broadening the textile industry by introducing man-made fibres. A series of government acts tried to attract new enterprises into the province with factory space and grants toward capital expenditure. But in relation to the rest of the United Kingdom, Northern Ireland is peripheral and partially isolated, and since 1968 civil disorder has been a great disincentive to incoming firms.

Changes in the economy

Change has continued, however. In common with the rest of the United Kingdom, there has been much restructuring. The number employed in manufacturing industries has fallen dramatically and now accounts for less than one-fifth of the total workforce. The construction industry has also suffered. But there has been an enormous increase in the number of employees in the service industries, now two-thirds of the total. This has been mainly in the public sector, an even greater change than in the United Kingdom as a whole because the province had previously lagged behind in this respect. Between 1960 and 1980 employment in the public sector more than doubled, mainly in education, medicine, and administration. Insurance, finance, and other services have also increased.

Local trade unions are affiliated with the Trades Union Congress, and a vast majority of members belong to English- or Scottish-based unions. Smaller numbers are members of unions based in Northern Ireland or in the Irish republic.

Finance. Northern Ireland is unified fiscally with the United Kingdom. It has three sources of revenue: first, a share of the United Kingdom's revenue from customs and excise, income, value-added, and capital gains taxes, as well as the national insurance surcharge; second, nontax revenue collected locally, such as rates (contributions toward the cost of government services) and land annuities; and third, specific and nonspecific payments from the United Kingdom to bridge the gap between income and expenditure.

Transportation. One of the more noteworthy features of the countryside of Northern Ireland is a close network of well-maintained roads. The majority are very minor, but all parts of the province are accessible. Public-road transport outside the Belfast municipal service has been nationalized since 1935, and since 1968 the Northern Ireland Transport Holding Company (formerly the Ulster Transport Authority) has also controlled the railways, bus companies, and Belfast airport. The railways diminished rapidly—from 824 to some 200 miles—in the economic reorganization following nationalization; a link with Londonderry and the line to Dublin, via Newry, are all that remain. Inland waterways have almost disappeared, although a little commercial traffic still uses the Lower Bann Navigation to Coleraine, and there is some recreational sailing.

The province's links with Britain by sea and air are of paramount importance. Belfast is one of the major ports of the British Isles and has several miles of quays with modern container-handling facilities. Larne and Londonderry are the other ports of significance. Coleraine and Warrenpoint handle some freight, and Larne and Belfast handle passenger transport. All services handle car ferries. The Belfast International Airport, near Aldergrove, has regular air service to major cities in Britain, Europe, and North America.

ADMINISTRATION AND SOCIAL CONDITIONS

Government. Under the Government of Ireland Act (1920) separate parliaments were planned for both northern and southern Ireland, but the plan became effective only for the north, where a Parliament was formally opened in 1921. When the Irish Free State emerged in 1922, Northern Ireland opted to continue its form of government, which was marked by strong ties with the central government of the United Kingdom at Westminster; a commission eventually confirmed a boundary with the south that would contain the six counties. When the Irish Free State seceded from the Commonwealth in 1949, the union of Northern Ireland with the United Kingdom was further confirmed. Paradoxically, its measure of self-government is the result of its determination to preserve union with Britain, and its links with London are safeguarded by the election to Westminster of 17 members. Northern Ireland elects three members to the European Parliament.

Political links with the United Kingdom

Until 1972, the crown exercised all legislative power in Northern Ireland through a governor who held office for six years. The Parliament of Northern Ireland consisted of a Senate and a House of Commons. The former had 26 members, with 24 elected by the Commons for eight years and the lord mayors of Belfast and Londonderry as ex officio members. The House of Commons had 52 members and a maximum life of five years and in all ways was closely modeled on Westminster, which was responsible for matters relating to the crown, war and peace, the armed forces, and foreign powers, as well as trade, navigation, coinage, and many others. Northern Ireland could make no laws interfering with religious equality. In almost all internal matters it followed Westminster policy.

In March 1972 years of political violence, accompanied by increased terror, led the British Parliament to suspend Northern Ireland's Parliament and government and place the country under direct rule from London. Eventually the administration was taken over by the secretary of state for Northern Ireland. Since 1972 a series of attempts has been made to introduce either a power-sharing executive or a new assembly, but they have proved fruitless. The only surviving initiative is the Anglo-Irish Agreement of 1985, which recognizes that the United Kingdom and the republic of Ireland have common problems that may demand a common solution.

Politics since the country's inception has been dominated by the issue of union or separation, and this split has followed religious lines. Protestants have been Unionists, supporting coalition with the United Kingdom, and always had a majority in the Parliament of Northern Ireland. Roman Catholics traditionally supported the former Nationalist Party, advocating reunion with the republic of Ireland, and were always in opposition in Parliament. Electoral geography followed almost exactly a map of the distribution of religious majorities, the east being a stronghold of Unionism and the west of Nationalism, with the concentration of people in and around Belfast assuring an overall Unionist majority. An increasing diversity of political views has produced a plethora of contemporary political parties among Unionists; most support now lies with the Official Unionist and the Democratic Unionist parties. The Social Democratic and Labour Party and Sinn Féin (the political wing of the Provisional Irish Republican Army) are the principal groups supporting unification with Ireland.

Political parties

The former two-tier system of local government—six counties and a county borough, and 24 urban and 26 rural districts—was replaced in 1973 by a single-tier system, paralleling similar changes in the remainder of the United Kingdom. A total of 26 districts were created, each with an elected council. The status of Belfast and Londonderry was maintained in their designation as city councils, and 12 other councils—Antrim, Ards, Ballymena, Ballymoney, Carrickfergus, Castlereagh, Coleraine, Craigavon, Larne, Lisburn, Newtownabbey, and North Down—have borough status. The councils' direct functions are centred

on local services; they have a representative function concerning regional services such as education, housing, and health and social welfare and a consultative function in planning, roads, and conservation.

In most respects the administration of justice reflects the system in the United Kingdom as a whole, being administered by the Crown Court, the High Court, and the Court of Appeal, with final recourse to the House of Lords. Minor offenses are dealt with by a magistrates' court, and others in county courts supervised by a judge and subject to a jury. The exception is that crimes dealing with terrorism are heard by a single judge with no jury.

Civil order is maintained by the Royal Ulster Constabulary (RUC). In 1972 regular troops of the British Army entered the country to quell civil strife and thereafter continued to be deployed in the province. They are assisted by the locally recruited Ulster Defence Regiment.

Education. The social services in Northern Ireland are closely patterned on those of the United Kingdom, although there is sometimes delay in implementing policy, partly because of historic inertia and partly because of the dual nature of the society. Nowhere is this more apparent than in education. The 1947 Education Act parallels the 1944 act of England and Wales, but its implementation is hampered by the traditional tenacious denominational control of education, which the population is generally reluctant to abandon. Northern Ireland has two universities. The Queen's University, established in 1845 as one of three in Ireland, has had a charter since 1908; the University of Ulster was established in 1984 by the merger of the New University of Ulster (at Coleraine) and the Ulster Polytechnic. It has campuses at Coleraine, Jordanstown, Londonderry, and Belfast.

Denominational education

Health and welfare. In health services, as in education, Northern Ireland follows the United Kingdom. The Department of Health and Social Services oversees. The Queen's University has a large medical faculty that supports the health service. Ulster is also known for its export of doctors and nurses.

As in other regions of Britain, one of the major social problems in Northern Ireland is housing. There is a legacy of much poor rural housing in the countryside, exacerbated by the isolation of many of the small dispersed farms, while in Belfast the province suffers from a grim legacy of 19th-century industrial housing.

Social conditions and standards of living in the province have not yet entirely overcome a rural tradition of subsistence farming and low cost of living. Nor has this been offset by conditions among industrial workers, whose employment has felt every depression early and has always been slow to recover. High unemployment and more marked differentials between skilled and unskilled workers' wages have aggravated social problems. Generally speaking, although the standard of incomes and living in Northern Ireland and in England and Wales is roughly comparable, there are still many in Northern Ireland who do not achieve those standards.

CULTURAL LIFE

In the arts and in cultural life generally it is difficult—as is the case with most aspects of life in Northern Ireland—to distinguish between native and imported. Few traces remain of any culture predating the Protestant Ascendancy: the occasional *caley,* or traditional Irish group entertainment, is only the faintest echo of a past that faded with the language. Gone, too, are the strolling players and rhyming weavers. Folk participation and recreation are periodically focused on religious ceremonies and processions—colourful, noisy, and (in recent years) tragically violent demonstrations of sectarian feelings. This fundamental division has also given rise to the nearest approach to folk art, the painting of William of Orange crossing the Boyne, a deeply symbolic representation, adorning the gable end of many a Protestant terrace. In the name of religion, paint, bunting, flags, and arches briefly splash colour over drab gray streets.

In other respects, the cultural milieu of the province is one shared with the remainder of the British Isles and has few distinctive regional characteristics. With the exception of a few country mansions—a legacy of 18th-century English architects—the buildings of Northern Ireland are undistinguished, though this same influence has lent a dignity of proportion and urbanity to some larger farmhouses and to the occasional terrace of houses in the towns. For the most part, however, the industrial growth of the 19th century has completely overshadowed the previous planned phase of growth without contributing any buildings of distinction. It remains to be seen how far larger projects—university buildings and hospitals, for example—have improved the situation.

Northern Ireland has an Arts Council, which successfully encourages all aspects of the arts. The reopening of the Grand Opera House in 1980 has confirmed the return of Belfast's dominance in the performing arts. It has two other theatres, and there is also a touring company based at the University of Ulster at Coleraine. Music is mainly imported, but Belfast has a symphony orchestra and a youth orchestra and has fostered one of the largest festivals in Britain. The council also sponsors art exhibitions. Belfast has a permanent gallery, as do Londonderry and Armagh, though some exhibitions tour the entire province. The 19th century saw little development in the visual arts, but a new interest in landscape emerged briefly at the beginning of the 20th. A provincial school of painting may now be emerging and some recognize a school of poetry. But the province is only slowly shaking off the utilitarian, no-nonsense approach to life that underlay its Victorian growth.

Arts and music

The British Broadcasting Corporation (BBC) has a near-total monopoly in sound broadcasting, but Ulster Television, Ltd., representing the Independent Broadcasting Authority, vies with the BBC. Northern Ireland shares the British press, but several daily newspapers are also published in Belfast.

The Ulster Museum is the national museum. The Ulster Folk and Transport Museum provides a particularly interesting link with peasant origins in Northern Ireland and includes an open-air folk museum. This reflects an appreciation of the past that is also shown by the state's care of ancient monuments, both historic and prehistoric.

Of other cultural institutions perhaps the most notable is the observatory at Armagh. Founded by Archbishop Richard Robinson (Lord Rokeby) in 1790, it has remained an independently governed institution, though it now receives considerable state aid. It has links with observatories in South Africa and also has one of the few astronomy libraries in the British Isles.

The people of Northern Ireland indulge in the same sports as do others in the United Kingdom, association football (soccer) and Rugby Union (amateur) football attracting great popular following; in rugby football the Ulster team contributes to an Irish 15 for international games. Fishing is among the most popular recreations, and the province is known for its coarse fishing, especially bream and roach; with salmon and trout fishing this attracts a European clientele. Many multipurpose recreational centres have been built, the Council of Physical Recreation planning some of the developments. Northern Ireland's hill-walking courses and excellent beaches might attract much greater numbers of overseas visitors were it not for political instability. (E.J.)

History

Out of the 19th- and early 20th- century ferment that produced a sovereign state of Ireland to its south, a separate state of Northern Ireland emerged in 1920–22 as a constituent part of the United Kingdom. Northern Ireland's early history is the history of the traditional Irish province of Ulster, six of whose nine counties Northern Ireland embraces.

EARLY ULSTER

The northernmost of Ireland's provinces has some geographic distinctness. A diagonal line from the northwestern point of Donegal Bay to the southeastern point of Dundalk Bay marks the narrow waist of the island. A belt of hills, lakes, and forests along this line provides a natural border

to the north, discouraging access to or from it. In the early Christian times (the 5th and 6th centuries) this region had a distinctive culture, known under the Celtic name Ulaid (Latin: Ultonia; English: Ulster). Its political centre was at Emain Macha, or Navan Fort, near the modern-day town of Armagh. The most successful Christian missionary in Ireland, the 5th-century Patrick, was predominantly based in the north and associated with its rulers. He established his ecclesiastical centre near Emain Macha, at Armagh, which is still the primatial see of both the Roman Catholic church in Ireland and of the Protestant Church of Ireland.

Mythic history. Ulster is of special importance in the mythic history of Ireland because its rulers and their champions played a prominent role in the rich Irish sagas of the Middle Ages. The Ulster cycle of these tales deals with the exploits of a King Conchobor and the prodigious warriors of the Red Branch, the most celebrated of whom was Cú Chulainn, called the Hound of Ulster. The best-known tale of this cycle is the *Táin Bó Cuailnge* (*The Cattle Raid of Cooley*), which recounts the invasion of Ulster by Queen Medb of Connaught (Connacht, the traditional western province; literally, the "descendants of Conn") in pursuit of a legendary bull. Eventually the men of Connacht are repulsed by the Ulstermen and their spectacular hero, Cú Chulainn.

Táin Bó Cuailnge

The oldest manuscript of the *Táin,* known as the *Book of the Dun Cow,* was compiled in the 12th century and contains language dated to the 8th century. But it is assumed that the story in oral form had existed for at least several centuries previously and that it includes descriptions of practices current in Celtic society in Ireland or Britain or on the Continent as long as several centuries before the birth of Christ. If it is mythic with respect to particular persons and events, the *Táin* is nevertheless an invaluable source for the early history of Irish society.

Gaelic Irish and Anglo-Normans (*c.* 600–*c.* 1300). The postmythic history of Ulster dates from the 7th century, when it begins to be available from Latin documents and chronicles created by churchmen. By this time the 100 or more *tuatha* (clans) of the island had loosely grouped themselves into the five provinces of Ulster (Ulaid), Meath (Midhe, which later dissolved), Leinster (Laigin), Munster (Muma), and Connaught (Connacht). By the 8th century Ulster was dominated by a dynasty called the Uí Néill (O'Neill), which claimed descent from a shadowy figure of the 5th century known as Niall of the Nine Hostages. Divided into a northern and southern branch, the Uí Néill asserted hegemony as high kings, to whom all other Irish kings owed deference. In the early 11th century the king of Munster, Brian Boru, effectively challenged the high kings of the Uí Néill dynasty, thereby ending Ulster's political dominance in early Irish history.

The Uí Néill dynasty

The dominance of Munster was short-lived. In the mid-12th century an incursion of Norman adventurers from England, South Wales, and the Continent greatly complicated the island's political pattern. The Norman beachhead was in Waterford, but from there they struck out both north and west. By 1177 a force of several hundred men under John de Courci, advancing north from Dublin, had established itself in northern County Down and southern Antrim. They built formidable castles at Downpatrick and Carrickfergus and established the northeast coast as the heart of Norman Ulster. De Courci became so threateningly independent that the English king, John Plantagenet, created an earldom of Ulster in 1205 and conferred it upon the more submissive Hugh de Lacy. The title passed to the Norman family of de Burgo, which was joined in the coastal sections of Down and Antrim in the later 13th century by Anglo-Norman families with names such as Mandeville, Savage, Logan, and Bisset. The hinterland of Ulster remained imperviously Gaelic. (For the subsequent fortunes of the Norman colony and the resurgence of Gaelic society in the 14th and 15th centuries, see IRELAND: *History: First centuries of English rule [c. 1166–c. 1600].*)

EARLY MODERN ULSTER

English and Scottish plantations. In the course of the 16th and 17th centuries the most isolated and undisturbed part of Ireland was transformed by immigration from Britain. The narrow North Channel separates northeastern Ulster from southwestern Scotland. Whereas in the early Middle Ages there had been significant eastward migration of Ulstermen to Scotland, in the late 16th century there began a pronounced westward flow of Scots to Ulster. The crucial preconditions of Ulster's transformation were expansion of English ambitions in Ireland from the 1530s, the defeat of Hugh O'Neill and the lords of the north in the opening years of the 17th century, and the determination of James I to "plant" six of Ulster's nine counties with immigrant English and Scottish colonists.

A few years after the defeat of the northern earls an excuse was found to plant the six counties of Ulster, which were judged to have escheated to the Crown. Only Monaghan, Down, and Antrim were excepted, the first because it had been subjected to a "native" plantation in the 1590s, the latter two because neither was held by the rebel earls and both were already areas of extensive de facto Scottish settlement. The plantation formalized and encouraged an immigration that had begun before the 17th century and that continued throughout and after it.

Religion and social structure. Religious differences accentuated the transforming effect of immigration. A half-hearted attempt to propagate Protestantism in Ireland had largely failed by the 1590s among both the Gaelic Irish and the so-called Old English (descendants of the Anglo-Normans). Despite its nominal proscription, the Roman Catholic church claimed the allegiance of almost the entire population, except for the British-born newcomers. English-born settlers gravitated to the Protestant Church of Ireland, modeled on the Church of England. Scottish settlers took with them the ardent Calvinism that had recently established itself in their homeland. Any affinity that Gaelic Irish and Gaelic Scots might once have shared was offset, in an age of doctrinal extremism and intolerance, by the polarities of their respective religions.

Immigration and religious polarities

Ulster became a province dominated by Protestant English and Scottish planters. Its landholding aristocracy was largely English, but beneath it lay a yeomanry of substantial tenant farmers drawn from both Scottish and English immigrants. The native Irish became a landless, displaced population for whom only menial vocations were available. The most violent reaction to this subjection was the rebellion of 1641, which originated in Ulster and took the form of a surprise attack upon English (and later Scottish) settlers. The plantation temporarily collapsed as colonists fled for their lives, but with the reconquest of Ireland by Oliver Cromwell in the 1650s and the restoration of the Stuart monarchy in the 1660s, the Ulster plantation was reestablished.

THE 18TH AND 19TH CENTURIES

Ulster in the 18th century. After the plantation of the 17th century Ulster was distinctive among the provinces of Ireland because its immigrant British (and Protestant) population was larger and more concentrated than that of any other region. When, in 1689, the Roman Catholic James II, expelled from England by the Revolution of the previous year, attempted to recover his fortunes in Ireland, he based his forces in Catholic Dublin, while his adversary, the Protestant William III, made Protestant Belfast his encampment. When James's forces invested the new town of Londonderry, its Protestant inhabitants withstood a long and painful siege rather than capitulate to a Catholic Stuart. Ulster had become the most British and most Protestant part of Ireland, but it contained a large population of non-British Catholics and was contiguous with a larger and decisively Catholic Ireland.

In the late 17th and early 18th century, Ulster, like many predominantly Protestant regions of Europe, became a refuge for Huguenots, Protestants who fled from France after the revocation of the Edict of Nantes in 1685. Many of these refugees brought commercial and industrial skills that contributed to the development of linen-cloth manufacture. Although the linen industry remained traditional and small-scale (and existed in other parts of Ireland as well), it established a foundation for the later industrialization of Belfast and the Lagan Valley in the 19th century.

Huguenot settlers

Eighteenth-century Ulster had two elite and two lower

classes. One of the elites was predominantly "English," contained the most influential landowners, and was affiliated with the Protestant Church of Ireland. The other elite was predominantly "commercial," contained Scots as well as English, and included Protestants affiliated with various sects, especially Calvinistic ones. The two lower classes were divided by religion; one was Catholic, the other Protestant. Among the lower-class Protestants there was substantial emigration to North America in the middle decades of the century. These so-called Scotch-Irish, frustrated by limited economic opportunity in Ulster, became a mainstay of the Middle Atlantic colonies and the Appalachian frontier. The lower-class Protestants who remained in Ulster competed with lower-class Catholics for favourable leases of land and, later, for favourable jobs. Over a period of time the elites gained the allegiance of the lower-class Protestants by playing upon their natural fear and jealousy of lower-class Catholics.

Late 18th-century Ulster exhibited diverse, contrary tendencies. On the one hand, Belfast was the seat of the Society of United Irishmen, whose cultivated members, under the influence of French philosophy, dreamed of an ecumenical Ireland freed of the corrupt Hanoverian monarchy, in which Catholic and Protestant would embrace as compatriots. On the other hand, exacerbated conditions in County Armagh gave rise to bitter sectarian strife, most memorably the pitched battle between Protestant and Catholic factions at the Diamond in September 1795. From that supposed Protestant triumph emerged the Orange Society (later known as the Orange Order), devoted to perpetuating the predominance of Protestants over Catholics that the victory of William III in 1690 (at the Battle of the Boyne and elsewhere) had made possible.

The Orange Order

After the British repression of the rebellion of 1798, which the United Irishmen had encouraged but which had its bloodiest episodes in the south, and the Act of Union of 1800, which put Irish Protestants under the formal protection of the British (now Union) Parliament, the Orange Order was theoretically anachronistic. Protestants no longer needed to band together for their defense. In the 1830s vigorous efforts were made to dissolve the Order, which had spread to England, Scotland, and Irish Protestant communities abroad. The Order became dormant but, feeding off several centuries of remembered conflict, fear, and guilt, it survived.

Home Rule. The Order lay ready to reanimate Protestant Ulster when a succession of 19th-century Irish crises (including the Great Famine of the 1840s) provoked discussion of Irish self-government and the consequent revision of the Union. The first Home Rule bill was introduced in Parliament in 1886 by the Liberal prime minister, William Ewart Gladstone. It was defeated in the House of Commons, but its mere formulation was sufficient to raise the spectre of the political domination of Irish Protestants, located mainly in the north, by Irish Catholics, spread throughout the island. Orangeism revived explosively and was adroitly exploited by the Conservative party, which now made "Unionism"—preservation of the Union of England and Ireland—its foremost concern.

A second Home Rule bill was introduced by the Liberals and was defeated in 1893, in the midst of a long period of Conservative preponderance. When the Liberals finally returned to power in 1905, their victory foretold another effort to establish a measure of self-government for Ireland. In 1912 the third, and final, Home Rule Bill twice passed the House of Commons, but both times it was defeated in the House of Lords. Protestant Ulster, under the leadership of a prominent barrister, Sir Edward Carson, prepared to resist incorporation into a self-governing Ireland. Oaths were sworn (the Solemn League and Covenant), and paramilitary forces were organized and armed. A civil war in Ireland (between Irish Nationalists in the south and Unionists in the north) seemed imminent. In 1914 the Home Rule Bill of 1912 passed the Commons for the third time, which, according to the Parliament Act of 1911, made ratification by the House of Lords unnecessary. When war broke out in Europe, Parliament, however, postponed the operation of the Home Rule Act until after the war, and the Liberal government of H.H. Asquith implied that special provision would be made for Ulster. Thousands of Irish Catholics and Protestants put aside their differences to join the British fighting forces in World War I. But the situation in Ireland was dramatically inflamed by the Easter Rising of 1916 and its harsh repression. The south was being radicalized, and it began to appear that, however offensive the third Home Rule Bill was for Protestant Ulster, it was too late and too little to satisfy Nationalist sentiment in Catholic Ireland.

The third Home Rule Bill

After the war the coalition government of David Lloyd George was obliged to deal with an almost impossible situation in which most of Ireland rejected the Union and most of Ulster rejected everything else. The intended remedy was the Government of Ireland Act of 1920, which created two modestly self-governing units: one comprising six of Ulster's nine counties (later to be known as Northern Ireland); the other comprising the three remaining counties of Ulster together with the 23 counties of the rest of Ireland. Although the Protestant majority of the six counties clearly preferred continuation of the Union for all of Ireland, it settled for Home Rule for itself. Paradoxically, the Catholic majority of the 26 counties, for whom Home Rule had originally been intended, rejected it as inadequate and fought a brief war with Britain before agreeing, through its provisional government, to the Anglo-Irish Treaty of 1921–22. This slightly enlarged the sovereignty of the new Irish state, but it also confirmed the right of the six counties of Northern Ireland to opt out of the arrangement, which they did.

NORTHERN IRELAND SINCE 1922

Precarious coexistence. The constitutional revisions of 1920–22 succeeded in creating a state within Ireland acceptable to the approximately one million Protestant unionists of the six counties. It did not provide a remedy for the several hundred thousand Protestant unionists who lived elsewhere in Ireland, many of whom eventually emigrated. More importantly, it did not provide significant protection for the 500,000 Roman Catholic Nationalists who resided within the six counties. Under the leadership of Sir James Craig (Lord Craigavon), prime minister from 1921 to 1940, Northern Ireland was an unapologetically sectarian state permanently dominated by its Protestant majority and governed in their special interest. Catholics expressed their disdain for the new state by withdrawing from the political arena almost entirely, thereby making even easier Protestant control of local government and the favouring of Protestants in the distribution of jobs, public housing, education, and social services.

Balancing these disadvantages for the Catholic minority was the industrial economy of the north, which had no parallel in the south. By the end of the 19th century Belfast was Ireland's largest city, with a population of nearly 350,000 and with numerous jobs in the textile industries and in shipbuilding. Although skilled jobs were systematically reserved for Protestants, Belfast's economic magnet drew lower-class Catholics from the impoverished countryside. Even if it housed them in appalling ghettoes and inflicted upon them sectarian harassment in the forms of assault, vandalism, discrimination, and occasional riot, Belfast's economic appeal endured even through the Great Depression of the 1930s and the doldrums of the 1960s and '70s.

Employment opportunities for Roman Catholics

Several factors help to explain the relatively minor emigration of Roman Catholics from the north. Not only did they fear that they would be economically worse off in the south, but World War II brought a measure of economic revival, especially in ship and aircraft manufacture. Moreover, the social-welfare provisions extended to Northern Ireland after the war exceeded by far the supports and protections available to individuals in the socially conservative south. Northern Catholics did not "vote with their feet," but neither were they reconciled to the glaring inequities of their state.

Distintegration of stability. By the mid-1960s the fragile stability of Northern Ireland began to erode. Liberal Unionists began to experience some doubts concerning the legitimacy of Protestant domination. Roman Catholics, who were already reentering politics, were impressed by the achievements of American blacks in the civil rights

struggles of that period, and they saw their own situation as analogous. Catholic civil rights protests in 1968 set the scene for violent confrontations that rekindled sectarian conflict between the two communities, especially in Belfast and Derry. The moribund Irish Republican Army (IRA) came back to life in the form of the Provisionals, urban guerrillas who undertook to protect the Catholic population from official and unofficial assault and whose political agenda called for the summary departure of the British armed forces and the end of Protestant domination. The Protestant response was the formation of its own paramilitary brigades.

British forces in Northern Ireland

British forces entered the province in the early 1970s, nominally to keep the peace. Soon, however, they came to be viewed by many Catholics as unwelcome agents of a foreign power. The constitution and parliament of Northern Ireland were suspended in March 1972 by the government of the Conservative British prime minister Edward Heath, and since that time a cabinet official, the minister for Northern Ireland, has been responsible for the province. The British Army has remained a major presence, and elements of martial law have permeated civil and judicial processes in an effort to stem undiminishing violence. The worst year, in terms of deaths, was 1972, when 467 people were killed, 321 of them civilians. The years from 1971 to 1976 represent the most acute period, with an average of 275 deaths per year due to sectarian violence. The figures for the 1980s showed considerable improvement, but 50 to 100 political murders and assassinations still occurred each year. Despite occasional efforts at conciliation and "power sharing" and despite British economic resurgence in the 1980s, which had some benefits for Northern Ireland, the basic social and political dilemmas of the state created in 1920 remained unresolved. (K.S.B.)

For later developments in the history of Northern Ireland, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

BIBLIOGRAPHY

British land and people: A graphic image of the United Kingdom is presented in *The Reader's Digest Complete Atlas of the British Isles* (1965). The evolution of the country's complex and varied landforms is explored in the series "The Geomorphology of the British Isles," ed. by ERIC H. BROWN and KEITH CLAYTON, which began in 1976 and is arranged by region. J. WREFORD WATSON and J.B. SISSONS (eds.), *The British Isles: A Systematic Geography* (1964), is an authoritative text created under the auspices of the 20th International Geographical Congress. Changes in the observable landscapes are detailed in several works by ROY MILLWARD and ADRIAN ROBINSON, published in the series "Landscapes of Britain" in the 1970s. A.G. TANSLEY, *The British Islands and Their Vegetation* (1939, reissued in 2 vol., 1965), is a monumental classic on the ecology of the land. Issues of human geography are introduced in J.W. HOUSE (ed.), *The UK Space: Resources, Environment, and the Future,* 3rd ed. (1982); and R.J. JOHNSTON and J.C. DOORNKAMP (eds.), *The Changing Geography of the United Kingdom* (1982). THOMAS W. FREEMAN, *The Conurbations of Great Britain,* 2nd rev. ed. (1966), analyzes the growth of Britain's cities. A comprehensive historical survey of demography and social conditions is provided in E.A. WRIGLEY and R.S. SCHOFIELD, *The Population History of England, 1541–1871: A Reconstruction* (1981). For information on recent developments, *Britain* (annual), published by the Central Office of Information, is very useful. (W.Ra.)

British economy: Historical studies of economic development and growth include R.C.O. MATTHEWS, C.H. FEINSTEIN, and J.C. ODLING-SMEE, *British Economic Growth, 1856–1973* (1982); PETER BROWNING, *The Treasury and Economic Policy, 1964–1985* (1986); and NICK GARDNER, *Decade of Discontent: The Changing British Economy Since 1973* (1987). General characteristics of the output in all sectors of the economy are studied in M.H. PESTON, *The British Economy: An Elementary Macroeconomic Perspective,* 2nd ed. (1984); and JOHN BLACK, *The Economics of Modern Britain,* 4th ed. (1985). J.C.R. DOW and I.D. SAVILLE, *A Critique of Monetary Policy: Theory and British Experience* (1988); and C.A.E. GOODHART, *Monetary Theory and Practice: The UK Experience* (1984), explore monetary issues. Special studies of economic performance and economic conditions include KEITH SMITH, *The British Economic Crisis: Its Past and Future* (1984); BERNARD ELBAUM and WILLIAM LAZONICK (eds.), *The Decline of the British Economy* (1986); F.V. MEYER (ed.), *Prospects for Recovery in the British Economy* (1985); BEN FINE and LAURENCE HARRIS, *The Peculiarities of the British Economy* (1985); JOHN KAY, COLIN MAYER, and DAVID THOMPSON (eds.), *Privatization and Regulation: The UK Experience* (1986); RUDIGER DORNBUSCH and RICHARD LAYARD (eds.), *The Performance of the British Economy* (1987); and CHRIS HAWKINS and GEORGE MCKENZIE, *The British Economy: What Will Our Children Think?* (1982). WILLIAM KEEGAN, *Mrs. Thatcher's Economic Experiment* (1984); ALAN WALTERS, *Britain's Economic Renaissance: Margaret Thatcher's Reforms, 1979–1984* (1986); and GEOFFREY MAYNARD, *The Economy Under Mrs. Thatcher* (1988), analyze the influence of contemporary political and governmental issues on economic policy. For both historical and the most current economic developments, see ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *OECD Economic Surveys: United Kingdom* (annual); *National Institute Economic Review* (quarterly); and LONDON BUSINESS SCHOOL CENTRE FOR ECONOMIC FORECASTING, *Economic Outlook* (monthly). (U.M.S.)

British administration and social conditions: RICHARD ROSE, *Politics in England: Change and Persistence,* 5th ed. (1989); and ALAN R. BALL, *British Political Parties: The Emergence of a Modern Party System,* 2nd ed. (1987), examine political activities. PHILIP NORTON, *The British Polity* (1984); PETER HENNESSY, *Cabinet* (1986); and MICHAEL RYLE and PETER G. RICHARDS (eds.), *The Commons Under Scrutiny,* 3rd rev. ed. (1988), discuss the constitutional framework. A detailed history of the civil service is offered in PETER HENNESSY, *Whitehall* (1989). The operations of local government are discussed in TONY BYRNE, *Local Government in Britain: Everyone's Guide to How It All Works,* 4th ed. (1986); TONY TRAVERS, *The Politics of Local Government Finance* (1986); and *Paying for Local Government* (1986), a report to Parliament. R.M. JACKSON, *Jackson's Machinery of Justice,* 8th rev. ed., edited by J.R. SPENCER (1989); and ROBERT REINER, *The Politics of the Police* (1985), look at the political aspects of the administration of law and law enforcement. On education, see KEITH EVANS, *The Development and Structure of the English School System* (1985), a study of school management and organization; and ROY LOWE, *Education in the Post-War Years: A Social History* (1988), which explores the change in policies in the mid-20th century that continue to influence the educational system. Contrasting views on the administration of welfare economics are discussed in JULIAN LE GRAND, *The Strategy of Equality: Redistribution and the Social Services* (1982); HOWARD GLENNERSTER, *Paying for Welfare* (1985); and NICHOLAS BARR, *The Economics of the Welfare State* (1987). The policies, politics, and organization of the health care system are studied in CHRISTOPHER HAM, *Health Policy in Britain: The Politics and Organisation of the National Health Service,* 2nd ed. (1985). Policies directed toward income maintenance and redistribution and provision of housing are analyzed in A.B. ATKINSON, *The Economics of Inequality,* 2nd ed. (1983); J.A. KAY and M.A. KING, *The British Tax System,* 4th ed. (1986); A.E. HOLMANS, *Housing Policy in Britain: A History* (1987); and PETER MALPASS and ALAN MURIE, *Housing Policy and Practice,* 2nd ed. (1987). (N.A.Ba.)

British cultural life: Useful information on all aspects of the cultural and social life of the United Kingdom over the centuries is found in ALAN ISAACS and JENNIFER MONK (eds.), *The Cambridge Illustrated Dictionary of British Heritage* (1986), an alphabetically arranged reference work. Historical studies of social and cultural customs include HUGH CUNNINGHAM, *Leisure in the Industrial Revolution: c. 1780–c. 1880* (1980); and SUSAN LASDUN, *Victorians at Home* (1981). Architecture is described in the text accompanying the excellent maps and photographs in *The National Trust Atlas,* 3rd ed. (1987). Other special analyses include ALASTAIR FOWLER, *A History of English Literature* (1987); PETER MILES and MALCOLM SMITH, *Cinema, Literature & Society* (1987); and IAIN CHAMBERS, *Popular Culture: The Metropolitan Experience* (1986), an exploration of the complexity of modern popular culture as influenced by the development and growth of cities. A broad picture of intellectual life is presented in RICHARD HOGGART, *An English Temper: Essays on Education, Culture, and Communications* (1982). (Ed.)

General history: The multivolume "Oxford History of England" series, with the individual works cited in the appropriate chronological sections below, provides a comprehensive survey and excellent bibliographies. More concise overviews include GEORGE MACAULAY TREVELYAN, *History of England,* new illustrated ed. (1973); ARVEL B. ERICKSON and MARTIN J. HAVRAN, *England; Prehistory to the Present* (1968); MAURICE ASHLEY, *Great Britain to 1688: A Modern History* (1961); K.B. SMELLIE, *Great Britain Since 1688: A Modern History* (1962); and CHRISTOPHER HIBBERT, *The English: A Social History, 1066–1945* (1986). For quick reference, see CHRISTOPHER HAIGH (ed.), *The Cambridge Historical Encyclopedia of Great Britain and Ireland* (1985); and E.B. FRYDE *et al.* (eds.), *Handbook of British Chronology,* 3rd ed. (1986).

Ancient Britain: STUART PIGGOTT, *Ancient Europe from the Beginnings of Agriculture to Classical Antiquity* (1965, reprinted 1973), a survey of the pre-Roman period; and TIMOTHY DARVILL, *Prehistoric Britain* (1987), a systematic account of the same five centuries, are both based on modern archaeological research. PETER SALWAY, *Roman Britain* (1981), from the above-mentioned Oxford series, and SHEPPARD FRERE, *Britannia: A History of Roman Britain,* 3rd rev. ed. (1987), provide detailed analyses. A.L.F. RIVET (ed.), *The Roman Villa in Britain* (1969), describes various aspects of the Roman villas of Britain, and the agricultural system and way of life they represent; ANTHONY BIRLEY, *Life in Roman Britain,* new ed. (1981), studies the government, institutions, life, and religions of Roman Britain as they are reflected in archaeological finds and works of the ancient historians; and ERIC BIRLEY, *Roman Britain and the Roman Army* (1953), reprinted 1961), explores the organization of the Roman army through the evidence of inscriptions.

The Anglo-Saxon period: A valuable translation of and commentary on essential records and narrative material preserved in primary sources is provided in DOROTHY WHITELOCK (ed.), *English Historical Documents, c. 500–1042,* 2nd ed. (1979), the first volume of a new edition of the noted multivolume series of sources. Modern general histories include FRANK STENTON, *Anglo-Saxon England,* 3rd ed. (1971, reprinted 1989); PETER HUNTER BLAIR, *An Introduction to Anglo-Saxon England,* 2nd ed. (1977); D.J.V. FISHER, *The Anglo-Saxon Age, c. 400–1042* (1973); DOROTHY WHITELOCK, *The Beginning of English Society,* 2nd ed. (1954, reprinted 1982); and R.I. PAGE, *Life in Anglo-Saxon England* (1970). For coverage of special topics, see J.N.L. MYRES, *The English Settlements* (1986); MARTYN J. WHITTOCK, *The Origins of England, 410–600* (1986); J.M. WALLACE-HADRILL, *Early Germanic Kinship in England and on the Continent* (1971, reprinted 1980); WILLIAM A. CHANEY, *The Cult of Kingship in Anglo-Saxon England: The Transition from Paganism to Christianity* (1970); HENRY MAYR-HARTING, *The Coming of Christianity to Anglo-Saxon England* (1972, reprinted 1977); JOHN GODFREY, *The Church in Anglo-Saxon England* (1962); FRANK BARLOW, *The English Church, 1000–1066: A History of the Later Anglo-Saxon Church,* 2nd ed. (1979); H.R. LOYN, *The Governance of Anglo-Saxon England, 500–1087* (1984), and *The Vikings in Britain* (1977); FRANK BARLOW, *Edward the Confessor* (1970, reprinted 1984); JAMES TAIT, *The Medieval English Borough: Studies on Its Origins and Constitutional History* (1936, reprinted 1968); CHARLES S. ORWIN and CHRISTABEL S. ORWIN, *The Open Fields,* 3rd ed. (1967); CHRISTINE FELL, CECILY CLARK, and ELIZABETH WILLIAMS, *Women in Anglo-Saxon England and the Impact of 1066* (1984); and JAMES CAMPBELL, *Essays in Anglo-Saxon History* (1986).

England from 1066 to 1485: (*Sources*): For translation of a wide range of sources, with commentary, see DAVID C. DOUGLAS and GEORGE W. GREENAWAY (eds.), *English Historical Documents, 1042–1189* (1953); HARRY ROTHWELL (ed.), *English Historical Documents, 1189–1327* (1975); and A.R. MYERS (ed.), *English Historical Documents, 1327–1485* (1969), all from the above-mentioned series. Other good anthologies are R. ALLEN BROWN, *The Norman Conquest* (1984); BERTIE WILKINSON, *The Constitutional History of England, 1216–1399,* 3 vol. (1948–58, reprinted as *The Constitutional History of Medieval England, 1216–1399,* 1965–67), and his *Constitutional History of England in the Fifteenth Century, 1399–1485* (1964). (*General works*): See HELEN M. CAM, *England Before Elizabeth,* 3rd ed. (1967); AUSTIN LANE POOLE, *From Domesday Book to Magna Carta, 1087–1216,* 2nd ed. (1955, reprinted 1975); MAURICE POWICKE, *The Thirteenth Century, 1216–1307,* 2nd ed. (1962); M.T. CLANCHY, *England and Its Rulers, 1066–1272: Foreign Lordship and National Identity* (1983); MAY MCKISACK, *The Fourteenth Century, 1307–1399* (1959, reprinted 1976); ANTHONY TUCK, *Crown and Nobility 1272–1461: Political Conflict in Late Medieval England* (1985); MICHAEL PRESTWICH, *The Three Edwards: War and State in England, 1272–1377* (1980); E.F. JACOB, *The Fifteenth Century, 1399–1485* (1961, reprinted 1978); and M.H. KEEN, *England in the Later Middle Ages: A Political History* (1973). Among studies of individual reigns, see DAVID C. DOUGLAS, *William the Conqueror: The Norman Impact upon England* (1964, reissued 1977); FRANK BARLOW, *William Rufus* (1983); R.H.C. DAVIS, *King Stephen, 1135–1154* (1967, reissued 1977); W.L. WARREN, *Henry II* (1973); MICHAEL PRESTWICH, *Edward I* (1988); G.L. HARRISS (ed.), *Henry V: The Practice of Kingship* (1985); BERTRAM WOLFFE, *Henry VI* (1973); CHARLES ROSS. *Edward IV* (1974, reprinted 1983), and *Richard III* (1981, reprinted 1988). For the history of government and administration, see W.L. WARREN, *The Governance of Norman and Angevin England, 1086–1272* (1987); S.B. CHRIMES, *An Introduction to the Administrative History of Mediaeval England,* 3rd ed. (1966); and T.F. TOUT, *Chapters in the Administrative History of Mediaeval England: The Wardrobe, the Chamber, and the Small Seals,* 6 vol. (1920–33, reprinted 1967), (*Special topics*): Recommended works include R. ALLEN BROWN, *The Normans and the Norman Conquest,* 2nd ed. (1985); V.H. GALBRAITH, *The Making of Domesday Book* (1961), superseded in many ways, but a classic; and PETER SAWYER (ed.), *Domesday Book: A Reassessment* (1985). On feudal society, see FRANK STENTON, *The First Century of English Feudalism, 1066–1166,* 2nd ed. (1961, reprinted 1979); AUSTIN LANE POOLE, *Obligations of Society in the XII and XIII Centuries* (1946, reprinted 1984); and J.C. HOLT, *Magna Carta* (1965), the best account of the Great Charter. See also his *Magna Carta and Medieval Government* (1985). For the nobility, see K.B. MCFARLANE, *The Nobility of Later Medieval England* (1973, reprinted 1980), a most influential book; and CHRIS GIVEN-WILSON, *The English Nobility in the Late Middle Ages: The Fourteenth-Century Political Community* (1987). On Parliament, see G.O. SAYLES, *The King's Parliament of England* (1974); G.L. HARRISS, *King, Parliament, and Public Finance in Medieval England to 1369* (1975); E.B. FRYDE and EDWARD MILLER (ed.), *Historical Studies of the English Parliament,* 2 vol. (1970); and R.G. DAVIES and J.H. DENTON (eds.), *The English Parliament in the Middle Ages* (1981). On the economy of the period, see J.L. BOLTON, *The Medieval English Economy, 1150–1500* (1980); EDWARD MILLER and JOHN HATCHER, *Medieval England: Rural Society and Economic Change, 1086–1348* (1978); M.M. POSTAN, *The Medieval Economy and Society: An Economic History of Britain in the Middle Ages* (1972); REGINALD LENNARD, *Rural England. 1086–1135: A Study of Social and Agrarian Conditions* (1959, reprinted 1966); and JOHN HATCHER, *Plague, Population, and the English Economy, 1348–1530* (1977). England's major trade is discussed in EILEEN POWER, *The Wool Trade in English Medieval History* (1941, reprinted 1987); and T.H. LLOYD, *The English Wool Trade in the Middle Ages* (1977). See also SUSAN REYNOLDS, *An Introduction to the History of English Medieval Towns* (1977, reprinted 1982); and MAURICE BERESFORD, *New Towns of the Middle Ages: Town Plantation in England, Wales and Gascony* (1967, reprinted 1988). For the church, see FRANK BARLOW, *The English Church, 1066–1154: A History of the Anglo-Norman Church* (1979); DAVID KNOWLES, *The Monastic Order in England: A History of Its Development from the Times of St. Dunstan to the Fourth Lateran Council, 940–1216,* 2nd ed. (1963), and *The Religious Orders in England,* 3 vol. (1948–59, reprinted 1979); W.A. PANTIN, *The English Church in the Fourteenth Century* (1955, reprinted 1980); and C.H. LAWRENCE (ed.), *The English Church and the Papacy in the Middle Ages* (1965, reprinted 1984). Studies of the law of the period include FREDERICK POLLOCK and FREDERIC WILLIAM MAITLAND, *The History of English Law Before the Time of Edward I,* 2nd ed., 2 vol. (1898, reissued 1982), still fundamental; DORIS M. STENTON, *English Justice Between the Norman Conquest and the Great Chapter, 1066–1215* (1964); ALAN HARDING, *The Law Courts of Medieval England* (1973); S.F.C. MILSOM, *The Legal Framework of English Feudalism* (1976, reprinted 1986); and JOHN BELLAMY, *Crime and Public Order in England in the Later Middle Ages* (1973).

England in the 16th century: For a good selection of the sources of the period, see C.H. WILLIAMS (ed.), *English Historical Documents, 1485–1558* (1967). The Oxford series volumes provide comprehensive but rather dated surveys: J.D. MACKIE, *The Earlier Tudors, 1485–1558* (1952, reprinted 1978); and J.B. BLACK, *The Reign of Elizabeth, 1558–1603,* 2nd ed. (1959, reprinted 1976). Standard bibliographies for 16th-century England are CONYERS READ (ed.), *Bibliography of British History: Tudor Period, 1485–1603,* 2nd ed. (1959, reissued 1978); and MORTIMER LEVINE, *Tudor England 1485–1603* (1968). Excellent political surveys include G.R. ELTON, *Reform and Reformation: England, 1509–1558* (1977); and WALLACE MACCAFFREY, *The Shaping of the Elizabethan Regime* (1968, reprinted 1971), and *Queen Elizabeth and the Making of Policy, 1572–1588* (1981). The Reformation and ensuing religious crisis are best studied in A.G. DICKENS, *The English Reformation,* rev. ed. (1967); G.R. ELTON, *Policy and Police: The Enforcement of the Reformation in the Age of Thomas Cromwell* (1972, reprinted 1985); PATRICK COLLINSON, *The Elizabethan Puritan Movement* (1967, reprinted 1982), and *The Religion of Protestants: The Church in English Society, 1559–1625* (1982); and CLAIRE CROSS, *Church and People, 1450–1660: The Triumph of the Laity in the English Church* (1976, reprinted 1983). Political theory is brilliantly condensed in CHRISTOPHER MORRIS, *Political Thought in England: Tyndale to Hooker* (1953, reprinted 1980). JOHN NEALE, *Elizabeth I and Her Parliaments,* 2 vol. (1953–57, reissued 1966), and *The Elizabethan House of Commons,* rev. ed. (1976), although under growing attack, constitute the standard authority on the subject. Social, psychological, and cultural history is best sampled in LOUIS B. WRIGHT, *Middle-Class Culture in Elizabethan England* (1935, reissued 1980); RALPH A. HOULBROOKE, *The English Family, 1450–1700* (1984); LAWRENCE STONE, *The Crisis of the Aristocracy, 1558–1641* (1965, reprinted 1979); E.M.W. TILLYARD, *The Elizabethan World Picture* (1943,

reprinted 1973); KEITH WRIGHTSON, *English Society, 1580–1680* (1982); FRITZ CASPARI, *Humanism and the Social Order in Tudor England* (1954, reissued 1968); LACEY BALDWIN SMITH, *Treason in Tudor England: Politics and Paranoia* (1986); KEITH THOMAS, *Religion and the Decline of Magic* (1971, reissued 1978); and MARGO TODD, *Christian Humanism and the Puritan Social Order* (1987). GARRETT MATTINGLY, *The Armada* (1959, reprinted 1987), remains the classic account. Foreign policy is surveyed by R.B. WERNHAM, *Before the Armada: The Emergence of the English Nation, 1485–1588* (1966, reissued 1972), and *The Making of Elizabethan Foreign Policy, 1558–1603* (1980). THOMAS WOODROOFFE, *Vantage at Sea: England's Emergence as an Oceanic Power* (1958), is an excellent treatment of the expansion. For the general European setting, see LACEY BALDWIN SMITH, *The Horizon Book of the Elizabethan World* (1967). L.A. CLARKSON, *The Pre-Industrial Economy in England, 1500–1750* (1971), provides an informative economic survey.

Britain in the 17th century: Documentary sources of the period are collected and commented upon in A. BROWNING (ed.), *English Historical Documents, 1660–1714* (1953). SAMUEL RAWSON GARDINER's magisterial work remains invaluable: his *History of England from the Accession of James I to the Outbreak of the Civil War, 1603–1642,* 10 vol. (1863–82), continued in *History of the Great Civil War, 1642–1649,* 3 vol. (1886–91), and *History of the Commonwealth and Protectorate,* 3 vol. (1894–1901), are available also in many later editions. THOMAS BABINGTON MACAULAY, *The History of England from the Accession of James II,* 5 vol. (1849–61), available in many later editions, some with modern commentary, is another brilliant classic, and should be read for its literary rather than historical value. Other good single-volume treatments are BARRY COWARD, *The Stuart Age: A History of England 1603–1714* (1980); and J.P. KENYON, *Stuart England,* 2nd ed. (1985). A good introduction to the economic history of the period is found in C.G.A. CLAY, *Economic Expansion and Social Change: England 1500–1700,* 2 vol. (1984). J.P. SOMMERVILLE, *Politics and Ideology in England, 1603–1640* (1986), analyzes political theory; and MARK A. KISHLANSKY, *Parliamentary Selection: Social and Political Choice in Early Modern England* (1986), studies politics in practice. The early period of the century is analyzed in CONRAD RUSSELL, *Parliaments and English Politics, 1621–1629* (1979); RICHARD CUST, *The Forced Loan and English Politics, 1626–1628* (1987); and NICHOLAS TYACKE, *Anti-Calvinists: The Rise of English Arminianism, c. 1590–1640* (1987). LAWRENCE STONE, *The Causes of the English Revolution, 1529–1642* (1972, reissued 1986), is a classic introductory account, somewhat superseded by more specialized studies. J.S. MORRILL, *The Revolt of the Provinces: Conservatives and Radicals in the English Civil War, 1630–1650* (1976), is an important statement of the provincialist thesis. DAVID STEVENSON, *The Scottish Revolution, 1637–1644: The Triumph of the Covenanters* (1973), deals with the crises of the Bishop's Wars. C.V. WEDGWOOD, *The Great Rebellion.* 2 vol. (1955–58), remains the most readable account of the civil war. RONALD HUTTON, *The Royalist War Effort 1642–1646* (1981); and MARK A. KISHLANSKY, *The Rise of the New Model Army* (1979), study military politics. An important analysis of the revolution is offered in DAVID UNDERDOWN, *Pride's Purge: Politics in the Puritan Revolution* (1971, reissued 1985). CHRISTOPHER HILL, *The World Turned Upside Down: Radical Ideas During the English Revolution* (1972, reissued 1984), sets out a Marxist interpretation. RONALD HUTTON, *The Restoration: A Political and Religious History of England and Wales, 1658–1667* (1985), examines the period of reestablishment of the monarchy, with all the ensuing confusion. A sustained political history is presented in DAVID OGG, *England in the Reign of Charles II,* 2nd ed., 2 vol. (1955, reissued 1984), and *England in the Reign of James II and William III* (1955, reissued 1984). Important specialized thematic analyses include C.D. CHANDAMAN, *The English Public Revenue, 1660–1688* (1975); JOHN MILLER, *Popery and Politics in England, 1660–1688* (1973); J.R. JONES, *The First Whigs: The Politics of the Exclusion Crisis, 1678–1683* (1961, reprinted 1985); and J.H. PLUMB, *The Growth of Political Stability in England: 1675–1725* (1967, reissued 1980; also published as *The Origins of Political Stability, England, 1675–1725*). J.R. JONES, *The Revolution of 1688 in England* (1972, reprinted 1984), is an excellent treatment of what was once called the Glorious Revolution. The period 1702–14 is masterfully explored in GEOFFREY HOLMES, *British Politics in the Age of Anne,* rev. ed. (1987). (M.A.K.)

Britain, 1714–1815: Documentary sources of the period are gathered in D.B. HORN and MARY RANSOME (eds.), *English Historical Documents, 1714–1783* (1957); and A. ASPINALL and E. ANTHONY SMITH (eds.), *English Historical Documents, 1783–1832* (1959). The Oxford series offers BASIL WILLIAMS, *The Whig Supremacy, 1714–1760,* 2nd rev. ed., edited by C.H. STUART (1962, reprinted 1982); and J. STEVEN WATSON, *The Reign of George III, 1760–1815* (1960, reprinted 1981). The most accessible later introductions to the period include ROY PORTER, *English Society in the Eighteenth Century* (1982); W.A. SPECK, *Stability and Strife: England, 1714–1760* (1977); ERIC J. EVANS, *The Forging of the Modern State: Early Industrial Britain, 1783–1870* (1983); PAUL LANGFORD, *The Eighteenth Century, 1688–1815* (1976); and JOHN CANNON (ed.), *The Whig Ascendancy: Colloquies on Hanoverian England* (1981). Interesting information and strongly opinionated criticism of accepted historiographical and political thought on the period, are found in J.C.D. CLARK, *English Society, 1688–1832: Ideology, Social Structure, and Political Practice During the Ancien Regime* (1985). For surveys of Wales and Scotland in this period, see PHILIP JENKINS, *The Making of a Ruling Class: The Glamorgan Gentry, 1640–1790* (1983), a broader study than its title suggests; DAVID J.V. JONES, *Before Rebecca: Popular Protests in Wales, 1793–1835* (1973); and BRUCE LENMAN, *Integration, Enlightenment, and Industrialization: Scotland 1746–1832* (1981). On the Walpole era, see H.T. DICKINSON, *Walpole and the Whig Supremacy* (1973); JEREMY BLACK (ed.), *Britain in the Age of Walpole* (1984); and, for the Opposition view, LINDA COLLEY, *In Defiance of Oligarchy: The Tory Party, 1714–60* (1982); and E.P. THOMPSON, *Whigs and Hunters: The Origin of the Black Act* (1975, reissued 1985). For the Pelhams, see JOHN B. OWEN, *The Rise of the Pelhams* (1957, reprinted 1971); and F.J. MCLYNN, *The Jacobite Army in England, 1745: The Final Campaign* (1983). Broader developments in British society at this time are explored in JOHN CANNON, *Aristocratic Century: The Peerage of Eighteenth-Century England* (1984); G.A. CRANFIELD, *The Development of the Provincial Newspaper, 1700–1760* (1962, reprinted 1978); P.J. CORFIELD, *The Impact of English Towns, 1700–1800* (1982); and NEIL MCKENDRICK, JOHN BREWER, and J.H. PLUMB, *The Birth of a Consumer Society: The Commercialization of Eighteenth-Century England* (1982). The poor and the obscure are treated in DOUGLAS HAY *et al., Albion's Fatal Tree: Crime and Society in Eighteenth-Century England* (1975); and DOROTHY MARSHALL, *The English Poor in the Eighteenth Century: A Study in Social and Administrative History* (1926, reprinted 1969). Developments after 1754 are studied in RICHARD MIDDLETON, *The Bells of Victory: The Pitt-Newcastle Ministry and the Conduct of the Seven Years' War, 1757–1762* (1985); LEWIS NAMIER, *The Structure of Politics at the Accession of George III,* 2nd ed. (1957, reissued 1973); JOHN BREWER, *Party Ideology and Popular Politics at the Accession of George III* (1976, reprinted 1981); JOHN CANNON, *Parliamentary Reform, 1640–1832* (1972); and ROBERT W. TUCKER and DAVID C. HENDRICKSON, *The Fall of the First British Empire: Origins of the War of American Independence* (1982), a detailed analysis of the imperial policy toward the American colonies, with a rather controversial critique of established historiographical views on the subject. Britain's recovery after the American Revolution is outlined in N.F.R. CRAFTS, *British Economic Growth During the Industrial Revolution* (1985); LINDA COLLEY, "The Apotheosis of George III: Loyalty, Royalty, and the British Nation, 1760–1820," in *Past & Present,* 102:94–129 (February 1984); and E.P. THOMPSON, *The Making of the English Working Class,* new ed. (1968, reissued 1980). Britain's defeat of Napoleon and rise to world dominance is examined in PAUL KENNEDY, *The Rise and Fall of the Great Powers: Economic Change and Military Conflict from 1500 to 2000* (1987); CLIVE EMSLEY, *British Society and the French Wars, 1793–1815* (1979); IAN R. CHRISTIE, *Stress and Stability in Late Eighteenth-Century Britain: Reflection on the British Avoidance of Revolution* (1984); and C.A. BAYLY, *Imperial Meridian: The British Empire and the World, 1780–1830* (1989). (L.J.Co.)

Britain, 1815–1914: For documentary sources of the period, see G.M. YOUNG and W.D. HANDCOCK (eds.), *English Historical Documents, 1833–1874* (1956); and W.D. HANDCOCK (ed.), *English Historical Documents, 1874–1914* (1977). ELIE HALEVY, *A History of the English People in the Nineteenth Century,* 2nd rev. ed., 6 vol. in 7 (1949–52; originally published in French, 5 vol., 1912–32), offers an excellent, massive survey, although there is a mid-century gap in coverage. The Oxford series volumes are LLEWELLYN WOODWARD, *The Age of Reform, 1815–1870,* 2nd ed. (1962, reprinted 1979); and R.C.K. ENSOR, *England, 1870–1914* (1936, reprinted 1985). Other comprehensive histories include R.K. WEBB, *Modern England: From the Eighteenth Century to the Present,* 2nd ed. (1980); G.M. YOUNG, *Portrait of an Age: Victorian England,* new annotated ed. (1977); and ASA BRIGGS, *The Age of Improvement* (1959, reissued 1979). See also the same author's *Victorian People: A Reassessment of Persons and Themes, 1851–67,* rev. ed. (1973), *Victorian Cities,* new ed. (1968), and *Victorian Things* (1988). WALTER E. HOUGHTON, *The Victorian Frame of Mind, 1830–1870* (1957); and SAMUEL HYNES, *The Edwardian Turn of Mind* (1968), observe cultural and intellectual life. GEOFFREY BEST, *Mid-Victorian Britain, 1851–1875,* rev. ed. (1973); and WILLIAM L. BURN, *The Age of Equipoise: A Study of the Mid-Victorian*

Generation (1964), focus on the middle period of the century. Economic conditions are surveyed in W.H.B. COURT, *A Concise Economic History of Britain, from 1750 to recent times* (1954, reprinted 1967). Foreign relations, continental and colonial, and a shift in influence are discussed in R.W. SETON-WATSON, *Britain in Europe, 1789–1914: A Survey of Foreign Policy* (1937, reprinted 1968); C.E. CARRINGTON, *The British Overseas: Exploits of a Nation of Shopkeepers,* 2nd ed. (1968); BERNARD PORTER, *Britain, Europe, and the World, 1850–1986: Delusions of Grandeur,* 2nd ed. (1987); and KEITH ROBBINS, *The Eclipse of a Great Power: Modern Britain, 1870–1975* (1983). The transition from Victorian Britain to the 20th century is examined in DONALD READ, *Edwardian England, 1901–15* (1972).

(As.B.)

Britain since 1914: J.H. BETTEY, *English Historical Documents, 1906–1939* (1967), offers a selection of documentary sources. Detailed exploration of the first half of the 20th century is presented in A.J.P. TAYLOR, *English History, 1914–1945* (1965, reissued 1985), from the Oxford series. Social conditions during World War I and its aftermath are examined in ARTHUR MARWICK, *The Deluge: British Society and the First World War* (1965). Interwar politics are the focus of CHARLES LOCH MOWAT, *Britain Between the Wars, 1918–1940* (1955, reissued 1971); and BENTLEY B. GILBERT, *Britain Since 1918,* 2nd rev. ed. (1980). Social and economic conditions before World War II are observed in ROBERT GRAVES and ALAN HODGE, *The Long Week-End: A Social History of Great Britain, 1918–1939* (1940, reissued 1985); and JOHN STEVENSON, *British Society, 1914–45* (1984). F.S. NORTHEDGE, *The Troubled Giant: Britain Among the Great Powers, 1916–1939* (1966), is a diplomatic history of the period. Political and social aspects of involvement in the war are analyzed in PAUL ADDISON, *The Road to 1945: British Politics and the Second World War* (1975); and ANGUS CALDER, *The People's War: Britain, 1939–1945* (1969, reissued 1986). PAUL KNAPLUND, *Britain: Commonwealth and Empire, 1901–1955* (1956, reprinted 1974), treats the decline of the empire, including the first decade of postwar developments. Other histories reaching into the postwar years are W.N. MEDLICOTT, *British Foreign Policy Since Versailles, 1919–1963,* 2nd rev. ed. (1968); SIDNEY POLLARD, *The Development of the British Economy, 1914–1980,* 3rd ed. (1983); and ALFRED F. HAVIGHURST, *Britain in Transition: The Twentieth Century,* 4th ed. (1985). DAVID BUTLER and GARETH BUTLER, *British Political Facts, 1900–1985,* 6th ed. (1986), is an informative reference source. C.J. BARTLETT, *A History of Postwar Britain, 1945–1974* (1977), is an informative, sustained narrative. Analyses of the postwar governments include KENNETH O. MORGAN, *Labour in Power, 1945–1951* (1984); and PETER HENNESSY and ANTHONY SELDON (eds.), *Ruling Performance: British Governments from Attlee to Thatcher* (1987).

(B.B.G.)

England: Many sources deal with Great Britain or the United Kingdom rather than with England exclusively. Historical surveys of physical geography include W.J. HOSKINS, *The Making of the English Landscape,* new ed. (1988); and H.C. DARBY (ed.), *A New Historical Geography of England* (1973). Physical geography is explored in detail in such illustrated studies as L. DUDLEY STAMP, *Britain's Structure and Scenery,* 6th ed. (1967, reprinted 1984); ARTHUR E. TRUEMAN, *Geology and Scenery in England and Wales,* new ed., rev. by J.B. WHITTOW and J.R. HARDY (1971); and GORDON MANLEY, *Climate and the British Scene* (1952, reprinted 1971). See also PAUL COONES and JOHN PATTEN, *The Penguin Guide to the Landscape of England and Wales* (1986); and J.A. STEERS, *The Sea Coast,* 4th ed. (1969). Rural conditions are outlined in VICTOR BONHAM-CARTER, *The Survival of the English Countryside* (1971); and HOWARD NEWBY, *Green and Pleasant Land?: Social Change in Rural England* (1979, reissued 1985). The impact of industrialization is studied in ROBIN H. BEST and J.T. COPPOCK, *The Changing Use of Land in Britain* (1962); and ROBERT ARVILL, *Man and Environment: Crisis and the Strategy of Choice,* 5th ed. (1983). L. DUDLEY STAMP and STANLEY H. BEAVER, *The British Isles: A Geographic and Economic Survey,* 6th ed. (1971), includes analysis of natural resources. Other works with an economic slant are B.E. COATES and E.M. RAWSTRON, *Regional Variations in Britain: Studies in Economic and Social Geography* (1971); and JIM LEWIS and ALAN TOWNSEND (eds.), *The North South Divide: Regional Change in Britain in the 1980s* (1989). Governmental organization and political developments are the subject of ANTHONY SAMPSON, *The New Anatomy of Britain* (1971); JOHN MOHAN (ed.), *The Political Geography of Contemporary Britain* (1989); HENRY DRUCKER (ed.), *Developments in British Politics 2,* rev. ed. (1988); DAVID BUTLER and DONALD STOKES, *Political Change in Britain: The Evolution of Electoral Choice,* 2nd ed. (1974); and DAVID BUTLER, *The Electoral System in Britain Since 1918,* 2nd ed. (1963, reprinted 1986). A glimpse of English national characteristics is offered in GEORGE ORWELL, *The English People* (1947, reprinted 1974); and a more detailed study in GEOFFREY GORER, *Exploring English Character* (1955). The interconnections of social and cultural history are examined in RAYMOND WILLIAMS, *Culture and Society, 1780–1950* (1958, reprinted 1983); RICHARD HOGGART, *The Uses of Literacy: Changing Patterns in English Mass Culture* (1957, reissued 1984); ROY LEWIS and ANGUS MAUDE, *The English Middle Classes* (1949, reissued 1973); MALCOLM MUGGERIDGE, *The Thirties: 1930–1940 in Great Britain,* new ed. (1967); J.B. ROSE *et al., Colour and Citizenship: A Report on British Race Relations* (1969); and MARTIN J. WIENER, *English Culture and the Decline of the Industrial Spirit, 1850–1980* (1981). The arts and material culture are discussed in NIKOLAUS PEVSNER, *The Englishness of English Art* (1956, reprinted 1976); WILLIAM GAUNT, *A Concise History of English Painting* (1964, reissued as *English Painting: A Concise History,* 1985); NIKOLAUS PEVSNER, *The Buildings of England,* 46 vol. (1951–1974); ALEC CLIFTON-TAYLOR, *The Pattern of English Building,* 4th ed. (1987); and FREDERIC J. OSBORN and ARNOLD WHITTICK, *New Towns: Their Origins, Achievements, and Progress,* 3rd ed. (1977).

(W.H.Th./Ed.)

Scotland: A good modern review of the geography of Scotland is CHALMERS M. CLAPPERTON (ed.), *Scotland, a New Study* (1983). W.H. MURRAY, *The Hebrides* (1966), and *The Companion Guide to the West Highlands of Scotland: The Seaboard from Kintyre to Cape Wrath,* 7th ed. (1977, reissued 1984), are informative regional accounts. F. FRAZER DARLING and J. MORTON BOYD, *The Highlands and Islands,* 2nd ed. (1969); J.B. WHITTOW, *Geology and Scenery in Scotland* (1977); and G.Y. CRAIG (ed.), *Geology of Scotland,* 2nd ed. (1983), are written with impressive authority. For the economy, GAVIN MCCRONE, *Scotland's Economic Progress, 1951–1960* (1965), is a useful account of the 1950s; as is for later years KEITH P.D. INGHAM and JAMES LOVE (eds.), *Understanding the Scottish Economy* (1983). W.H. MARWICK, *A Short History of Labour in Scotland* (1967), analyzes trade unions. *Scottish Abstract of Statistics* (annual); and *Scottish Economic Bulletin* (semiannual), are mines of information. Political characteristics are surveyed in JAMES G. KELLAS, *Modern Scotland,* 2nd ed. (1980), and *The Scottish Political System,* 4th ed. (1989). The self-government debate has inspired a prolific literature, including JACK BRAND, *The National Movement in Scotland* (1978); H.J. HANHAM, *Scottish Nationalism* (1969); and CHRISTOPHER HARVIE, *Scotland and Nationalism: Scottish Society and Politics, 1707–1977* (1977). Governmental issues are surveyed in *The Scottish Government Yearbook;* and JOHN S. GIBSON, *The Thistle and the Crown: A History of the Scottish Office* (1985). THOMAS MACKAY COOPER (LORD COOPER OF CULROSS), *The Scottish Legal Tradition,* 4th ed., rev. by MICHAEL C. MESTON (1977); and SCOTTISH LAW COMMISSION, *The Legal System of Scotland,* 3rd ed. (1981), are brief introductions to Scots law. Trends in health care are outlined in *Health in Scotland* (annual), and *Scottish Health Statistics* (annual). The national cultural heritage is explored in diverse works, including DAVID DAICHES (ed.), *A Companion to Scottish Culture* (1981), a comprehensive reference source; JOHN TELFER DUNBAR, *History of Highland Dress,* 2nd ed. (1979); ROBERT BAIN, *The Clans and Tartans of Scotland,* 5th ed. (1976); FRANCIS COLLINSON, *The Traditional and National Music of Scotland* (1966); J.F. FLETT and T.M. FLETT, *Traditional Dancing in Scotland* (1964, reprinted 1985); JENNY CARTER and JANET RAE, *Chambers Guide to Traditional Crafts of Scotland* (1988); JOHN G. DUNBAR, *The Architecture of Scotland,* 2nd rev. ed. (1978); JAMES L. CAW, *Scottish Painting, Past and Present, 1620–1908* (1908, reprinted 1975); WILLIAM HARDIE, *Scottish Painting, 1837–1939* (1976); EDWARD GAGE, *The Eye in the Wind: Scottish Painting Since 1945* (1977); RODERICK WATSON, *The Literature of Scotland* (1984); CAIRNS CRAIG (ed.), *The History of Scottish Literature,* 4 vol. (1987–88); TREVOR ROYLE, *The Macmillan Companion to Scottish Literature* (1983); and DAVID HUTCHISON, *The Modern Scottish Theatre* (1977).

(I.C.M.)

General histories include ROSALIND MITCHISON, *A History of Scotland,* 2nd ed. (1982); MICHAEL JENNER, *Scotland Through the Ages* (1987); and GORDON DONALDSON, *Scotland: The Shaping of a Nation,* 2nd rev. ed. (1980). Early periods are studied in W. CROFT DICKINSON, *Scotland from the Earliest Times to 1603,* 3rd rev. ed., edited by ARCHIBALD A.M. DUNCAN (1977); and DERICK S. THOMSON, *The Companion to Gaelic Scotland* (1983). The following are more specialized histories: R.A. HOUSTON and I.D. WHYTE (eds.), *Scottish Society, 1500–1800* (1988); T.C. SMOUT, *A History of the Scottish People, 1560–1830* (1969, reprinted 1985), and *A Century of the Scottish People, 1830–1950* (1986); R.H. CAMPBELL, *The Rise and Fall of Scottish Industry, 1707–1939* (1980); and I.G.C. HUTCHISON, *A Political History of Scotland, 1832–1924: Parties, Elections, and Issues* (1986).

(J.M.S./Ed.)

Wales: H. CARTER and H.M. GRIFFITHS (eds.), *National Atlas of Wales* (1980), provides wide-ranging coverage, with explanatory text in both English and Welsh. E.G. BOWEN, *Wales: A Physical, Historical, and Regional Geography* (1957), is a

comprehensive volume, now somewhat out of date but still valuable; DAVID THOMAS (ed.), *Wales: A New Study* (1977), offers a similar approach in a more recent work; and ERIC H. BROWN, *The Relief and Drainage of Wales: A Study in Geomorphological Development* (1960), examines the physiographical evolution of the Welsh landscape. F.V. EMERY, *Wales* (1969); MOELWYN WILLIAMS, *The Making of the South Wales Landscape* (1975); and D. HUW OWEN (ed.), *Settlement and Society in Wales* (1988), are more general treatments. H. CARTER, *The Towns of Wales: A Study in Urban Geography* (1965), discusses the growth, functions, and morphology of urban areas; ELWYN DAVIES and ALWYN D. REES (eds.), *Welsh Rural Communities,* 2nd ed. (1962), examines four rural communities; and IAN HUME and W.T.R. PRYCE (eds.), *The Welsh and Their Country* (1986), treats, in a series of essays, social topics. On population, see PETER S. HARPER and ERIC SUNDERLAND (eds.), *Genetic and Population Studies in Wales* (1986). Several special issues of *Cambria: A Welsh Geographical Review* (semiannual), itself an informative source, comprise the collection *Human Geography from Wales,* ed. by WAYNE K.D. DAVIES (1986). Language issues are analyzed in J.W. AITCHISON and H. CARTER, *The Welsh Language, 1961–1981: An Interpretative Atlas* (1985); MEIC STEPHENS (ed.), *The Welsh Language Today* (1973); and COLIN BAKER, *Aspects of Bilingualism in Wales* (1985). For the economy, see GRAHAM L. REES (ed.), *Survey of the Welsh Economy* (1973); and *Digest of Welsh Statistics* (annual). Analysis of government is found in J.A. ANDREWS (ed.), *Welsh Studies in Public Law* (1970).

(H.Ca.)

General historical surveys include GWYN A. WILLIAMS, *When Was Wales?: A History of the Welsh* (1985); R. BRINLEY JONES (ed.), *Anatomy of Wales* (1972); and PRYS MORGAN and DAVID THOMAS, *Wales: The Shaping of a Nation* (1984). The early period is examined in COLIN RENFREW, *Archaeology and Language: The Puzzle of Indo-European Origins* (1987), reflecting recent scholarly thought; while T.G.E. POWELL, *The Celts,* new ed. (1980); NORA K. CHADWICK, *The Celts* (1970); and MÝLES DILLON and NORA K. CHADWICK, *The Celtic Realms,* 2nd ed. (1972), represent more traditional views. Important studies of Roman and post-Roman Wales include V.E. NASH-WILLIAMS, *The Roman Frontier in Wales,* 2nd ed., rev. by MICHAEL G. JARRETT (1969), and *The Early Christian Monuments of Wales* (1950); E.G. BOWEN, *Saints, Seaways, and Settlements in the Celtic Lands* (1969, reprinted with revisions, 1977); and LESLIE ALCOCK, *Arthur's Britain: History and Archaeology, AD 367–634* (1971). For the Middle Ages, JOHN EDWARD LLOYD, *A History of Wales from the Earliest Times to the Edwardian Conquest,* 2 vol. (1911), available also in many later editions, is a classic work still not superseded. Later studies include WENDY DAVIES, *Wales in the Early Middle Ages* (1982); R.R. DAVIES, *Conquest, Coexistence, and Change: Wales, 1063–1415* (1987), and *Lordship and Society in the March of Wales, 1282–1400* (1978); and GLANMOR WILLIAMS, *The Welsh Church from Conquest to Reformation,* rev. ed. (1976). The key works for modern Welsh history are GLANMOR WILLIAMS, *Recovery, Reorientation, and Reformation: Wales, c. 1415–1642* (1987); GERAINT H. JENKINS, *The Foundations of Modern Wales: Wales 1642–1780* (1987); and KENNETH O. MORGAN, *Rebirth of a Nation: Wales, 1880–1980* (1981). Valuable specialist histories include GERAINT H. JENKINS, *Literature, Religion, and Society in Wales, 1660–1730* (1978); DAVID WILLIAMS, *The Rebecca Riots: A Study in Agrarian Discontent* (1955, reprinted 1986); IEUAN GWYNEDD JONES, *Explorations and Explanations: Essays in the Social History of Victorian Wales* (1981), and *Communities: Essays in the Social History of Victorian Wales* (1987); KENNETH O. MORGAN, *Wales in British Politics, 1868–1922,* 3rd ed. (1980); and PETER SMITH, *Houses of the Welsh Countryside: A Study in Historical Geography,* 2nd enlarged ed. (1988). For material evidence in general the "Inventories" of the Royal Commission on Ancient and Historical Monuments in Wales are invaluable. For historical information on people, JOHN EDWARD LLOYD and R.T. JENKINS (eds.), *The Dictionary of Welsh Biography down to 1940* (1959; originally published in Welsh, 1953); and MEIC STEPHENS (ed.), *The Oxford Companion to the Literature of Wales* (1986), are useful reference works.

(J.B.Sm.)

Northern Ireland: The geomorphological background of the island of Ireland as a whole is found in G.L. HERRIES DAVIES and NICHOLAS STEPHENS, *Ireland* (1978); and A.R. ORME, *Ireland* (1970). Treatment of Northern Ireland in particular is found in R. COMMON (ed.), *Northern Ireland from the Air* (1964), a photographic atlas; and R.H. BUCHANAN and B.M. WALKER (eds.), *Province, City, & People: Belfast and Its Region* (1987). EMRYS JONES, *A Social Geography of Belfast* (1960), discusses the historical and environmental background. The character of another part of Northern Ireland is admirably portrayed in E. ESTYN EVANS, *Mourne Country: Landscape and Life in South Down,* 2nd ed. (1967). PAUL A. COMPTON, *Northern Ireland: A Census Atlas* (1978), provides demographic statistics; see also the author's *Demographic Trends in Northern Ireland* (1986). An introduction to the economy is provided by THOMAS WILSON, *Economic Development in Northern Ireland* (1965). The background to the cultural and political history of Ulster is studied in M.W. HESLINGA, *The Irish Border as a Cultural Divide: A Contribution to the Study of Regionalism in the British Isles,* 3rd ed. (1979). DENIS P. BARRITT and CHARLES F. CARTER, *The Northern Ireland Problem: A Study in Group Relations,* 2nd ed. (1972), deals with the social and political prelude to the current situation. The contemporary political situation is treated in RICHARD ROSE, *Northern Ireland: A Time of Choice* (1976), and JOHN OLIVER, *Ulster Today and Tomorrow* (1978). Useful material is found in the *Ulster Year Book* (annual), as well as in *Social and Economic Trends in Northern Ireland* (annual).

(E.J.)

For general historical surveys, see PATRICK BUCKLAND, *A History of Northern Ireland* (1981); CONSTANTINE FITZGIBBON, *Red Hand: The Ulster Colony* (1971); A.T.Q. STEWART, *The Narrow Ground: Aspects of Ulster, 1609–1969* (1977), and *The Ulster Crisis* (1967); T.W. MOODY, *The Ulster Question, 1603–1973,* 4th ed. (1980); and SEAN CRONIN, *Irish Nationalism: A History of its Roots and Ideology* (1980). The following studies of specific historical events and periods can be recommended: M. PERCEVAL-MAXWELL, *The Scottish Migration to Ulster in the Reign of James I* (1973); PHILIP S. ROBINSON, *The Plantation of Ulster: British Settlement in an Irish Landscape, 1600–1670* (1984); RAYMOND GILLESPIE, *Colonial Ulster: The Settlement of East Ulster, 1600–1641* (1985); PATRICK MACRORY, *The Siege of Derry* (1980); CONRAD GILL, *The Rise of the Irish Linen Industry* (1925, reprinted 1964); J.C. BECKETT and R.E. GLASSCOCK (eds.), *Belfast: The Origin and Growth of an Industrial City* (1967); HEREWARD SENIOR, *Orangeism in Ireland and Britain, 1795–1836* (1966); W.A. MAGUIRE, *The Downshire Estates in Ireland, 1801–1845: The Management of Irish Landed Estates in the Early Nineteenth Century* (1972); DAVID W. MILLER, *Queen's Rebels: Ulster Loyalism in Historical Perspective* (1978); and PAUL ARTHUR, *Government and Politics of Northern Ireland,* 2nd ed. (1984). Extensive discussion of the problems of the 20th century include R.J. LAWRENCE, *The Government of Northern Ireland: Public Finance and Public Services, 1921–1964* (1965); PAUL ARTHUR and KEITH JEFFERY, *Northern Ireland Since 1968* (1988); JOHN BOWMAN, *De Valera and the Ulster Question, 1917–1973* (1982); and PADRAIG O'MALLEY, *The Uncivil Wars: Ireland Today* (1983).

(K.S.B.)

United Nations

The name United Nations was used during World War II to denote the nations allied against Germany, Italy, and Japan and later was adopted as the name of the postwar world organization. The Declaration of the United Nations, signed by 26 states on Jan. 1, 1942, set forth the war aims of the Allied powers.

The first major step toward the formulation of a permanent organization was taken at the Dumbarton Oaks Conference, a meeting of diplomatic experts of the Big Four powers (the United States, the United Kingdom, the Soviet Union, and China) held Aug. 21–Oct. 7, 1944, at Dumbarton Oaks, an estate in Washington, D.C. Though substantial progress was made, there was disagreement on two essential issues: the voting system of the proposed Security Council, which later became famous as the "veto problem," and membership, because the Soviet Union demanded seats in the General Assembly for all of its constituent republics. Roosevelt, Churchill, and Stalin finally resolved these two issues at the Yalta Conference and also agreed that the new agency would include a trusteeship system to succeed the League of Nations mandate system.

The Dumbarton Oaks proposals, as modified at the Yalta Conference, thus formed the basis of negotiations at the United Nations Conference on International Organization (UNCIO), which convened at San Francisco on April 25, 1945, and drafted the Charter of the United Nations (UN). The Charter was signed on June 26 and entered into force on Oct. 24, 1945.

The San Francisco Conference was attended by representatives of the 26 states that had signed the Declaration of the United Nations and the 20 others that had declared war against the Axis powers by March 1945. Four other states (the Ukrainian S.S.R., the Belorussian S.S.R., Argentina, and Denmark) were admitted during the conference. Poland, not present at the conference, was permitted to become an original member of the UN.

The San Francisco Conference was the first major international conference that was not dominated by Europe; only nine continental European states west of the Soviet Union were represented. The 21 American republics, seven Middle Eastern states, six Commonwealth nations, three Soviet republics, two East Asian nations, and two African states that convened with the European states at San Francisco represented all geographical areas of the world.

The international Secretariat provided interpreters and translators and distributed documents and speeches daily in the five official languages (English, French, Spanish, Russian, and Chinese). The chairmanship of the plenary sessions rotated among the Big Four. The private consultations of the Big Four, to which France was later added, exerted great influence. The rule of unanimity, usually adopted by political conferences, was abandoned. Measures under consideration could be carried in committees, commissions, and plenary sessions by a two-thirds vote.

Political issues arose, especially between the Western powers and the Soviet Union, over the admission of the Ukrainian S.S.R., the Belorussian S.S.R., and Argentina; the recognition of a government of Poland; and the extension of the great-power veto in the Security Council to discussion as well as to recommendations and decisions. These controversies were settled by compromises. The issues concerning domestic jurisdiction versus international competence for the protection of human rights and the promotion of economic and social welfare, the status of colonial areas and of regional and defense arrangements, and great-power dominance versus the equality of states involved other groupings. In these matters the small states and the Asian and Latin-American states succeeded in modifying the Dumbarton Oaks proposals.

(Q.W./E.Lu./Ed.)

For coverage of related topics in the *Macropædia* and *Micropædia,* see the *Propædia,* sections 544 and 971, and the *Index.*

This article is divided into the following sections:

Organization of the United Nations 140
- Purposes and membership 140
- Principal organs 141
  - General Assembly
  - Security Council
  - Economic and Social Council
  - Trusteeship Council
  - International Court of Justice
  - Secretariat

Functions of the United Nations 142
- Maintenance of international peace and security 142
  - Settlement of disputes
  - Regulation of armaments
  - Revision of the Charter
- Development of international law 143
- Economic and social cooperation 143
  - Economic reconstruction
  - Technical assistance
  - Financing economic development
  - Trade and development
  - Refugees
  - Human rights
  - Control of narcotics
  - Specialized agencies
- Dependent areas 147
  - Non-self-governing territories
  - Trusteeship system

Administration 148
- Finances 148
- Secretariat 148
- Privileges and immunities 148
- Headquarters 148

Bibliography 148

Organization of the United Nations

PURPOSES AND MEMBERSHIP

The first article of the Charter outlines the purposes of the organization, declaring that the primary objective of the United Nations is the maintenance of international peace and security. The organization is also dedicated to the development of friendly relations among nations, based on the principle of equal rights and self-determination of peoples; to the achievement of international cooperation in solving international economic, social, cultural, or humanitarian problems; and to serving as a centre for harmonizing the actions of nations in the attainment of these common ends. Some of the basic principles of the United Nations, as outlined in Article 2 of the Charter, are the following: the United Nations is based on the sovereign equality of its members; disputes are to be settled by peaceful means; members undertake not to use force or the threat of force in contravention of the purposes of the United Nations; each member must assist the organization in any action it takes under the Charter; and states that are not members of the United Nations are required to act in accordance with these principles insofar as necessary for the maintenance of international peace and security. Article 2 also stipulates that, except to take enforcement measures, the organization shall not intervene in matters within the domestic jurisdiction of any state.

New members are admitted to the United Nations on

recommendation of the Security Council and by a two-thirds vote of the General Assembly. The requirement of concurrence of the permanent members of the Security Council was for 10 years a serious obstacle to the admission of new members. By 1950 only nine of 31 applicants had been admitted. A number of efforts were made by the General Assembly to break the deadlock. In the 10th assembly session (1955) there was wide support for a package deal sponsored by 29 members under the leadership of Canada. The proposal was approved by the Security Council in modified form, and the General Assembly voted to admit 16 new members. Thereafter, new members were admitted almost yearly.

The People's Republic of China

The question of China's representation in the UN was before the General Assembly at every session from 1950, when India introduced a resolution declaring that the Communist regime was entitled to representation in the assembly. This resolution failed, largely because of the opposition led by the United States. For 20 years efforts to bring the People's Republic of China into the organization and to expel the representatives of Nationalist China (Taiwan) were frustrated. Finally, in 1971, the United States acquiesced to the idea of admitting Communist China to the Security Council and General Assembly if Nationalist China could retain a seat in the General Assembly. This compromise failed. The General Assembly on Oct. 25, 1971, voted to admit the representatives of the People's Republic and to remove the Nationalist representatives. Nationalist China's permanent Security Council seat was given to Communist China.

PRINCIPAL ORGANS

The United Nations has six principal organs: the General Assembly, Security Council, Economic and Social Council, Trusteeship Council, International Court of Justice, and Secretariat.

General Assembly. The General Assembly is the only body in which all of the UN members are represented. A member may send as many as five representatives, but each member has only one vote. Decisions on substantive questions are taken by a majority or by a two-thirds vote, depending on the importance of the matters involved. Procedural questions are decided by majority vote. Through its deliberative, supervisory, financial, and elective functions, the General Assembly occupies a central position in the functioning of the United Nations.

The General Assembly exercises control over the activities of UN organs in the economic and social fields and in dealing with the problems of non-self-governing territories. It receives annual reports from the secretary general, Security Council, Economic and Social Council, and Trusteeship Council and may make recommendations to these organs. The assembly approves trust agreements and through the Trusteeship Council supervises the administration of trust territories. It exercises general supervision and control over the operations of the Secretariat. The financial power of the assembly is exercised through its control over the budget of the United Nations and the scale of assessments levied on members.

The General Assembly chooses all the members of the Economic and Social Council and the elective members of the Security Council and Trusteeship Council. Along with the Security Council, it participates in the election of judges of the International Court of Justice and the appointment of the secretary general. The assembly shares with the Security Council the power to propose amendments to the Charter, as well as the right to convene a conference to revise the Charter.

Evolution of the General Assembly

During its first 10 years the General Assembly increased in importance as an organ of deliberation and political influence. At the same time, the influence and the effectiveness of the Security Council declined because of the inability of the permanent members to cooperate. The broad language of Articles 10, 11, and 14 of the Charter permitted the role of the General Assembly to increase far beyond what was originally envisaged, and most of its members encouraged this increase. Members became more concerned with ways to strengthen the assembly in both organization and procedure.

The assembly convenes annually, but its rules also permit the calling of special sessions on short notice. It works through a complex structure of main committees, procedural committees, standing committees, and subsidiary and ad hoc bodies.

Security Council. The Charter assigns the Security Council primary responsibility for the maintenance of international peace and security. The council consisted originally of 11 members—five permanent members (the Republic of China, France, the United Kingdom, the Soviet Union, and the United States) and six nonpermanent members elected by the General Assembly for two-year terms. From the beginning, nonpermanent members of the Security Council were elected with a view to giving representation to certain regions or groups of states. This practice ran into difficulty as the number of UN members increased. In 1965, when an amendment to the Charter was ratified, the council became a 15-member body consisting of the original five permanent members and 10 nonpermanent members. The latter were to be chosen as follows: five from African and Asian states, one from eastern European states, two from Latin-American states, and two from western European and other states. The presidency is held by each member in turn for one month.

Under the Charter members of the United Nations agree to carry out the decisions of the Security Council. This agreement refers primarily to decisions pertaining to the maintenance of international peace and security. The council may investigate any dispute that might threaten international peace and security, but it can make recommendations only for its peaceful settlement. The council may, however, require members of the UN to apply sanctions against any state found guilty of a threat to the peace, breach of the peace, or act of aggression, or of failing to perform the obligations incumbent upon it under a judgment of the International Court of Justice. Recommendations for the regulation of national armaments may also be made by the council.

On procedural matters decisions by the Security Council are made by an affirmative vote of any nine (seven until 1965) of its members. On substantive matters, including the investigation of a dispute and the application of sanctions, nine (seven until 1965) affirmative votes (including those of the five permanent members) are required, but in practice a permanent member may abstain without impairing the validity of the decision. A vote on whether a matter is procedural or substantive is itself a substantive question. Because the council is required to function continuously, each member is represented at all times at UN headquarters.

Economic and Social Council. The Economic and Social Council directs and coordinates the economic, social, humanitarian, and cultural activities of the United Nations. Its 54 members (originally 18) are elected by the General Assembly for three-year terms with the possibility of reelection. The council meets at least twice a year.

Commissions

The council is assisted by commissions organized on functional or geographic bases. The functional commissions, including an economic, a social, and a human rights commission, carry out studies and otherwise assist the council in the performance of its duties. They formulate resolutions, recommendations, and international conventions on which the council and General Assembly take action. Four regional commissions—for Europe, Asia and the Far East, Latin America, and Africa—that were established to deal with regional economic problems have assumed roles of considerable importance both as advisory organs and as organs with operational responsibilities.

Trusteeship Council. The Trusteeship Council was established to supervise the administration of trust territories by the administering states. It is composed of UN members who administer trust territories, the permanent members of the Security Council not administering trust territories, and as many other nonadministering members elected by the General Assembly for three-year terms as are necessary to ensure an equal number of administering and nonadministering members in total membership. The Trusteeship Council is authorized to send inspection missions into the trust territories, to receive and examine

petitions, to consider reports submitted by the administering authorities, and to make recommendations.

World Court

International Court of Justice. The International Court of Justice, popularly known as the World Court, is the principal judicial organ of the United Nations. The 15 judges of the court are elected by the General Assembly and the Security Council voting independently. No two judges may be nationals of the same state. The main forms of civilization and the major legal systems of the world are to be represented. Judges serve for nine years and are eligible for reelection. The seat of the World Court is The Hague.

The jurisdiction of the World Court comprises "all cases which parties refer to it and all matters specially provided for in the Charter of the United Nations or in treaties and conventions in force." By formal declaration states may accept the compulsory jurisdiction of the World Court in specified categories of disputes. The World Court may give advisory opinions at the request of the General Assembly or the Security Council or at the request of other organs and specialized agencies authorized by the General Assembly.

Secretariat. The Secretariat is headed by the secretary general, who is appointed by the General Assembly upon the recommendation of the Security Council. In addition to being the chief administrative officer of the United Nations, the secretary general also has important political functions, being specifically charged with bringing before the organization any matter that threatens international peace and security (see below *Administration*).

Functions of the United Nations

MAINTENANCE OF INTERNATIONAL PEACE AND SECURITY

Settlement of disputes. The preamble to the Charter begins with the declaration that the peoples of the United Nations are determined "to save succeeding generations from the scourge of war." The primary responsibility for the maintenance of international peace and security is placed on the Security Council. The founders realized that without basic agreement among the great powers on important international issues there could be no effective cooperation in the maintenance of peace or the application of sanctions against an aggressor. For this reason the Charter provided that substantive decisions would require the unanimous vote of the five permanent members of the Security Council. But this requirement means that when there is disagreement on substantive matters among the great powers the Security Council is unable to act. It was assumed by those who drafted the Charter that the five great powers who were given permanent membership in the council would cooperate in most such matters. Almost immediately there developed a sharp rift between the Soviet Union and the Western powers; and the Soviet Union, being in a minority, resorted to frequent use of the veto.

The "veto problem"

Article 11, paragraph 2, of the Charter authorizes the General Assembly to "discuss any questions relating to the maintenance of international peace and security" and to "make recommendations with regard to any such questions to the state or states concerned or to the Security Council or to both." This authorization is restricted by the provision of Article 12 that, "while the Security Council is exercising in respect of any dispute or situation the functions assigned to it in the present Charter, the General Assembly shall not make any recommendations with regard to that dispute or situation unless the Security Council so requests." While these provisions grant the assembly a broad secondary role, there is an important difference between the two bodies. The Security Council can make decisions that bind all the members, whereas the General Assembly can only make recommendations.

The General Assembly in November 1950 adopted the Uniting for Peace Resolution. It provided for immediate consideration by the General Assembly of a threat to the peace, breach of the peace, or act of aggression, with a view to recommending measures, including the use of armed force, "if the Security Council, because of lack of unanimity of the permanent members, fails to exercise its primary responsibility." The resolution established a Peace Observation Commission to "observe and report on the situation in any area where there exists international tension the continuance of which is likely to endanger the maintenance of international peace and security." It also established a Collective Measures Committee to "study and report on methods which the Assembly might use in strengthening international peace and security."

The Security Council must first seek to bring about a pacific settlement of disputes, in accordance with Chapter VI of the Charter. If the parties to a dispute that threatens the peace fail to settle it by peaceful means of their own choice, the Security Council must call upon them to settle it. The council may investigate any dispute or situation in order to determine whether its continuance is likely to endanger international peace and security. Any state, whether it is a member of the UN or not, may bring any such dispute or situation to the attention of the Security Council or the General Assembly. At any stage of the dispute or situation the Security Council may recommend appropriate procedures or methods of adjustment, and if the parties fail to settle the dispute by peaceful means the Security Council may recommend terms of settlement.

Whenever the Security Council determines that a threat to the peace exists or that a breach of the peace or act of aggression has taken place, it may decide upon measures to be taken. The Charter envisages the application of graduated measures, from economic and diplomatic sanctions to military action. By subscribing to the Charter, all the members undertook to place at the disposal of the Security Council armed forces and facilities for military sanctions against aggressors or disturbers of the peace; but this provision did not become operative because no agreements to give it effect were concluded.

A large number of disputes and situations involving the maintenance of peace have come before the Security Council and the General Assembly, reflecting the unsettled conditions in the world since World War II. The United Nations has succeeded in helping to settle some of them, such as the removal of Soviet troops from northern Iran in 1946; others have remained unsolved.

Regulation of armaments. The Security Council is responsible for formulating plans to regulate armaments. The General Assembly may discuss and recommend principles governing "disarmament and the regulation of armaments." The development of the atomic bomb during World War II created a situation in which it seemed to the United States and other governments that the international control of atomic energy demanded consideration even in advance of any perfection of collective security arrangements. Consequently, the first act of the United Nations was to establish the Atomic Energy Commission in January 1946. In December 1946 the General Assembly adopted a resolution providing for the urgent consideration of the control of atomic weapons and other weapons of mass destruction and for the regulation and reduction of all armaments and armed forces.

Atomic Energy Commission

The Atomic Energy Commission began its deliberations in June 1946 and submitted reports to the Security Council in 1946, 1947, and 1948. The majority of the members of the commission called for international managerial control or ownership of atomic energy facilities, international inspection by a proposed international atomic development authority, and the elimination of the veto from enforcement provisions of the agreement. The majority also insisted that the control system should be in operation before the existing stockpile of atomic bombs was destroyed. The Soviet Union, however, refused to agree to international ownership and to the kind of international inspection demanded by the majority; it demanded that the veto be applicable to enforcement measures and that the destruction of atomic stockpiles accompany the establishment of a control system. The commission recorded an impasse in negotiations in its 1948 report.

To deal with armaments other than weapons of mass destruction, the Security Council organized the Commission for Conventional Armaments, but progress in this field was also blocked by disagreement between the Soviet Union and the Western powers. Although the commission and the General Assembly in 1949 approved a plan

whereby each state would submit full information to the commission on its conventional armaments and armed forces, the refusal of the Soviet Union to accept the plan prevented any implementation.

Disarmament Commission

In January 1952 the General Assembly voted to merge the Atomic Energy Commission and the Conventional Armaments Commission into a Disarmament Commission. The Disarmament Commission, which consisted of the members of the Security Council and Canada, was directed to prepare proposals for the regulation, limitation, and balanced reduction of all the armed forces and armaments, for the elimination of all weapons of mass destruction, and for the international control of atomic energy to ensure its use for peaceful purposes only. In spite of vigorous efforts to achieve constructive results both through the commission and through the General Assembly itself, little progress was made. In 1957, however, the International Atomic Energy Agency was established to promote the peaceful uses of atomic energy.

A Ten-Nation Disarmament Committee, composed of five members from eastern Europe and five from the West, began deliberations in March 1959, but negotiations soon reached an impasse. The committee was enlarged to 18 members by the addition of neutral countries, but one member, France, did not participate. Both the Soviet Union and the Western powers declared that general and complete disarmament was their goal, but little agreement was reached on how to achieve that goal.

The General Assembly in 1961 adopted a resolution declaring the use of nuclear or thermonuclear weapons to be contrary to international law, to the UN Charter, and to the laws of humanity. The Nuclear Test-Ban Treaty was signed on Aug. 5, 1963, by the Soviet Union, the United Kingdom, and the United States. This agreement—to which more than 100 states later adhered—prohibited nuclear tests or explosions in the atmosphere, in outer space, and underwater. In 1966 the General Assembly unanimously approved a treaty prohibiting the placement of weapons of mass destruction in orbit, on the Moon, or on other celestial bodies and recognizing the use of outer space for peaceful purposes only. In January 1968 the assembly approved the draft of a "nonproliferation treaty" that banned the spread of nuclear weapons from nuclear to nonnuclear powers; the treaty has not been ratified by most of the near-nuclear powers, such as Argentina, Brazil, Egypt, Israel, Pakistan, and South Africa. In 1971 the assembly approved a treaty banning the emplacement of weapons of mass destruction on the seabed.

Since 1962 negotiations on disarmament questions have taken place in Geneva: in 1962–68, by the Eighteen-Nation Committee on Disarmament; in 1968–78, by the Conference of the Committee on Disarmament; and since then, by the Committee on Disarmament. In 1978 and 1982 the UN also organized two special sessions of the General Assembly on disarmament. The first of these sessions agreed on a general statement of organizational changes and a program of negotiations; the second broke down with little agreement.

Revision of the Charter. In accordance with Article 109, a proposal to call a conference to review the Charter was placed on the agenda of the General Assembly at its 10th session in 1955. The assembly decided that such a conference should be held in the future, but no date was set. Three proposals for Charter amendment were urged by a number of Latin-American states as early as 1956: to increase the number of nonpermanent members of the Security Council, to increase the membership of the Economic and Social Council, and to increase the number of judges on the International Court of Justice. As the membership of the United Nations rose to more than 100, the demand for enlargement of the councils became more insistent; some regions, particularly Africa, felt they were not given adequate representation.

The General Assembly adopted, on Dec. 17, 1963, a resolution proposing an amendment to the Charter to enlarge the Security Council to 15 (by increasing the number of nonpermanent members to 10) and to raise the membership of the Economic and Social Council to 27. Under the proposed revision, decisions of the Security Council on procedural matters would require an affirmative vote of any nine members, and decisions on all other matters would require nine affirmative votes, including those of the five permanent members. The proposed amendment provided for the geographic distribution of the seats of the nonpermanent members as follows: Africa and Asia, five; eastern Europe, one; Latin America, two; and western Europe and other regions, two. The Soviet Union, which previously had opposed any amendment to the Charter so long as China was barred from the United Nations, changed its position, and by September 1965 the two amendments had come into effect.

DEVELOPMENT OF INTERNATIONAL LAW

International Law Commission

In November 1947 the General Assembly established the International Law Commission of 15 members to make recommendations for the progressive development and codification of international law. The General Assembly directed the commission to formulate the principles of international law recognized at the Nürnberg trial of Nazi war criminals and to prepare a draft code of offenses against the peace and security of mankind. In 1950 the commission submitted its formulation of the Nürnberg principles, which covered crimes against the peace, war crimes, and crimes against humanity. The commission presented to the assembly in 1951 draft articles on offenses against the peace and security of mankind, which enumerated 12 crimes against international law, including any act of aggression, threat of or preparation for aggression, annexation of territory, and genocide. The commission also prepared a draft declaration on the rights and duties of states.

The commission made studies of the possibility of codifying certain branches of international law, giving priority to the law of treaties and the law of the sea (see INTERNATIONAL LAW). It also prepared a study on the legal aspects of reservations to multilateral conventions. Draft conventions were prepared on statelessness, the peaceful settlement of disputes, the law of the sea, diplomatic relations, consular relations, and the law of treaties. The commission was also concerned about the subjects of arbitral procedure and international criminal jurisdiction.

After the commission has completed a draft convention it submits it to the General Assembly, which may either convene an international conference to draw up formal conventions based on the draft or merely commend the text to the states. A conference to consider the draft convention on the law of the sea was held at Geneva in 1958; conferences held at Vienna in 1961 and 1963 completed conventions on diplomatic relations and optional protocols on the acquisition of nationality and the compulsory settlement of disputes and on consular relations. The commission in 1966 completed a draft convention on the law of treaties, which was placed before a conference meeting in split sessions in 1968 and 1969. The commission has since drafted conventions on special missions and on state succession and discussed the law relating to watercourses, the relations between states and international organizations, and other questions.

Prolonged attempts to arrive at a definition of aggression, undertaken both by the commission and by special committees set up by the General Assembly, eventually resulted in agreement on such a definition, passed without dissent in 1974. In 1970 the General Assembly passed a Declaration on Friendly Relations among States, which set out certain principles that should govern relations between states. In 1968 the UN Commission on International Trade Law (UNCITRAL) was established to discuss the harmonization and unification of international trade law.

ECONOMIC AND SOCIAL COOPERATION

A major purpose of the United Nations is "to achieve international cooperation in solving international problems of an economic, social, cultural, or humanitarian character, and in promoting and encouraging respect for human rights and for fundamental freedoms for all without distinction as to race, sex, language, or religion." The General Assembly, the Economic and Social Council, the Secretariat, and the specialized agencies are primarily re-

sponsible for action in this field. An important part of this aspect of UN activity consists of research, publication of reports, and technical assistance to governments. The United Nations has no authority to legislate or to enforce measures of economic and social cooperation; it can only make recommendations to the member governments.

Major annual publications

The UN publishes carefully prepared statistics and surveys of economic and social developments. Among the more significant of its yearly publications are the *World Economic Survey,* the *Statistical Yearbook,* the *Demographic Yearbook,* and the *Yearbook on Human Rights.*

Economic reconstruction. The devastation of large areas of the world and the disruption of economic relations during World War II resulted in the establishment of the United Nations Relief and Rehabilitation Administration in 1943. To assist in dealing with regional problems, the Economic and Social Council in 1947 established the Economic Commission for Europe and the Economic Commission for Asia and the Far East. Similar commissions were established for Latin America in 1948 and for Africa in 1958.

Technical assistance. A program of technical assistance to less developed countries was undertaken in 1946, when the General Assembly passed a resolution calling for the establishment of machinery in the Secretariat for giving such aid. The Expanded Program of Technical Assistance was approved by the General Assembly in 1949. It is financed by voluntary contributions from members.

Financing economic development. In 1954 the General Assembly recommended that both capital-importing and capital-exporting countries examine their policies and practices with a view to encouraging the flow of private capital. In April 1955 the International Bank submitted to its members the draft charter of the International Finance Corporation, to come into operation as soon as ratified by 30 states and 75 percent of the $100 million capital had been subscribed. The corporation was to make direct loans to private enterprises without government guarantees and was to be allowed to make loans for other than fixed returns. The charter entered into force on July 20, 1956.

In 1960 the International Development Association was established as an affiliate of the International Bank. It was created to make loans to less developed countries on terms that were more flexible than bank loans. It was authorized to finance any project "which will make an important contribution to the development of an area or areas concerned, whether or not the project is revenue producing or directly productive."

To assist in the financing of non-self-liquidating development projects a proposal was submitted to the General Assembly for a Special United Nations Fund for Economic Development (SUNFED). The General Assembly in 1957 unanimously adopted a resolution to set up a separate fund to provide systematic assistance in fields essential to technical, economic, and social development of less developed countries. The special fund began operations in 1959. The Expanded Program of Technical Assistance and the Special Fund were merged in 1965 to become the UN Development Programme (UNDP).

Trade and development. Because they were dissatisfied with the existing international trading system and because they believed that the General Agreement on Tariffs and Trade (GATT) was primarily interested in the trade problems of rich countries, developing countries demanded the establishment of a new organization to consider their trade problems. In 1964 a United Nations Conference on Trade and Development (UNCTAD) took place in Geneva and decided to establish the discussions on a continuing basis. There would be full conferences every three or four years, a Trade and Development Board of 55 members elected by the conference, a permanent secretariat under a secretary general (to be appointed by the UN secretary general, with the confirmation of the General Assembly), and committees to discuss specialized topics.

UNCTAD

Discussion in these bodies resulted in agreements on a "generalized system of preferences," providing for lower tariff rates for some exports of poorer countries, and on the creation of a Common Fund to help finance buffer stocks for commodity agreements. UNCTAD has also discussed questions related to shipping, insurance, commodities, the transfer of technology, and ways of assisting the exports of developing countries.

Refugees. The International Refugee Organization (IRO) was established in 1946 to take over the refugee functions of the United Nations Relief and Rehabilitation Administration, which expired in 1947. The IRO was successful in resettling, repatriating, transporting, and maintaining more than one million European refugees. It was abolished in 1952 and replaced by a new refugee structure. A United Nations high commissioner for refugees was appointed and was directed to act under the Convention Relating to the Status of Refugees, drawn up by the Economic and Social Council and approved by the General Assembly in 1951. An Advisory Committee on Refugees was appointed by the council in 1951 to assist the high commissioner. The commission has undertaken major operations to help refugees in western Europe, Africa, Asia, and Central America. A separate organization, the UN Relief and Works Agency (UNRWA), is responsible for aiding refugees in the Middle East.

Human rights. The General Assembly in 1948 adopted the Universal Declaration of Human Rights, which was prepared by the Commission on Human Rights. In 1948 the commission began to draft a Covenant of Human Rights that, upon ratification by governments, would become legally binding upon them. Wide differences in economic and social philosophies hampered efforts to achieve agreement, but the Covenants on Economic, Social, and Cultural Rights and the Covenant on Civil and Political Rights were agreed upon by the General Assembly in 1966. In 1976 the latter covenant, having been ratified by 35 states, entered into force, including provisions for regular cross-examination of ratifying countries concerning their implementation of the covenant. The General Assembly has adopted a Convention on the Elimination of all Forms of Racial Discrimination, which has come into force, and has passed declarations on many other human rights questions.

Commission on Human Rights

The UN Commission on Human Rights and its subcommission meet annually in Geneva to consider a wide range of human rights issues. Since the early 1970s these bodies have examined the human rights situations in particular countries, and in many cases they have passed resolutions concerning violations of human rights in those countries. They have also considered such questions as slavery, religious intolerance, detention without trial, and forcible incarceration in mental hospitals. The General Assembly has established a Special Committee on Apartheid, concerned with violations of human rights in South Africa.

Control of narcotics. The Commission on Narcotic Drugs was authorized by the General Assembly in 1946 to assume the functions entrusted by international conventions to the League of Nations Advisory Committee on Traffic in Opium and Other Dangerous Drugs. The agreements, conventions, and protocols on the control of narcotic drugs concluded in 1912, 1925, 1931, and 1936 were amended in a draft protocol approved by the assembly in November 1946 and subsequently came into force among its signatories. In addition to reestablishing the pre-World War II system of narcotics control, which had been disrupted by the war, the UN addressed new problems resulting from the development of synthetic drugs. Efforts were made to simplify the system of control by drafting one convention incorporating all the agreements in force.

Commission on Narcotic Drugs

Specialized agencies. The League of Nations provided a focal point for intergovernmental organizations and for the growing network of nongovernmental agencies. After World War II the United Nations became a kind of "roof organization" for the major enterprises in systematic international cooperation. The UN entered into coordinating agreements with intergovernmental agencies operating in economic, social, cultural, educational, health, and related fields. The Economic and Social Council has a consultative relationship with about 300 nongovernmental organizations.

General structure of UN agencies

Each agency has a general conference in which all the members are represented, and the conference elects an executive council that initiates proposals and carries out

decisions of the general conference. Each agency also has a permanent secretariat headed by a director. Many agencies have regional subcommissions operating in various parts of the world. Some of the specialized agencies were in existence before the United Nations was organized; some were in the process of establishment during World War II; and some agencies were organized under the auspices of the United Nations. (L.M.G./A.Va./E.Lu./Ed.)

International Labour Organisation. In 1919 the Peace Conference of Paris set up a commission for international labour legislation headed by Samuel Gompers, president of the American Federation of Labor. The commission put aside the proposals for a body with legislative authority and proposed a body with powers of recommendation to national governments, in composition a tripartite body in which half the representation would be by governments, and one-fourth each by labour and by employers. The peace conference adopted these proposals and, by inserting them in the Treaty of Versailles, set up the International Labour Organisation (ILO).

In its first decade the ILO was concerned primarily with research, with the definition and promotion of minimum standards of labour legislation, and with "mutual education" and some forms of collaboration among worker, employer, and government delegates and the office professional staff. During the 1930s the ILO sought ways to combat worldwide unemployment and economic depression.

After 1945 the Cold War, the breakup of European colonial empires, and the claims of the developing nations placed new tasks in the foreground for an organization the membership of which was no longer chiefly that of European, economically developed states but increasingly that of the less developed states. The ILO's major emphases shifted to human rights and to technical assistance and other work of an operating character in the interest of the less developed countries of Asia, Africa, Latin America, and Europe itself. The ILO's ideal has been universality of membership. It admitted Germany in 1919, long before the League of Nations did. Some nations—Brazil and, for a time, Japan—remained in the ILO after resigning from the League. Others became members of the ILO without joining the League or the UN. The United States, which did not join the League, joined the ILO in 1934; it withdrew in 1977 but rejoined in 1980. After World War II the ILO admitted the Federal Republic of Germany and in 1951, Japan. But South Africa was practically forced out of the organization because of its racial policies. Post-World War II amendments made it easier to join the ILO. Members of the UN do not have automatic membership in the ILO, as did League members, but may join by filing a declaration accepting the obligations of the ILO constitution. Nonmembers of the UN must be accepted by a vote of two-thirds of the conference, including that of two-thirds of the government delegates.

Membership of the ILO

The International Labour Office is the secretariat of the organization and its research staff. By 1970 it had international civil servants from more than 90 nations, including technical-assistance experts on temporary missions, working in Geneva, or in the field, all over the world. Their work was made possible by the ILO's own budget plus funds from other sources, chiefly UN Development Programme funds.

In 1962 the ILO helped to establish the International Institute for Labour Studies at Geneva. In 1965, in cooperation with the Italian government, it opened an International Centre for Technical and Vocational Training in Turin. During the 1970s the ILO launched its World Employment Programme, designed to encourage all countries, especially poor ones, to adopt development and industrial strategies likely to maximize employment for their populations. (V.R.L./E.Lu./Ed.)

International Bank for Reconstruction and Development and International Monetary Fund. The International Bank (better known as the World Bank) and the International Monetary Fund (IMF) had their origin in wartime preparations for postwar international financial and economic cooperation that culminated in the United Nations Monetary and Financial Conference held in July 1944 at Bretton Woods, N.H., U.S., and attended by 44 nations. The principal purposes of the World Bank are to assist in the reconstruction and development of its member countries by facilitating the investment of capital for productive purposes, thereby promoting the growth of international trade and improvement of standards of living, to promote private foreign investment by guarantees of and participation in loans and other investments made by private investors, and to make loans for productive purposes out of its own resources or funds borrowed by it when private capital is not available on reasonable terms.

World Bank

The World Bank came into existence on Dec. 27, 1945, when its articles of agreement were signed by 29 governments. The bank officially began operations at its headquarters in Washington, D.C., in June 1946. By 1988 the bank had 151 member countries. It has an executive board of 20 states, most of them representing a group of members, and a managing director. Voting in the bank is weighted according to initial contributions to the bank's capital.

The World Bank may lend funds directly, either from its capital funds or from funds that it borrows in private investment markets; it may guarantee loans made by others; or it may participate in such loans. Loans may be made to member countries directly or to any of their political subdivisions or to private business or agricultural enterprises in the territories of members. When the member government in whose territory the project is located is not itself the borrower, however, this member government must guarantee the loan.

The World Bank obtains its funds for loans from paid-in capital subscription, from borrowings in the capital markets of the world, and from net earnings. Sales to investors of portions of the bank's loan portfolio and repayments of loans to the bank, while representing only a recovery of funds originally derived from one of the above sources, have the same effect as new capital in that they reduce the amount the bank would otherwise have to obtain from other sources. The capital markets of the world provide the largest amount of funds for loans. The bank's soft-loan affiliate, the International Development Association (IDA), lends to low-income countries on favourable terms, free of interest except for a small service charge. The International Finance Corporation (IFC), another affiliate, provides loans to private business in developing countries.

Most of the World Bank's loans at first went for large-scale infrastructure projects—roads, railways, ports, power projects, and telecommunications. Since about 1970 an increasing proportion of World Bank lending has been for agricultural, educational, and population programs.

International Monetary Fund

The International Monetary Fund was designed to stabilize international monetary rates. It came into existence in March 1946 after the ratification and appropriation of funds by national governments had been completed, but the IMF was not actually opened until March 1947, and the first transactions were made in May of that year. Operating funds of the IMF are subscribed by member governments. Each member has a quota, of which an amount equal either to 25 percent of the quota or 10 percent of the member's holdings of gold and U.S. dollars, whichever is smaller, is subscribed in gold and the remainder in national currency. The IMF assists members over temporary difficulties in their international balance of payments. It does so partly by consultation and technical advice, but also by allowing members to purchase from it with their own national currencies the gold or foreign exchange they need. In principle, not more than 25 percent of the member's quota may be purchased in any one year, but in an emergency this limit may be waived. Not more than 125 percent of the quota may be purchased in all. Charges begin at 0.5 percent for the first 25 percent for the first year, rising by 0.5 percent for each additional year, with a limit of 5 percent. Members may arrange standby credits to use as and if necessary. The IMF's lending is intended to be for short-term purposes while the borrower overcomes balance-of-payment problems. Most lending is to developing countries, which—with the increasing cost of imports, uncertain export markets, and heavy debt burdens—have become especially dependent on this help. The IMF has created a number of "facilities" under which

it provides loans for countries facing particular difficulties. Members are required to repurchase the excess of their own currencies over their quotas within an agreed period as their balances improve. Members agree not to alter the exchange value of their currencies (except once by no more than 10 percent) without prior IMF agreement.

The expansion of world trade, coupled with a succession of international financial crises, created a demand for additional reserves that could be used in settlement of international balances. In October 1969 the IMF voted to distribute a total of $9.5 billion in Special Drawing Rights. These enlarged members' quotas without any additional subscription either in gold or in national currencies, and thus created a base for new credit expansion. But demands upon international liquidity continued to increase. A realignment of currencies agreed upon in December 1971 included a devaluation of the U.S. dollar and the upvaluation of other major currencies. Another realignment came after the United States announced a new devaluation in February 1973. A number of countries allowed their currencies to float—*i.e.,* to change in value according to market conditions and without formal announcement.

(J.H.A./J.B.C./E.Lu./Ed.)

Food and Agriculture Organization. The FAO, first of the permanent specialized agencies of the United Nations to be founded after World War II, came into formal being in October 1945 with the signing of its constitution at a conference held in Quebec. The immediate factor leading to its foundation was the Conference on Food and Agriculture convened at the request of U.S. President Franklin D. Roosevelt at Hot Springs, Va., in 1943. In 1951 the organization was transferred from its temporary headquarters in Washington, D.C., to a permanent seat in Rome.

Basis of the FAO

The ideas underlying the foundation of the FAO came from two sources. First was the International Institute of Agriculture founded in Rome in 1905. The IIA had been designed to protect farmers against the effects of sudden slumps and gluts and was, therefore, concerned with information about market trends and agricultural statistics. Second was the League of Nations, which in the period immediately before World War II had been interested in problems of nutrition and their relationship to health. Both the IIA and the League had been principally concerned with the more advanced countries; in the foundation of the FAO, however, several new and developing countries took great interest and played an active part.

Because the FAO was in part an answer to the problem of feeding vast populations in countries whose economies had been disrupted by World War II, it was natural that the first few years should be devoted to trying to bring about a rapid increase in the world's food supply. The FAO's member countries also had an eye on the possible emergence of surpluses of some commodities in certain countries should the distribution system break down or should needy governments not be able to pay for the food they required. An appreciation of the rapid rise in population in virtually all parts of the world in the immediate postwar period added urgency to the organization's work.

With the creation of the United Nations Expanded Programme of Technical Assistance, further extrabudgetary sums became available for the FAO's field operations and from 1950 onward an increasing number of countries received technical assistance from the FAO under the terms of that fund. The range of projects included agriculture, nutrition, forestry, fisheries, and economics.

Most of the experts working under the technical-assistance program are on individual assignments, advising member governments on specific problems. The FAO also helps its member governments through regional projects, including plant and animal disease control schemes. A broad educational program offers training centres and seminars in many subjects and a fellowship program run in connection with the assignment of technical-assistance experts.

The FAO publishes the *Plant Protection Bulletin,* covering the work of a worldwide reporting service on plant diseases and quarantine regulations. There is a parallel informational service for animal diseases. The FAO also produces a monthly bulletin of agricultural statistics and allied yearbooks on production and trade in agriculture, fisheries, and forest products.

United Nations Educational, Scientific and Cultural Organization. The roots of UNESCO lie in the intellectual cooperation efforts of the League of Nations and in the 1945 UN founding conferences at San Francisco and London, at which scholars who wanted to help increase the exchange of knowledge drafted UNESCO's program.

Through conferences, seminars, and publications, through the promotion of research and exchange of information and knowledge, and through technical advisory services, UNESCO assists and supports the efforts of member states. The significance of its activities in any individual nation is limited. Even the allocation of the total budget among the neediest members would provide small amounts compared to what is spent by the countries themselves. The real effect is to be measured by the influence of internationally formulated standards of excellence, by the effectiveness of scholarly communication and exchanges, and by the quality of technical assistance provided.

Administration of UNESCO

UNESCO is administered by a director general and an international civil service of about 1,500 persons, subject to advice and direction from an executive board of 30 (originally 18) members elected by the General Conference and designated by individual member states. Its permanent headquarters are in Paris. The constitution provides for citizen advisory commissions in member states.

Emphasis has been on strengthening international professional (nongovernmental) organizations in education, science, and culture; establishing clearinghouses for the exchange of information; promoting international professional conferences, symposia, and seminars; and publishing scholarly abstracts. A Universal Copyright Convention and an Agreement on the Importation of Educational, Scientific, and Cultural Materials were drafted and submitted for ratification by member states. Since 1960 an increasing proportion of UNESCO's work has consisted of programs to assist developing countries, for example, in eliminating illiteracy, training teachers, and developing educational institutions. Part of this is financed by the UN Development Programme. UNESCO also provides assistance in developing library resources and fellowships for study in more highly developed areas of the world.

Although UNESCO's contributions have been mainly indirect, exceptions have been the annual *Study Abroad,* a comprehensive index to foreign study opportunities; seminars for teachers on methods to increase international understanding; some modest promotion of social science research into national character and the nature of race; a project urging member states to increase understanding between Asia and the West; and some efforts to promote international exchanges in the arts.

UNICEF

United Nations Children's Fund. The General Assembly established the United Nations International Children's Emergency Fund (later styled the UN Children's Fund but still officially called UNICEF) in December 1946 to provide for the needs of children in devastated areas. Financed by contributions from member states, UNICEF helps to feed children in more than 100 countries; works to prevent such diseases as tuberculosis, whooping cough, and diphtheria; and provides clothing and other needs. It was made a permanent UN organization in 1953.

(E.Lu./Ed.)

World Health Organization. WHO, established in 1948, is a specialized agency designed to further international cooperation for improved health conditions. WHO inherited various duties from the Health Organization of the League of Nations, set up in 1923, and from the International Office of Public Health at Paris, established in 1909. These duties included epidemic control, quarantine measures, and the standardization of drugs. Under its constitution, however, WHO is given a much broader mandate—to promote the attainment of "the highest possible level of health" by all peoples. Health is defined as "a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity," and good health is held to be fundamental to world peace and security.

The membership of WHO consists of sovereign states and

non-self-governing territories. The organization is financed primarily from annual contributions made by member governments on the basis of relative capacity to pay. In addition, after 1951 WHO was allocated substantial resources from the expanded technical-assistance program of the UN. Grants for special purposes have been also made to WHO from time to time by private foundations and individual governments.

Activities of WHO

The work of WHO embraces three fairly distinct categories of activities:

1. The provision of central clearinghouse and research services. Information about the occurrence of pestilential disease anywhere in the world is broadcast over an international radio network to national health authorities, seaports, airports, and ships at sea. In 1952, 13 outdated international sanitary agreements were replaced by a codified set of international sanitary regulations designed to standardize quarantine measures without interfering unnecessarily with trade and air travel across national boundaries. The central WHO secretariat issues numerous publications, sets statistical standards, and keeps member countries informed of the latest developments in use of vaccines, cancer research, nutrition, control of drug addiction, and health hazards of nuclear radiation.

2. Measures for the control of epidemic and endemic disease. This category consists chiefly of mass campaigns promoted by WHO against communicable diseases. These campaigns have been successful in eliminating smallpox from the world, largely eliminating plague, and reducing the incidence of several other diseases. Among the techniques employed in these campaigns are nationwide vaccination programs, instruction in the use of antibiotics and insecticides, the improvement of laboratory and clinical facilities for early diagnosis and prevention of disease, assistance in providing pure water supplies and adequate sanitation systems, and health education for rural communities. WHO has promoted the use of paramedics and primary health care in developing countries and has published a code of practice concerning the advertisement of baby foods and a list of simple generic drugs that can be substituted for more expensive proprietary medicines.

3. Efforts to expand the public health administrations of member nations. As its program developed, WHO set as its most important task the strengthening of national and local health services, especially in Africa, Asia, and Latin America. The organization provides technical advice to governments in the preparation of long-term national health plans, sends out to the field international teams of experts to conduct surveys and demonstration projects, helps plan and set up local health centres, offers aid in the development of national training institutions for medical and nursing personnel, makes available teachers for on-the-spot short-course training experiments, and makes traveling fellowship awards to doctors, public health administrators, nurses, sanitary inspectors, and laboratory technicians. (W.R.S./E.Lu./Ed.)

DEPENDENT AREAS

The United Nations has maintained concern for people living in non-self-governing territories on two different levels: under principles and procedures developed for making administering states internationally accountable for the treatment of their non-self-governing territories, and under the United Nations trusteeship system.

Non-self-governing territories. Under Article 73 of the Charter, members of the United Nations responsible for administering non-self-governing territories agreed to ensure the people of such territories just treatment, protection against abuses, and advancement toward self-government. They also agreed to transmit to the secretary general information concerning the economic, social, and educational conditions in their territories. In 1947 the General Assembly created a special committee to examine information on these territories and to make recommendations to the administering authorities on the basis of information received. Attempts of the committee to obtain political information on territories from the administering states met with resistance in many instances.

There was controversy over the questions of which territories were non-self-governing and when they ceased to be non-self-governing. France stopped transmitting information on several territories on the ground that they had become either self-governing or an integral part of France. Later, Great Britain ceased to send information on Malta and Ghana, the United States on the Panama Canal Zone and Puerto Rico, Denmark on Greenland, and The Netherlands on Suriname and the Antilles. In 1952 the General Assembly adopted a resolution containing a list of factors to guide in deciding whether a territory had obtained a full measure of self-government. In 1947 the committee received information on 74 territories with about 215 million inhabitants; by 1956 this had declined to 55 territories with 115 million inhabitants. In 1963 the committee was dissolved and its functions were transferred to a Special Committee on Colonialism that had been created in 1961.

Decolonization

The anticolonial movement in the United Nations reached a high point in 1960, when the General Assembly adopted a resolution sponsored by 43 African and Asian states. This resolution, called the Declaration on the Granting of Independence to Colonial Countries and Peoples, condemned "the subjection of peoples to alien subjugation, domination, and exploitation" and declared that "immediate steps shall be taken . . . to transfer all powers" to the peoples in the colonies "without any conditions or reservations, in accordance with their freely expressed will and desire . . . in order to enable them to enjoy complete independence and freedom . . . ."

Spain, Portugal, South Africa, and Southern Rhodesia (whose unilateral declaration of independence in 1965 as Rhodesia was not recognized by the UN) were the chief targets of the anticolonial drive. Spain and Portugal had refused to transmit information on their overseas territories on the ground that they were not dependencies but provinces enjoying constitutional equality with the provinces in the homeland. Southern Rhodesia was a peculiar case; the situation in South Africa was somewhat similar, except that South Africa was completely independent. Hostility toward South Africa, Southern Rhodesia, and Portugal on the part of the other African countries became intense. In the 1970s the Portuguese territories won their independence, and in 1980 Southern Rhodesia was recognized as the independent nation of Zimbabwe. Only about 20 dependent territories remained, mostly small, often isolated islands, many of which did not want independence. (L.M.G./A.Va./E.Lu./Ed.)

Trusteeship system. The trusteeship system of the UN, the successor to the mandate system of the League of Nations, was established on the principle that colonial territories wrested from defeated enemies should not be annexed by any victorious nation but should be administered by mandatory or trust power under international supervision until they were able to determine their own future status.

Eleven such territories taken from Germany, Italy, and Japan were brought under the trusteeship system after 1945. With the attainment of independence by Togo, the British Cameroons (part of which joined Nigeria, the remainder becoming part of Cameroon), the French Cameroons (now Cameroon), Somalia, Tanganyika (now Tanzania), Western Samoa, Ruanda-Urundi (which became the separate countries of Rwanda and Burundi), New Guinea (now part of Papua New Guinea), and Nauru, by 1980 only the Trust Territory of the Pacific Islands remained under trusteeship. Failure to utilize the trusteeship system for other colonial areas was attributed to two factors. First, after 1945 many colonies attained independence without being made subject to the trusteeship system. Second, the colonial powers were reluctant to bring their colonial territories under the system because of the critical attitude shown in the debates of the United Nations toward any form of colonial relationship.

Differences between the mandate and trusteeship systems

The UN trusteeship system differs in two important respects from the mandate system of the League of Nations. The first difference is the provision that facilitates the submission of petitions from inhabitants of the trust territory or from any other source. In addition, the Charter provides that a petitioner may appear in person to present his case

before the Trusteeship Council or the General Assembly. Tens of thousands of petitions have been submitted to the United Nations. The second difference is the Charter provision for visits by UN missions to the trust territories. In practice, each territory is visited by such a mission every three years. The purpose of the missions is to supplement information provided in the annual reports of the administering authorities and to provide objective evaluations of the conditions in each territory. Each mission is composed of two persons drawn from administering countries and two persons selected from nonadministering countries, accompanied by five or six members of the UN Secretariat. The two members from administering countries are not drawn from among the nationals of the country administering the territory.

The Trusteeship Council was established by the Charter under the authority of the General Assembly. The council is empowered to consider reports made to it by the administering authorities, accept and examine petitions in consultation with the appropriate administering authority, provide visiting missions to trust territories, and take such other actions as are in conformity with the terms of trusteeship agreements between each administering authority and the General Assembly or, in the case of a strategic trust territory, with the Security Council.

The Trusteeship Council consists of an equal number of members administering trust territories and members not administering such territories. All of the permanent members of the Security Council are also members of the Trusteeship Council. To keep the council in balance, as many nonadministering members are elected for three-year terms as is necessary to ensure that the council is equally divided between the administering and nonadministering members.

All of the former mandatory powers entered into UN trusteeship agreements after 1945 except South Africa, which refused to enter into such an agreement for the mandated territory of South West Africa.

Strategic areas

The UN Charter (Article 82) provides that certain areas, which may include part or all of a trust territory, can be designated as strategic areas. This provision was placed in the Charter in 1945 on the initiative of the United States, and in 1947 a strategic trusteeship agreement was concluded between the United States and the United Nations that placed the Pacific Islands (the Marshalls, Marianas, and Carolines) under the trusteeship system. All of the other trust agreements were made between the administering authority and the General Assembly, but that for the Pacific Islands was made with the Security Council. In practice there is no appreciable difference in the way the two types of territories are handled. In the agreement with the Security Council the United States provided that it might close off any area for security reasons. This provision has never been put into effect, and the United States permits UN visiting missions to go through the territory and allows the Trusteeship Council to examine the annual reports on political, economic, and social conditions in the territory. (O.B.G./E.Lu./Ed.)

Administration

FINANCES

The secretary general must submit a biennial budget includingestimated expenditures to the General Assembly for its approval. The Charter stipulates that the expenses of the organization shall be borne by members as apportioned by the General Assembly. The Committee on Contributions prepares a scale of assessments for each member, based on the general economic level and capacity of each state, and this scale is submitted to the General Assembly for approval. The United States is the largest contributor, but other members make a larger per capita contribution.

When the cost of the special programs, specialized agencies, and peacekeeping operations is added to the regular budget, the total annual cost of the United Nations system becomes much greater. The special programs are financed by voluntary contributions from UN members, and the specialized agencies have their own budgets. The rapid increase in membership after 1955 did not help UN finances, for most of the new members were economically weak. Nearly all of them needed help, and their financial contributions to the UN were minimal.

SECRETARIAT

The organization is served by a substantial Secretariat, which is responsible for preparing numerous reports, studies, and investigations, in addition to the major tasks of providing services for large numbers of meetings, translating, interpreting, and other work. Under the Charter the staff is to be recruited mainly on the basis of merit, but "due regard is to be paid to the importance of recruiting the staff on as wide a geographical basis as possible." Some members of the Secretariat are engaged on permanent contracts, while others serve on temporary secondment from their national governments. In both cases, they must take an oath of loyalty to the UN and are not to receive instructions from member governments.

A United Nations Field Service and a United Nations Panel of Field Observers were organized by the General Assembly in 1949. The Field Service looks after the safety of UN missions in many parts of the world, and the Panel of Field Observers assists the various missions in supervising truces and observing plebiscites. Both are under the supervision of the secretary general.

PRIVILEGES AND IMMUNITIES

A General Convention on Privileges and Immunities of the United Nations, approved by the General Assembly in February 1946 and accepted by most of the members except the United States, asserts that the UN possesses juridical personality. The convention also provides for such matters as immunity from legal process of the property and officials of the United Nations. An agreement between the UN and the United States, signed in June 1947, defines the privileges and immunities of the UN headquarters in New York City.

HEADQUARTERS

The General Assembly decided during the second part of its first session in New York City to locate its permanent headquarters there. John D. Rockefeller, Jr., gave land for a building site in Manhattan. Temporary headquarters were established at Lake Success on Long Island, N.Y. The permanent Secretariat building was completed and occupied in 1951. The building providing accommodations for the General Assembly and the councils was completed and occupied in early 1952.

The UN flag, adopted in 1947, consists of the official emblem of the UN (a circular world map as seen from the North Pole surrounded by a wreath of olive branches) in white centred on a light blue background. The assembly designated October 24 as United Nations Day.

(L.M.G./A.Va./E.Lu./Ed.)

BIBLIOGRAPHY. See INIS L. CLAUDE, *Swords into Plowshares: The Problems and Progress of International Organization,* 4th ed. (1971); HERBERT G. NICHOLAS, *The United Nations as a Political Institution,* 5th ed. (1975); ROSALYN HIGGINS, *United Nations Peacekeeping: Documents and Commentary,* 4 vol. (1969–81); MARK W. ZACHER, *Dag Hammarskjöld's United Nations* (1970); PETER I. HAJNAL, *Guide to United Nations Organization, Documentation & Publishing for Students, Researchers, Librarians* (1978); EVAN LUARD, *International Agencies: The Emerging Framework of Interdependence* (1977), and *The United Nations: How It Works and What It Does* (1979); and UNITED NATIONS. DEPARTMENT OF PUBLIC INFORMATION, *Everyone's United Nations,* 10th ed. (1986). Contemporary information is found in the *Yearbook of the United Nations*; the *UN Chronicle* (monthly); and the *Report of the Secretary-General on the Work of the Organization* (annual). See also *Your United Nations: The Official Guidebook* (1982); *United Nations, Image and Reality: Questions and Answers About Management, Finance, and People* (1981); EVAN LUARD, *A History of the United Nations: The Years of Western Domination, 1945–1955* (1982); JOHN F. MURPHY, *The United Nations and the Control of International Violence: A Legal and Political Analysis* (1982); HANNA NEWCOMBE, *Design for a Better World* (1983); PETER I. HAJNAL, *Guide to UNESCO* (1983); and TOBY TRISTER GATI (ed.), *The US, the UN, and the Management of Global Change* (1983). (Ed.)

United States of America

The foremost country in the Western Hemisphere in population and economic development, the United States of America is a federal republic composed of 50 states. It is usually referred to simply as the United States or, colloquially, as America. The 48 coterminous states, which occupy the central one-third of North America, are bounded on the north by Canada, on the east by the Atlantic Ocean, on the south by the Gulf of Mexico and Mexico, and on the west by the Pacific Ocean. The newest states, Alaska and Hawaii, lie at the northwestern extremity of the continent and in the mid-Pacific, respectively. The national capital is Washington, which is coextensive with the District of Columbia, the federal capital region created in 1790.

The total area of the country is 3,679,192 square miles (9,529,063 square kilometres), making it the fourth largest in the world in area (after the Soviet Union, Canada, and China). Outlying territories and other politically associated areas in the Pacific Ocean and the Caribbean Sea add approximately 4,000 square miles to this figure.

The major characteristic of the United States is probably its great variety. Its physical environment ranges from the Arctic to the subtropical, from the moist rain forest to the arid desert, from the rugged mountain peak to the flat prairie. Although its total population is large by world standards, its overall population density is relatively low; the country embraces some of the world's largest urban concentrations as well as some of the most extensive areas that are almost devoid of habitation. The United States, like the Soviet Union and China, contains a highly diverse population; but unlike these other two countries—which largely incorporated indigenous peoples—its diversity has to a great degree come from an immense and sustained global immigration. Probably no other country has a wider range of racial, ethnic, and cultural types as does the United States. In addition to the presence of surviving native Americans (including American Indians, Aleuts, and Eskimo) and the descendants of Africans brought as slaves to America, the national character has been enriched, tested, and constantly redefined by the tens of millions of immigrants who by and large came to America hoping for greater social, political, and economic opportunities than they had in the places they left.

The United States is the world's greatest economic power, measured in terms of gross national product (GNP). The nation's wealth is partly a reflection of its rich natural resources and its enormous agricultural output, but it owes more to the country's highly developed industry. Despite its relative economic self-sufficiency in many areas, the United States is the most important single factor in world trade by virtue of the sheer size of its economy. Its exports and imports represent major proportions of the world total. The United States also impinges on the global economy as a source of and as a destination for investment capital. The country continues to sustain an economic life that is more diversified than any other on Earth, providing the majority of its people with one of the world's highest standards of living.

The United States is relatively young by world standards, being barely 200 years old; it achieved its current size only in the mid-20th century. America was the first of the European colonies to separate successfully from its motherland, and it was the first nation to be established on the premise that sovereignty rests with its citizens and not with the government. In its first century and a half, the country was mainly preoccupied with its own territorial expansion and economic growth and with social debates that ultimately led to civil war and a healing period that is still not complete. In the 20th century the United States emerged as a world power, and since World War II it has been one of the preeminent powers. It has not accepted this mantle easily nor always carried it willingly; the principles and ideals of its founders have been tested by the pressures and exigencies of its dominant status. Athough the United States still offers its residents opportunities for unparalleled personal advancement and wealth, the depletion of its resources, contamination of its environment, and continuing social and economic inequality that perpetuates areas of poverty and blight all threaten the fabric of the country. (Ed.)

This article first discusses the physical and human geography and the history of the United States as a whole. The 50 states, organized by region, are then treated individually in detail. The District of Columbia is discussed in the *Macropædia* article WASHINGTON, D.C. For discussion of other major U.S. cities, see the articles BOSTON, CHICAGO, LOS ANGELES, NEW ORLEANS, NEW YORK CITY, PHILADELPHIA, and SAN FRANCISCO. Political units in association with the United States include Puerto Rico, discussed in the article WEST INDIES, and several Pacific islands, discussed in PACIFIC ISLANDS.

The article is divided into the following sections:

Physical and human geography 150
- The land 150
  - Relief
  - Drainage
  - Climate
  - Plant life
  - Animal life
  - Settlement patterns
  - Traditional regions of the United States
- The people 191
  - Ethnic distribution
  - Religious groups
  - Immigration
- The economy 193
  - Resources
  - Sources of national income
  - Transportation
- Administration and social conditions 195
  - Government
  - Social services
- Cultural life 197
  - Literature
  - The visual arts
  - The theatre
  - Motion pictures
  - Dance
  - Audiences

History 202
- Colonial America to 1763 202
  - The European background
  - Settlement
  - Imperial organization
  - The growth of provincial power
  - Cultural and religious development
  - America, England, and the wider world
- The American Revolution and the early federal republic 209
  - Prelude to revolution
  - The War of Independence
  - Foundations of the American republic
  - The social revolution
  - The United States from 1789 to 1816
- The United States from 1816 to 1850 220
  - The Era of Mixed Feelings
  - The economy
  - Social developments
  - Jacksonian democracy
  - An age of reform
  - Expansionism and political crisis at midcentury
- Civil War 226

Prelude to war, 1850–60
Secession and the politics of the Civil War, 1860–65
The military background of the war
The land war
The naval war
Foreign affairs
The cost and significance of the Civil War
Reconstruction and the New South, 1865–1900 237
Reconstruction, 1865–77
The New South, 1877–90
The transformation of American society, 1865–1900 240
National expansion
Industrialization of the U.S. economy
National politics
Imperialism, the Progressive era, and the rise to world power, 1896–1920 248
American imperialism
The Progressive era
The rise to world power
The United States from 1920 to 1945 254
The postwar Republican administrations
The New Deal
World War II
The United States since 1945 259
The peak Cold War years, 1945–60
The late 20th century
New England 266
Connecticut 266
Maine 269
Massachusetts 273
New Hampshire 277
Rhode Island 280
Vermont 284
The Middle Atlantic region 287
Delaware 287
District of Columbia 290
Maryland 290
New Jersey 293
New York 296
Pennsylvania 302
The South 306
Alabama 306
Arkansas 310
Florida 313
Georgia 318
Kentucky 323
Louisiana 327
Mississippi 331
North Carolina 335
South Carolina 338
Tennessee 342
Virginia 346
West Virginia 350
The Midwest 353
Illinois 353
Indiana 357
Iowa 360
Kansas 363
Michigan 366
Minnesota 371
Missouri 375
Nebraska 378
North Dakota 382
Ohio 385
South Dakota 389
Wisconsin 392
The Southwest 396
Arizona 396
New Mexico 400
Oklahoma 404
Texas 407
The Mountain region 412
Colorado 412
Idaho 415
Montana 418
Nevada 421
Utah 424
Wyoming 428
The Pacific Coast 431
Alaska 431
California 435
Hawaii 441
Oregon 445
Washington 448
Bibliography 452

PHYSICAL AND HUMAN GEOGRAPHY

The land

The two great sets of elements that mold the physical environment of the United States are, first, the geologic, which determines the main patterns of landforms, drainage, and mineral resources and influences soils to a lesser degree, and, second, the atmospheric, which dictates not only climate and weather but also in large part the distribution of soils, plants, and animals. Although these elements are not entirely independent of one another, each produces on a map patterns that are so profoundly different that essentially they remain two separate geographies. (Since this portion of the article covers only the coterminous United States, see also the sections *Alaska* and *Hawaii*.)

RELIEF

The grand geologic pattern

The centre of the coterminous United States is a great sprawling interior lowland, reaching from the ancient shield of central Canada on the north to the Gulf of Mexico on the south. To east and west this lowland rises, first gradually and then abruptly, to mountain ranges that divide it from the sea on both sides. The two mountain systems differ drastically. The Appalachian Mountains on the east are low, almost unbroken, and in the main set well back from the Atlantic. From New York to the Mexican border stretches the low Coastal Plain, which faces the ocean along a swampy, convoluted coast. The gently sloping surface of the plain extends out beneath the sea, where it forms the continental shelf, which, although submerged beneath shallow ocean water, is geologically identical to the Coastal Plain. Southward the plain grows wider, swinging westward in Georgia and Alabama to truncate the Appalachians along their southern extremity and separate the interior lowland from the Gulf.

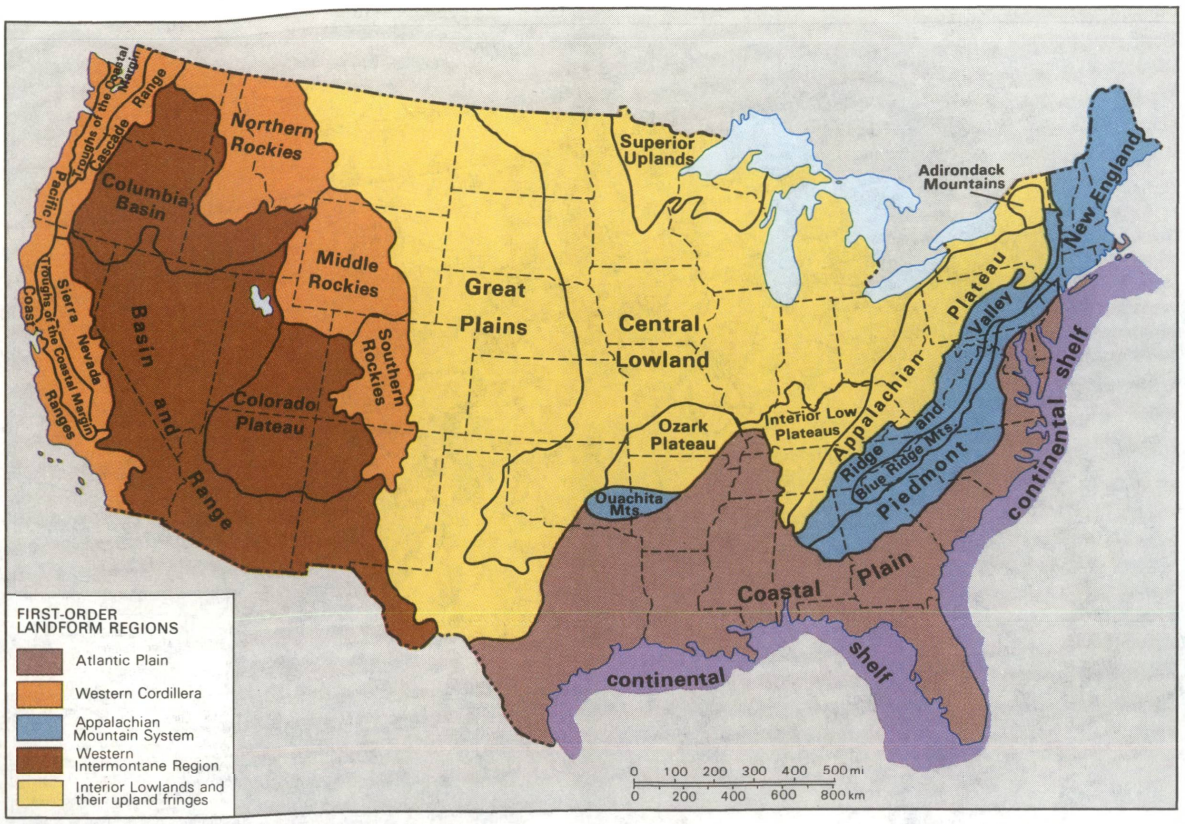
West of the Central Lowland is the mighty Cordillera, part of a global mountain system that rings the Pacific Basin. The Cordillera encompasses fully one-third of the United States, with an internal variety commensurate with its size. At its eastern margin lie the Rocky Mountains, a high, diverse, and discontinuous chain that stretches all the way from New Mexico to the Canadian border. The Cordillera's western edge is a Pacific coastal chain of rugged mountains and inland valleys, the whole rising spectacularly from the sea without benefit of a coastal plain. Pent between the Rockies and the Pacific chain is a vast intermontane complex of basins, plateaus, and isolated ranges so large and remarkable that they merit recognition as a region separate from the Cordillera itself.

These regions—the Interior Lowlands and their upland fringes, the Appalachian Mountain system, the Atlantic Plain, the Western Cordillera, and the Western Intermontane Region—are so various that they require further division into 24 major subregions, or provinces (see map).

The Interior Lowlands and their upland fringes. Andrew Jackson is supposed to have remarked that the United States begins at the Alleghenies, implying that only west of the mountains, in the isolation and freedom of the great Interior Lowlands, could people finally escape Old World influences. Whether or not the lowlands constitute the country's cultural core is debatable, but there can be no doubt that they comprise its geologic core and in many ways its geographic core as well.

The mid-continent of the nation

This enormous region rests upon an ancient, much-eroded platform of complex crystalline rocks that have for the most part lain undisturbed by major orogenic (mountain-building) activity for more than 600,000,000 years. Over much of central Canada, these Precambrian rocks are exposed at the surface and form the continent's



Physiographic regions of the United States.

single largest topographical region, the formidable and ice-scoured Canadian Shield.

In the United States most of the crystalline platform is concealed under a deep blanket of sedimentary rocks. In the far north, however, the naked Canadian Shield extends into the United States far enough to form two small but distinctive landform regions: the rugged and occasionally spectacular Adirondack Mountains of northern New York; and the more subdued but austere Superior Uplands of northern Minnesota, Wisconsin, and Michigan. As in the rest of the shield, glaciers have stripped soils away, strewn the surface with boulders and other debris, and obliterated preglacial drainage systems. Most attempts at farming in these areas have been abandoned, but the combination of a comparative wilderness in a northern climate, clear lakes, and white-water streams has fostered the development of both regions as year-round outdoor recreation areas.

Mineral wealth in the Superior Uplands is legendary. Iron lies near the surface and close to the deepwater ports of the upper Great Lakes. Iron is mined both north and south of Lake Superior, but best known are the colossal deposits of Minnesota's Mesabi Range, for more than a century one of the world's richest and a vital element in America's rise to industrial power. In spite of depletion, the Minnesota and Michigan mines still yield a major proportion of the country's iron and a significant percentage of the world's supply.

The Central Lowland

South of the Adirondack Mountains and Superior Uplands lies the boundary between crystalline and sedimentary rocks; abruptly, everything is different. The core of this sedimentary region—the heartland of the United States—is the great Central Lowland, which stretches for 1,500 miles (2,400 kilometres) from New York to central Texas and north another 1,000 miles to the Canadian province of Saskatchewan. To some, the landscape may seem dull, for heights of more than 2,000 feet (600 metres) are unusual, and truly rough terrain is almost lacking. Landscapes are varied, however, largely as the result of glaciation that directly or indirectly affected most of the subregion. North of the Missouri–Ohio river line, the advance and readvance of continental ice left an intricate mosaic of boulders, sand, gravel, silt, and clay and a complex pattern of lakes and drainage channels, some abandoned, some still in use. The southern part of the Central Lowland is quite different, covered mostly with loess (wind-deposited silt) that further subdued the already low relief surface. Elsewhere, especially near major rivers, postglacial streams carved the loess into rounded hills, and visitors have aptly compared their billowing shapes to the waves of the sea. Above all, the loess produces soil of extraordinary fertility. As the Mesabi iron was a major source of America's industrial wealth, its agricultural prosperity has been rooted in Midwestern loess.

The Central Lowland resembles a vast saucer, rising gradually to higher lands on all sides. Southward and eastward, the land rises gradually to three major plateaus. Beyond the reach of glaciation to the south, the sedimentary rocks have been raised into two broad upwarps, separated from one another by the great valley of the Mississippi River. The Ozark Plateau lies west of the river and occupies most of southern Missouri and northern Arkansas; on the east the Interior Low Plateaus dominate central Kentucky and Tennessee. Except for two nearly circular patches of rich limestone country—the Nashville Basin of Tennessee and the Kentucky Bluegrass region—most of both plateau regions consists of sandstone uplands, intricately dissected by streams. Local relief runs to several hundreds of feet in most places, and visitors to the region must travel winding roads along narrow stream valleys. The soils there are poor, and mineral resources are scanty.

Eastward from the Central Lowland the Appalachian Plateau—a narrow band of dissected uplands that strongly resembles the Ozark Plateau and Interior Low Plateaus in steep slopes, wretched soils, and endemic poverty—forms a transition between the interior plains and the Appalachian Mountains. Usually, however, the Appalachian Plateau is considered a subregion of the Appalachian Mountains, partly on grounds of location, partly because of geologic structure. Unlike the other plateaus, where rocks are warped upward, the rocks there form an elongated basin, wherein bituminous coal has been preserved from erosion. This Appalachian coal, like the Mesabi iron that it complements in U.S. industry, is extraordinary. Extensive, thick, and close to the surface, it has stoked the furnaces of northeastern steel mills for

©Encyclopædia Britannica Inc.
PACIFIC OCEAN
GREAT BASIN
HARNEY BASIN
COLORADO PLATEAU
GREAT PLAINS
ESTACADO
LLANO
MEXICO
CANADA
ALASKA
HAWAII
PACIFIC OCEAN
Scale 1:13,966,000
1 inch equals approx. 220 miles
Scale 1:31,657,000
1 inch equals approx. 500 miles
ALEUTIAN ISLANDS
BERING SEA
Gulf of Alaska
ARCTIC OCEAN

Pacific Coast Map 9
Mountain Map 7
Middle West Map 5
New England Map 1
Middle Atlantic Map 2
Mountain Map 8
The South Map 3
Pacific Coast Map 10
The Southwest Map 6
The South Map 4
Hawaii
Alaska Map 11
For maps of specific U.S. regions, see pp. 173-188.
Cities over 1,000,000
Cities 500,000 to 1,000,000
Cities 100,000 to 500,000
Cities under 100,000
National capitals
State capitals
OREGON State names
International boundaries
State boundaries
Canals
Aqueducts
Intracoastal waterway
Intermittent rivers
Dams
Intermittent lakes
Salt lakes
Glaciers
Swamps and marshes
Sand areas
National parks
Indian Reservations
Spot elevations in metres (1 m = 3.28 ft)
Albers Equal-Area Projection
Scale 1:13,966,000
1 inch equals approx. 220 miles
0 50 100 200 300 400 mi
0 100 200 300 400 500 600 km
Gulf of Mexico
ATLANTIC OCEAN
Straits of Florida
CUBA
Havana
BAHAMAS
Lake Superior
Lake Michigan
Lake Huron
Lake Erie
Lake Ontario
Chicago
Detroit
Cleveland
Toronto
Montreal
Ottawa
Quebec
New York City
Philadelphia
Baltimore
Washington, D.C.
Boston
Pittsburgh
Buffalo
Milwaukee
Minneapolis
St. Paul
Indianapolis
Columbus
Cincinnati
Memphis
Nashville
Atlanta
Birmingham
Jacksonville
Miami
Tampa
Orlando
New Orleans
Houston
Dallas
Kansas City
St. Louis
Denver
Omaha
Des Moines
Seattle
Portland
San Francisco
Los Angeles
San Diego
Las Vegas
Phoenix
Albuquerque
Salt Lake City
Boise
Billings
Bismarck
Fargo
Sioux Falls
Cheyenne
Oklahoma City
Wichita
Little Rock
Jackson
Shreveport
Charleston
Columbia
Raleigh
Norfolk
Virginia Beach
Richmond
Hartford
Providence
Albany
Syracuse
Rochester
Buffalo
Niagara Falls
Grand Rapids
Lansing
Flint
Saginaw
Toledo
Akron
Canton
Youngstown
Dayton
Louisville
Lexington
Knoxville
Chattanooga
Huntsville
Mobile
Pensacola
Tallahassee
Savannah
Charleston
Greensboro
Winston-Salem
Durham
Chapel Hill
Charlotte
Greenville
Spartanburg
Asheville
Roanoke
Harrisburg
Scranton
Allentown
Reading
Trenton
Atlantic City
Wilmington
Dover
Ocean City
Cape Hatteras
Cape Fear
Cape Canaveral
Key West
Key Largo
Everglades National Park
Florida Keys
Okefenokee Swamp
Great Smoky Mts.
Mount Mitchell 2037
Pine Mountain
Great Dismal Swamp
Chesapeake
Virginia Beach
Norfolk
Newport News
Hampton
Williamsburg
Petersburg
Lynchburg
Harpers Ferry
Frederick
Hagerstown
Gettysburg
Shenandoah
Ozark Plateau
Boston Mts.
Ouachita Mts.
Appalachian Mountains
Blue Ridge Mts.
Great Smoky Mts.
Pine Mountain
Mississippi
Missouri
Ohio
Rio Grande
Colorado
Columbia
Snake
Yukon
Mackenzie
Hudson Bay
Lake of the Woods
Lake Winnipeg
Great Slave Lake
Great Bear Lake
Lake Athabasca
Lake Winnebago
Lake Okeechobee
Lake Champlain
Lake George
Lake Pontchartrain
Lake Maracaibo
Lake Titicaca
Lake Nicaragua
Lake Managua
Lake Texcoco
Lake Chapala
Lake Patzcuaro
Lake Cuitzeo
Lake Xochimilco
Lake Tezcoco
Lake Izabal
Lake Atitlan
Lake Amatitlan
Lake Ilopango
Lake Coatepeque
Lake Yojoa
Lake Managua
Lake Nicaragua
Lake Cocibolca
Lake Xolotlan
Lake Enriquillo
Lake Azuei
Lake Peten Itza
Lake Izabal
Lake Atitlan
Lake Amatitlan
Lake Ilopango
Lake Coatepeque
Lake Yojoa
Lake Managua
Lake Nicaragua
Lake Cocibolca
Lake Xolotlan
Lake Enriquillo
Lake Azuei
Lake Peten Itza

MAP INDEX

This index includes all place-names that appear on the general U.S. map. All names are followed by a grid coordinate reference to be used in locating the subject. For rivers, the coordinates are those for the mouth; for physical regions, they refer approximately to the midpoint of the feature. The inset key on the general map outlines the regions depicted in greater detail on pages 173 to 188.

Political subdivisions

Alabama . . . . . 32 50 N 87 00 W
Alaska . . . . . 65 00 N 153 00 W
Arizona . . . . . 34 00 N 112 00 W
Arkansas . . . . . 34 50 N 92 30 W
California . . . . . 37 30 N 119 30 W
Colorado . . . . . 39 00 N 105 30 W
Connecticut . . . . 41 45 N 72 45 W
Delaware . . . . . 39 10 N 75 30 W
Florida . . . . . 28 00 N 82 00 W
Georgia . . . . . 32 50 N 83 15 W
Hawaii . . . . . 20 00 N 157 45 W
Idaho . . . . . 45 00 N 115 00 W
Illinois . . . . . 40 00 N 89 00 W
Indiana . . . . . 40 00 N 86 15 W
Iowa . . . . . 42 15 N 93 15 W
Kansas . . . . . 38 45 N 98 15 W
Kentucky . . . . . 37 30 N 85 15 W
Louisiana . . . . . 31 15 N 92 15 W
Maine . . . . . 45 15 N 69 15 W
Maryland . . . . . 39 00 N 76 45 W
Massachusetts . . 42 15 N 71 50 W
Michigan . . . . . 44 00 N 85 00 W
Minnesota . . . . . 46 00 N 94 15 W
Mississippi . . . . . 32 50 N 89 30 W
Missouri . . . . . 38 30 N 93 30 W
Montana . . . . . 47 00 N 110 00 W
Nebraska . . . . . 41 30 N 100 00 W
Nevada . . . . . 39 00 N 117 00 W
New Hampshire . 43 35 N 71 40 W
New Jersey . . . . . 40 15 N 74 30 W
New Mexico . . . . 34 30 N 106 00 W
New York . . . . . 43 00 N 75 00 W
North Carolina . . . 35 30 N 80 00 W
North Dakota . . . 47 30 N 100 15 W
Ohio . . . . . 40 15 N 82 45 W
Oklahoma . . . . . 35 30 N 98 00 W
Oregon . . . . . 44 00 N 121 00 W
Pennsylvania . . . . 40 45 N 77 30 W
Rhode Island . . . . 41 40 N 71 30 W
South Carolina . . 34 00 N 81 00 W
South Dakota . . . 44 15 N 100 00 W
Tennessee . . . . . 35 50 N 85 30 W
Texas . . . . . 31 30 N 99 00 W
Utah . . . . . 39 30 N 111 30 W
Vermont . . . . . 43 50 N 72 45 W
Virginia . . . . . 37 30 N 78 45 W
Washington . . . . . 47 30 N 120 30 W
West Virginia . . . . 38 45 N 80 30 W
Wisconsin . . . . . 44 45 N 89 30 W
Wyoming . . . . . 43 00 N 107 30 W

Cities and towns

Aberdeen, S.D. . . 45 28 N 98 29 W
Aberdeen, Wash. . . . . 46 59 N 123 50 W
Abilene, Kan. . . . 38 55 N 97 13 W
Abilene, Tex. . . . . 32 28 N 99 43 W
Ada, Okla. . . . . 34 46 N 96 41 W
Adrian, Mich. . . . . 41 54 N 84 02 W
Akron, Ohio . . . . . 41 05 N 81 31 W
Alamogordo, N.M. . . . . 32 54 N 105 57 W
Alamosa, Colo. . . 37 28 N 105 52 W
Albany, Ga. . . . . 31 35 N 84 10 W
Albany, N.Y. . . . . 42 39 N 73 45 W
Albert Lea, Minn. . 43 39 N 93 22 W
Albuquerque, N.M. . . . . 35 05 N 106 39 W
Alexandria, La. . . 31 18 N 92 27 W
Alexandria, Va. . . 38 48 N 77 03 W
Allentown, Pa. . . . 40 37 N 75 29 W
Alliance, Neb. . . . 42 06 N 102 52 W
Alpena, Mich. . . . 45 04 N 83 27 W
Alton, Ill. . . . . 38 53 N 90 11 W
Alturas, Calif. . . . 41 29 N 120 32 W
Altus, Okla. . . . . 34 38 N 99 20 W
Amarillo, Tex. . . . 35 13 N 101 50 W
Americus, Ga. . . . 32 04 N 84 14 W
Ames, Ia. . . . . 42 02 N 93 37 W
Anaconda, Mont. . . . . 46 08 N 112 57 W
Anchorage, Alaska . . . . . 61 13 N 149 54 W
Andalusia, Ala. . . 31 18 N 86 29 W
Anderson, Ind. . . . 40 10 N 85 41 W
Anderson, S.C. . . 34 31 N 82 39 W
Ann Arbor, Mich. . 42 17 N 83 45 W
Annapolis, Md. . . 38 59 N 76 30 W
Anniston, Ala. . . . 33 39 N 85 50 W
Appleton, Wis. . . 44 16 N 88 25 W
Arcata, Calif. . . . . 40 52 N 124 05 W
Arlington, Tex. . . . 32 44 N 97 07 W
Arlington, Va. . . . 38 53 N 77 07 W
Asbury Park, N.J. . . . . 40 13 N 74 01 W
Asheville, N.C. . . . 35 36 N 82 33 W
Ashland, Ky. . . . . 38 28 N 82 38 W
Ashland, Wis. . . . 46 35 N 90 53 W
Ashtabula, Ohio . 41 51 N 80 47 W
Aspen, Colo. . . . . 39 11 N 106 49 W
Astoria, Ore. . . . . 46 11 N 123 50 W
Athens, Ga. . . . . 33 57 N 83 23 W
Athens, Tenn. . . . 35 27 N 84 36 W
Atlanta, Ga. . . . . 33 45 N 84 23 W
Atlantic City, N.J. . . . . 39 21 N 74 27 W
Augusta, Ga. . . . . 33 28 N 81 58 W
Augusta, Me. . . . 44 19 N 69 47 W
Aurora, Colo. . . . . 39 43 N 104 49 W
Austin, Minn. . . . . 43 40 N 92 58 W
Austin, Tex. . . . . 30 17 N 97 45 W
Baker, Mont. . . . . 46 22 N 104 17 W
Baker, Ore. . . . . 44 47 N 117 50 W
Bakersfield, Calif. . . . . 35 23 N 119 01 W
Baltimore, Md. . . . 39 17 N 76 37 W
Bangor, Me. . . . . 44 48 N 68 46 W
Bar Harbor, Me. . . 44 23 N 68 13 W
Baraboo, Wis. . . . 43 28 N 89 45 W
Barrow, Alaska . . 71 18 N 156 47 W
Bartlesville, Okla. . . . . 36 45 N 95 59 W
Baton Rouge, La. . . . . 30 27 N 91 11 W
Battle Creek, Mich. . . . . 42 19 N 85 11 W
Bay City, Mich. . . 43 36 N 83 54 W
Beaumont, Tex. . . 30 05 N 94 06 W
Beckley, W.Va. . . 37 46 N 81 11 W
Bedford, Ind. . . . 38 52 N 86 29 W
Bellefontaine, Ohio . . . . . 40 22 N 83 46 W
Belleville, Ill. . . . . 38 31 N 89 59 W
Bellingham, Wash. . . . . 48 46 N 122 29 W
Beloit, Wis. . . . . 42 31 N 89 01 W
Bemidji, Minn. . . . 47 28 N 94 52 W
Bend, Ore. . . . . 44 04 N 121 19 W
Bennington, Vt. . . 42 53 N 73 12 W
Benton, Ark. . . . . 34 33 N 92 35 W
Benton Harbor, Mich. . . . . 42 07 N 86 26 W
Berkeley, Calif. . . 37 52 N 122 16 W
Berlin, N.H. . . . . 44 28 N 71 11 W
Bethel, Alaska . . . 60 48 N 161 45 W
Bethlehem, Pa. . . 40 37 N 75 23 W
Beulah, N.D. . . . . 47 15 N 101 46 W
Biddeford, Me. . . 43 30 N 70 28 W
Billings, Mont. . . . 45 47 N 108 30 W
Biloxi, Miss. . . . . 30 24 N 88 53 W
Binghamton, N.Y. . . . . 42 06 N 75 55 W
Birmingham, Ala. . 33 31 N 86 48 W
Bismarck, N.D. . . 46 48 N 100 47 W
Blakely, Ga. . . . . 31 23 N 84 56 W
Bloomington, Ill. . . 40 29 N 88 59 W
Bloomington, Ind. . . . . 39 10 N 86 32 W
Blythe, Calif. . . . 33 37 N 114 36 W
Boca Raton, Fla. . 26 21 N 80 05 W
Bogalusa, La. . . . 30 47 N 89 52 W
Boise, Idaho . . . . 43 37 N 116 13 W
Boston, Mass. . . . 42 22 N 71 04 W
Boulder, Colo. . . . 40 01 N 105 17 W
Boulder City, Nev. . . . . 35 59 N 114 50 W
Bowling Green, Ky. . . . . 36 59 N 86 27 W
Bozeman, Mont. . . 45 41 N 111 02 W
Bradenton, Fla. . . 27 30 N 82 34 W
Brady, Tex. . . . . 31 09 N 99 20 W
Brainerd, Minn. . . 46 22 N 94 12 W
Bremerton, Wash. . . . . 47 34 N 122 38 W
Bridgeport, Conn. . . . . 41 11 N 73 12 W
Brigham City, Utah . . . . . 41 31 N 112 01 W
Broadus, Mont. . . 45 27 N 105 25 W
Brookings, S.D. . . 44 19 N 96 48 W
Brownsville, Tex. . . . . 25 54 N 97 30 W
Brunswick, Ga. . . 31 10 N 81 30 W
Brunswick, Me. . . 43 55 N 69 58 W
Bryan, Tex. . . . . 30 40 N 96 22 W
Buffalo, N.Y. . . . . 42 53 N 78 53 W
Buffalo, Tex. . . . . 31 28 N 96 04 W
Bunkie, La. . . . . 30 57 N 92 11 W
Burbank, Calif. . . 34 11 N 118 18 W
Burlington, Colo. . 39 18 N 102 16 W
Burlington, Ia. . . . 40 48 N 91 06 W
Burlington, Vt. . . . 44 29 N 73 12 W
Burns, Ore. . . . . 43 35 N 119 03 W
Butte, Mont. . . . . 46 00 N 112 32 W
Cadillac, Mich. . . 44 15 N 85 24 W
Cairo, Ill. . . . . 37 00 N 89 11 W
Caldwell, Idaho . . 43 40 N 116 41 W
Cambridge, Mass. . . . . 42 22 N 71 06 W
Camden, N.J. . . . 39 55 N 75 07 W
Canton, Miss. . . . 32 37 N 90 02 W
Canton, Ohio . . . . 40 48 N 81 23 W
Cape Girardeau, Mo. . . . . 37 19 N 89 32 W
Carbondale, Ill. . . 37 44 N 89 13 W
Carlisle, Pa. . . . . 40 12 N 77 12 W
Carlsbad, N.M. . . 32 25 N 104 14 W
Carrollton, Ga. . . . 33 35 N 85 05 W
Carson City, Nev. . . . . 39 10 N 119 46 W
Casa Grande, Ariz. . . . . 32 53 N 111 45 W
Casper, Wyo. . . . 42 51 N 106 19 W
Cedar City, Utah . 37 41 N 113 04 W
Cedar Falls, Ia. . . 42 32 N 92 27 W
Cedar Rapids, Ia. . . . . 41 59 N 91 40 W
Chadron, Neb. . . . 42 50 N 103 00 W
Champaign, Ill. . . 40 07 N 88 15 W
Chapel Hill, N.C. . 35 55 N 79 04 W
Charleston, S.C. . 32 46 N 79 56 W
Charleston, W.Va. . . . . 38 21 N 81 39 W
Charlotte, N.C. . . 35 13 N 80 51 W
Charlottesville, Va. . . . . 38 02 N 78 30 W
Chattanooga, Tenn. . . . . 35 03 N 85 19 W
Chesapeake, Va. . 36 50 N 76 17 W
Chester, Pa. . . . . 39 51 N 75 22 W
Cheyenne, Wyo. . 41 08 N 104 49 W
Chicago, Ill. . . . . 41 53 N 87 38 W
Chico, Calif. . . . . 39 44 N 121 50 W
Chula Vista, Calif. . . . . 32 38 N 117 05 W
Cincinnati, Ohio . . 39 06 N 84 31 W
Claremore, Okla. . 36 19 N 95 36 W
Clarksdale, Miss. . 34 12 N 90 35 W
Clarkston, Wash. . 46 25 N 117 03 W
Clarksville, Tenn. . 36 31 N 87 21 W
Clayton, N.M. . . . 36 27 N 103 11 W
Clearwater, Fla. . . 27 58 N 82 48 W
Cleveland, Ohio . . 41 30 N 81 42 W
Clinton, Okla. . . . 35 31 N 98 58 W
Clovis, N.M. . . . . 34 24 N 103 12 W
Cody, Wyo. . . . . 44 32 N 109 03 W
Coeur d'Alene, Idaho . . . . . 47 41 N 116 46 W
College Station, Tex. . . . . 30 37 N 96 21 W
Colorado Springs, Colo. . . 38 50 N 104 49 W
Columbia, Mo. . . . 38 57 N 92 20 W
Columbia, S.C. . . 34 00 N 81 03 W
Columbia, Tenn. . 35 37 N 87 02 W
Columbus, Ga. . . 32 29 N 84 59 W
Columbus, Miss. . 33 30 N 88 25 W
Columbus, Neb. . 41 26 N 97 22 W
Columbus, Ohio . 39 58 N 83 00 W
Concord, N.H. . . . 43 12 N 71 32 W
Conway, S.C. . . . 33 51 N 79 03 W
Coos Bay, Ore. . . 43 22 N 124 12 W
Coral Gables, Fla. . . . . 25 45 N 80 16 W
Cordele, Ga. . . . . 31 58 N 83 47 W
Cordova, Alaska . 60 33 N 145 45 W
Corinth, Miss. . . . 34 56 N 88 31 W
Corpus Christi, Tex. . . . . 27 47 N 97 24 W
Corsicana, Tex. . . 32 06 N 96 28 W
Corvallis, Ore. . . . 44 34 N 123 16 W
Council Bluffs, Ia. . . . . 41 16 N 95 52 W
Covington, Ky. . . 39 05 N 84 31 W
Crescent City, Calif. . . . . 41 45 N 124 12 W
Cripple Creek, Colo. . . . . 38 45 N 105 11 W
Crystal City, Tex. . . . . 28 41 N 99 50 W
Dalhart, Tex. . . . . 36 04 N 102 31 W
Dallas, Tex. . . . . 32 47 N 96 49 W
Dalton, Ga. . . . . 34 46 N 84 58 W
Danville, Ill. . . . . 40 08 N 87 37 W
Danville, Va. . . . . 36 36 N 79 23 W
Davenport, Ia. . . . 41 32 N 90 35 W
Davis, Calif. . . . . 38 33 N 121 44 W
Dayton, Ohio . . . . 39 45 N 84 12 W
Daytona Beach, Fla. . . . . 29 13 N 81 01 W
De Kalb, Ill. . . . . 41 56 N 88 46 W
De Ridder, La. . . . 30 51 N 93 17 W
Decatur, Ala. . . . . 34 36 N 86 59 W
Decatur, Ill. . . . . 39 51 N 88 57 W
Decorah, Ia. . . . . 43 18 N 91 48 W
Defiance, Ohio . . 41 17 N 84 22 W
Denver, Colo. . . . 39 44 N 104 59 W
Des Moines, Ia. . . 41 35 N 93 37 W
Detroit, Mich. . . . 42 20 N 83 03 W
Detroit Lakes, Minn. . . . . 46 49 N 95 51 W
Devils Lake, N.D. . 48 07 N 98 52 W
Dickinson, N.D. . . 46 53 N 102 47 W
Dillingham, Alaska . . . . . 59 03 N 158 28 W
Dillon, Mont. . . . . 45 13 N 112 38 W
Dodge City, Kan. . 37 45 N 100 01 W
Dothan, Ala. . . . . 31 13 N 85 24 W
Dover, Del. . . . . 39 10 N 75 32 W
Dover, N.H. . . . . 43 12 N 70 53 W
Dublin, Ga. . . . . 32 32 N 82 54 W
Dubuque, Ia. . . . . 42 30 N 90 41 W
Duluth, Minn. . . . . 46 47 N 92 07 W
Duncan, Okla. . . . 34 30 N 97 57 W
Durango, Colo. . . 37 17 N 107 53 W
Durham, N.C. . . . 36 00 N 78 54 W
Dutch Harbor, Alaska . . . . . 53 53 N 166 32 W
East St. Louis, Ill. . . . . 38 37 N 90 09 W
Eau Claire, Wis. . . 44 49 N 91 30 W
Effingham, Ill. . . . . 39 07 N 88 33 W
El Cajon, Calif. . . 32 48 N 116 58 W
El Dorado, Ark. . . 33 12 N 92 40 W
El Paso, Tex. . . . . 31 45 N 106 29 W
Elk City, Okla. . . . 35 25 N 99 25 W
Elko, Nev. . . . . 40 50 N 115 46 W
Ely, Minn. . . . . 47 55 N 91 51 W
Ely, Nev. . . . . 39 15 N 114 54 W
Emporia, Kan. . . . 38 25 N 96 11 W
Enid, Okla. . . . . 36 24 N 97 53 W
Enterprise, Ore. . . 45 25 N 117 17 W
Erie, Pa. . . . . 42 08 N 80 05 W
Escanaba, Mich. . 45 45 N 87 04 W
Escondido, Calif. . 33 07 N 117 05 W
Eugene, Ore. . . . . 44 05 N 123 04 W
Eunice, La. . . . . 30 30 N 92 25 W
Eureka, Calif. . . . 40 47 N 124 09 W
Eustis, Fla. . . . . 28 51 N 81 41 W
Evanston, Ill. . . . 42 03 N 87 41 W
Evanston, Wyo. . . 41 16 N 110 58 W
Evansville, Ind. . . 37 58 N 87 35 W
Everett, Wash. . . 47 59 N 122 12 W
Fairbanks, Alaska, ... 64 51 N 147 45 W
Falls City, Neb. . . 40 03 N 95 36 W
Fargo, N.D. . . . . 46 53 N 96 48 W
Farmington, N.M. . 36 44 N 108 12 W
Fayetteville, Ark. . 36 03 N 94 09 W
Fayetteville, N.C. . 35 03 N 78 53 W
Fergus Falls, Minn. . . . . 46 17 N 96 04 W
Flagstaff, Ariz. . . . 35 12 N 111 39 W
Flint, Mich. . . . . 43 01 N 83 41 W
Florence, Ala. . . . 34 47 N 87 40 W
Florence, S.C. . . . 34 12 N 79 46 W
Forrest City, Ark. . 35 01 N 90 47 W
Fort Bragg, Calif. . 39 26 N 123 48 W
Fort Collins, Colo. . . . . 40 35 N 105 05 W

Fort Dodge, Ia. . . 42 30 N 94 11 W
Fort Lauderdale, Fla. . . . . 26 07 N 80 08 W
Fort Madison, Ia. . 40 38 N 91 27 W
Fort Myers, Fla. . . 26 39 N 81 53 W
Fort Pierce, Fla. . . 27 26 N 80 19 W
Fort Smith, Ark. . . 35 23 N 94 25 W
Fort Wayne, Ind. . 41 04 N 85 09 W
Fort Worth, Tex. . 32 45 N 97 18 W
Frankfort, Ky. . . . 38 12 N 84 52 W
Frederick, Md. . . . 39 25 N 77 25 W
Fredericksburg, Va. . . . . 38 18 N 77 28 W
Freeport, Ill. . . . . 42 17 N 89 36 W
Freeport, Tex. . . . 28 57 N 95 21 W
Fremont, Neb. . . . 41 26 N 96 30 W
Fresno, Calif. . . . . 36 44 N 119 47 W
Gadsden, Ala. . . . 34 01 N 86 01 W
Gainesville, Fla. . . 29 40 N 82 20 W
Gainesville, Ga. . . 34 17 N 83 49 W
Galena, Alaska . . 64 44 N 156 56 W
Galesburg, Ill. . . . 40 57 N 90 22 W
Gallup, N.M. . . . . 35 31 N 108 45 W
Galveston, Tex. . . 29 18 N 94 48 W
Garden City, Kan. . . . . 37 58 N 100 52 W
Garland, Tex. . . . 32 54 N 96 38 W
Gary, Ind. . . . . 41 36 N 87 20 W
Gatlinburg, Tenn. . . . . 35 43 N 83 31 W
Georgetown, S.C. . . . . 33 23 N 79 17 W
Germantown, Ohio . . . . . 39 38 N 84 22 W
Gillette, Wyo. . . . 44 18 N 105 30 W
Glasgow, Ky. . . . 37 00 N 85 55 W
Glasgow, Mont. . . 48 12 N 106 38 W
Glendive, Mont. . . 47 07 N 104 43 W
Glenwood Springs, Colo. . . 39 33 N 107 19 W
Goliad, Tex. . . . . 28 40 N 97 23 W
Goodland, Kan. . . 39 21 N 101 43 W
Grand Forks, N.D. . . . . 47 55 N 97 03 W
Grand Island, Neb. . . . . 40 55 N 98 21 W
Grand Junction, Colo. . . . . 39 04 N 108 33 W
Grand Rapids, Mich. . . . . 42 58 N 85 40 W
Granite Falls, Minn. . . . . 44 49 N 95 33 W
Great Falls, Mont. . . . . 47 30 N 111 17 W
Greeley, Colo. . . . 40 25 N 104 42 W
Green Bay, Wis. . 44 31 N 88 00 W
Green River, Wyo. . . . . 41 32 N 109 28 W
Greensboro, N.C. . . . . 36 04 N 79 48 W
Greenville, Ala. . . 31 50 N 86 38 W
Greenville, Miss. . 33 24 N 91 04 W
Greenville, S.C. . . 34 51 N 82 23 W
Greenwood, S.C. . 34 12 N 82 10 W
Griffin, Ga. . . . . 33 15 N 84 16 W
Gulfport, Miss. . . 30 22 N 89 06 W
Guymon, Okla. . . 36 41 N 101 29 W
Hagerstown, Md. . 39 39 N 77 43 W
Hammond, Ind. . . 41 38 N 87 30 W
Hammond, La. . . 30 30 N 90 28 W
Hampton, Va. . . . 37 02 N 76 21 W
Hannibal, Mo. . . . 39 42 N 91 22 W
Harlingen, Tex. . . 26 12 N 97 42 W
Harpers Ferry, W.Va. . . . . 39 20 N 77 44 W
Harrisburg, Pa. . . 40 16 N 76 53 W
Hartford, Conn. . . 41 46 N 72 41 W
Hattiesburg, Miss. . . . . 31 20 N 89 17 W
Havre, Mont. . . . . 48 33 N 109 41 W
Hays, Kan. . . . . 38 53 N 99 20 W
Helena, Mont. . . . 46 36 N 112 02 W
Henderson, Nev. . 36 02 N 114 59 W
Hialeah, Fla. . . . . 25 51 N 80 16 W
Hilo, Hawaii . . . . . 19 44 N 155 05 W
Hobbs, N.M. . . . . 32 42 N 103 08 W
Holly Springs, Miss. . . . . 34 46 N 89 27 W
Hollywood, Fla. . . 26 01 N 80 09 W
Honokaa, Hawaii . 20 05 N 155 28 W
Honolulu, Hawaii . 21 19 N 157 52 W
Hope, Ark. . . . . 33 40 N 93 36 W
Hopkinsville, Ky. . 36 52 N 87 29 W
Hot Springs, Ark. . 34 31 N 93 03 W
Hot Springs, S.D. . . . . 43 26 N 103 29 W
Houghton, Mich. . 47 07 N 88 34 W
Houston, Tex. . . . 29 46 N 95 22 W
Hugo, Okla. . . . . 34 01 N 95 31 W
Huntington, W.Va. . . . . 38 25 N 82 27 W
Huntsville, Ala. . . 34 44 N 86 35 W
Hutchinson, Kan. . 38 05 N 97 56 W
Idaho Falls, Idaho . . . . . 43 30 N 112 02 W
Independence, Mo. . . . . 39 05 N 94 24 W
Indianapolis, Ind. . 39 46 N 86 09 W
International Falls, Minn. . . . . 48 36 N 93 25 W
Iowa City, Ia. . . . . 41 40 N 91 32 W
Iron Mountain, Mich. . . . . 45 49 N 88 04 W
Ironwood, Mich. . . 46 27 N 90 09 W
Ithaca, N.Y. . . . . 42 26 N 76 30 W
Jackson, Mich. . . 42 15 N 84 24 W
Jackson, Miss. . . 32 18 N 90 12 W
Jackson, Tenn. . . 35 37 N 88 49 W
Jacksonville, Fla. . 30 20 N 81 39 W
Jacksonville, N.C. . . . . 34 45 N 77 26 W
Jamestown, N.Y. . 42 06 N 79 14 W
Jamestown, N.D. . 46 54 N 98 42 W
Jefferson City, Mo. . . . . 38 34 N 92 10 W
Jersey City, N.J. . 40 44 N 74 04 W
Johnstown, Pa. . . 40 20 N 78 55 W
Joliet, Ill. . . . . 41 32 N 88 05 W
Jonesboro, Ark. . . 35 50 N 90 42 W
Jonesboro, Ga. . . 33 31 N 84 22 W
Junction City, Kan. . . . . 39 02 N 96 50 W
Juneau, Alaska . . 58 20 N 134 27 W
Kailua, Hawaii . . . 21 24 N 157 45 W
Kaktovik, Alaska . 70 08 N 143 38 W
Kalamazoo, Mich. . . . . 42 17 N 85 35 W
Kalispell, Mont. . . 48 12 N 114 19 W
Kankakee, Ill . . . . 41 07 N 87 52 W
Kannapolis, N.C. . 35 30 N 80 37 W
Kansas City, Kan. . . . . 39 07 N 94 38 W
Kansas City, Mo. . 39 06 N 94 35 W
Kapaa, Hawaii . . . 22 05 N 159 19 W
Kearney, Neb. . . . 40 42 N 99 05 W
Kenai, Alaska . . . 60 33 N 151 16 W
Ketchikan, Alaska . . . . . 55 21 N 131 39 W
Key Largo, Fla. . . 25 06 N 80 27 W
Key West, Fla. . . . 24 33 N 81 49 W
King City, Calif. . . 36 13 N 121 08 W
Kingman, Ariz. . . 35 12 N 114 04 W
Kingsville, Tex. . . 27 31 N 97 52 W
Kirksville, Mo. . . . 40 12 N 92 35 W
Klamath Falls, Ore. . . . . 42 12 N 121 46 W
Knoxville, Tenn. . . 35 58 N 83 55 W
Kodiak, Alaska . . 57 47 N 152 24 W
Kokomo, Ind. . . . . 40 30 N 86 08 W
La Crosse, Wis. . . 43 48 N 91 15 W
La Junta, Colo. . . 37 59 N 103 33 W
Lafayette, La. . . . 30 14 N 92 01 W
Lake Charles, La. . . . . 30 14 N 93 13 W
Lake Havasu City, Ariz. . . . . 34 29 N 114 19 W
Lakeland, Fla. . . . 28 03 N 81 57 W
Lamoni, Ia. . . . . 40 37 N 93 56 W
Lancaster, Ohio . . 39 43 N 82 36 W
Lancaster, Pa. . . . 40 02 N 76 19 W
Lansing, Mich. . . . 42 44 N 84 33 W
Laramie, Wyo. . . . 41 19 N 105 35 W
Laredo, Tex. . . . . 27 30 N 99 30 W
Largo, Fla. . . . . 27 54 N 82 47 W
Las Cruces, N.M. . . . . 32 19 N 106 47 W
Las Vegas, Nev. . 36 01 N 115 09 W
Las Vegas, N.M. . 35 36 N 105 13 W
Laurel, Miss. . . . . 31 41 N 89 08 W
Lawton, Okla. . . . 34 37 N 98 25 W
Lebanon, N.H. . . . 43 39 N 72 15 W
Lewiston, Idaho . . 46 25 N 117 01 W
Lewiston, Me. . . . 44 06 N 70 13 W
Lewiston, Mont. . . 47 03 N 109 25 W
Lexington, Ky. . . . 38 01 N 84 30 W
Liberal, Kan. . . . . 37 02 N 100 55 W
Lihue, Hawaii . . . 21 59 N 159 23 W
Lima, Ohio . . . . . 40 44 N 84 06 W
Lincoln, Me. . . . . 45 22 N 68 30 W
Lincoln, Neb. . . . . 40 50 N 96 41 W
Little Rock, Ark. . 34 45 N 92 17 W
Logan, Utah . . . . 41 44 N 111 50 W
Logansport, Ind. . 40 45 N 86 22 W
Lompoc, Calif. . . . 34 38 N 120 28 W
Long Beach, Calif. . . . . 33 47 N 118 11 W
Longview, Tex. . . 32 30 N 94 44 W
Lorain, Ohio . . . . 41 28 N 82 11 W
Los Alamos, N.M. . . . . 35 53 N 106 19 W
Los Angeles, Calif. . . . . 34 04 N 118 15 W
Los Banos, Calif. . 37 04 N 120 51 W
Louisville, Ky. . . . 38 15 N 85 46 W
Lowell, Mass. . . . 42 38 N 71 19 W
Lubbock, Tex. . . . 33 35 N 101 51 W
Lynchburg, Va. . . 37 25 N 79 09 W
McAlester, Okla. . 34 56 N 95 46 W
McAllen, Tex. . . . 26 12 N 98 14 W
McCall, Idaho . . . 44 55 N 116 06 W
McCook, Neb. . . . 40 12 N 100 38 W
Macomb, Ill. . . . . 40 27 N 90 40 W
Macon, Ga. . . . . 32 51 N 83 38 W
Madison, Wis. . . . 43 04 N 89 24 W
Manchester, N.H. . . . . 43 00 N 71 28 W
Mandan, N.D. . . . 46 50 N 100 54 W
Manitowoc, Wis. . 44 05 N 87 39 W
Mankato, Minn. . . 44 10 N 94 00 W
Marietta, Ohio . . . 39 25 N 81 27 W
Marinette, Wis. . . 45 06 N 87 38 W
Marion, Ind. . . . . 40 32 N 85 40 W
Marion, Ohio . . . . 40 35 N 83 08 W
Marquette, Mich. . 46 33 N 87 24 W
Marshalltown, Ia. . 42 02 N 92 54 W
Martinsville, Va. . . 36 41 N 79 52 W
Marysville, Calif. . 39 09 N 121 35 W
Massillon, Ohio . . 40 48 N 81 32 W
Mattoon, Ill. . . . . 39 28 N 88 22 W
Medford, Ore. . . . 42 19 N 122 52 W
Meeker, Colo. . . . 40 02 N 107 55 W
Melbourne, Fla. . . 28 05 N 80 37 W
Memphis, Tenn. . 35 08 N 90 03 W
Meridian, Miss. . . 32 22 N 88 42 W
Mesa, Ariz. . . . . 33 25 N 111 49 W
Miami, Fla. . . . . 25 47 N 80 11 W
Miami Beach, Fla. . . . . 25 47 N 80 08 W
Middlebury, Vt. . . 44 01 N 73 10 W
Midland, Mich. . . . 43 36 N 84 14 W
Midland, Tex. . . . 32 00 N 102 05 W
Miles City, Mont. . 46 25 N 105 51 W
Milledgeville, Ga. . . . . 33 05 N 83 14 W
Milwaukee, Wis. . 43 02 N 87 55 W
Minden, La. . . . . 32 37 N 93 17 W
Minneapolis, Minn. . . . . 44 59 N 93 16 W
Minot, N.D. . . . . 48 14 N 101 18 W
Missoula, Mont. . . 46 52 N 114 01 W
Mitchell, S.D. . . . 43 43 N 98 02 W
Moab, Utah . . . . . 38 35 N 109 33 W
Mobile, Ala. . . . . 30 41 N 88 03 W
Mobridge, S.D. . . 45 32 N 100 26 W
Modesto, Calif. . . 37 39 N 121 00 W
Moline, Ill. . . . . 41 30 N 90 31 W
Monroe, La. . . . . 32 30 N 92 07 W
Monroe, Mich. . . . 41 55 N 83 24 W
Monterey, Calif. . . 36 37 N 121 55 W
Montgomery, Ala. . . . . 32 23 N 86 19 W
Montpelier, Vt. . . . 44 16 N 72 35 W
Montrose, Colo. . . 38 29 N 107 53 W
Morehead City, N.C. . . . . 34 43 N 76 43 W
Morgan City, La. . 29 42 N 91 12 W
Morgantown, W.Va. . . . . 39 38 N 79 57 W
Moscow, Idaho . . 46 44 N 117 00 W
Moses Lake, Wash. . . . . 47 08 N 119 17 W
Mount Pleasant, Mich. . . . . 43 36 N 84 46 W
Mount Vernon, Ill. . . . . 38 19 N 88 55 W
Mount Vernon, Wash. . . . . 48 25 N 122 20 W
Muncie, Ind. . . . . 40 12 N 85 23 W
Murfreesboro, Ark. . . . . 34 04 N 93 41 W
Murfreesboro, Tenn. . . . . 35 50 N 86 23 W
Muscle Shoals, Ala. . . . . 34 45 N 87 40 W
Muskegon, Mich. . 43 14 N 86 16 W
Muskogee, Okla. . 35 45 N 95 22 W
Myrtle Beach, S.C. . . . . 33 42 N 78 53 W
Napa, Calif. . . . . 38 18 N 122 17 W
Naples, Fla. . . . . 26 08 N 81 48 W
Nashville-Davidson, Tenn. . . . . 36 10 N 86 47 W
Natchez, Miss. . . 31 34 N 91 24 W
Natchitoches, La. . . . . 31 46 N 93 05 W
Needles, Calif. . . . 34 51 N 114 37 W
Nevada, Mo. . . . . 37 51 N 94 22 W
New Albany, Ind. . 38 18 N 85 49 W
New Bedford, Mass. . . . . 41 38 N 70 56 W
New Bern, N.C. . . 35 07 N 77 03 W
New Harmony, Ind. . . . . 38 08 N 87 56 W
New Haven, Conn. . . . . 41 18 N 72 55 W
New Iberia, La. . . 30 01 N 91 49 W
New Madrid, Mo. . 36 36 N 89 32 W
New Orleans, La. . . . . 29 58 N 90 04 W
New Rochelle, N.Y. . . . . 40 55 N 73 47 W
New York City, N.Y. . . . . 40 43 N 74 00 W
Newark, N.J. . . . . 40 44 N 74 10 W
Newcastle, Wyo. . 43 50 N 104 11 W
Newport, Ore. . . . 44 39 N 124 03 W
Newport, R.I. . . . . 41 29 N 71 18 W
Newport News, Va. . . . . 36 59 N 76 25 W
Niagara Falls, N.Y. . . . . 43 06 N 79 03 W
Nogales, Ariz. . . . 31 20 N 110 56 W
Nome, Alaska . . . 64 30 N 165 25 W
Norfolk, Va. . . . . 36 51 N 76 17 W
Norman, Okla. . . . 35 13 N 97 26 W
North Augusta, S.C. . . . . 33 30 N 81 59 W
North Platte, Neb. . . . . 41 08 N 100 46 W
O'Neill, Neb. . . . . 42 27 N 98 39 W
Oakland, Calif. . . 37 49 N 122 16 W
Ocala, Fla. . . . . 29 11 N 82 08 W
Ocean City, Md. . 38 20 N 75 05 W
Oceanside, Calif. . 33 12 N 117 23 W
Odessa, Tex. . . . 31 52 N 102 23 W
Ogallala, Neb. . . . 41 08 N 101 43 W
Ogden, Utah . . . . 41 13 N 111 58 W
Ogdensburg, N.Y. . . . . 44 42 N 75 29 W
Oklahoma City, Okla. . . . . 35 30 N 97 30 W
Olympia, Wash. . . 47 03 N 122 53 W
Omaha, Neb. . . . . 41 17 N 96 01 W
Oneida, N.Y. . . . . 43 06 N 75 39 W
Orem, Utah . . . . . 40 18 N 111 42 W
Orlando, Fla. . . . . 28 33 N 81 23 W
Oshkosh, Wis. . . . 44 01 N 88 33 W
Ottawa, Ill. . . . . 41 20 N 88 50 W
Ottumwa, Ia. . . . . 41 01 N 92 25 W
Overton, Nev. . . . 36 33 N 114 27 W
Owensboro, Ky. . 37 46 N 87 07 W
Paducah, Ky. . . . 37 05 N 88 37 W
Pahala, Hawaii . . 19 12 N 155 29 W
Palm Springs, Calif. . . . . 33 50 N 116 33 W
Palo Alto, Calif. . . 37 27 N 122 10 W
Panama City, Fla. . . . . 30 10 N 85 40 W
Paragould, Ark. . . 36 03 N 90 29 W
Paris, Tex. . . . . 33 40 N 95 33 W
Parkersburg, W.Va. . . . . 39 16 N 81 34 W
Parsons, Kan. . . . 37 20 N 95 16 W
Pasadena, Calif. . 34 09 N 118 09 W
Pasadena, Tex. . . 29 43 N 95 13 W
Pascagoula, Miss. . . . . 30 21 N 88 33 W
Pasco, Wash. . . . 46 14 N 119 06 W
Paterson, N.J. . . . 40 55 N 74 11 W
Pecos, Tex. . . . . 31 26 N 103 30 W
Pendleton, Ore. . . 45 40 N 118 47 W
Pensacola, Fla. . . 30 25 N 87 13 W
Peoria, Ill. . . . . 40 42 N 89 36 W
Perryville, Ky. . . . 37 39 N 84 57 W
Petersburg, Va. . . 37 14 N 77 24 W

Petoskey, Mich. . . 45 22 N 84 57 W
Philadelphia, Pa. . . . . 39 57 N 75 10 W
Phoenix, Ariz. . . . 33 27 N 112 04 W
Pierre, S.D. . . . . 44 22 N 100 21 W
Pine Bluff, Ark. . . 34 13 N 92 01 W
Pittsburgh, Pa. . . 40 26 N 80 01 W
Plano, Tex. . . . . 33 01 N 96 41 W
Plattsburgh, N.Y. . 44 42 N 73 27 W
Pocatello, Idaho . 42 52 N 112 27 W
Point Hope, Alaska . . . . . 68 21 N 166 41 W
Pontiac, Mich. . . . 42 38 N 83 18 W
Port Angeles, Wash. . . . . 48 07 N 123 27 W
Port Arthur, Tex. . 29 54 N 93 56 W
Port Gibson, Miss. . . . . 31 58 N 90 59 W
Port Huron, Mich. . . . . 42 58 N 82 26 W
Port Lavaca, Tex. . . . . 28 37 N 96 38 W
Port Royal, S.C. . . 32 23 N 80 42 W
Portage, Wis. . . . 43 33 N 89 28 W
Portland, Me. . . . 43 39 N 70 16 W
Portland, Ore. . . . 45 32 N 122 37 W
Portsmouth, N.H. . . . . 43 05 N 70 45 W
Portsmouth, Ohio . . . . . 38 44 N 83 00 W
Portsmouth, Va. . 36 50 N 76 18 W
Poughkeepsie, N.Y. . . . . 41 42 N 73 56 W
Prairie du Chien, Wis. . . . . 43 03 N 91 09 W
Prattville, Ala. . . . 32 28 N 86 29 W
Prescott, Ariz. . . . 34 33 N 112 28 W
Presque Isle, Me. . . . . 46 41 N 68 01 W
Providence, R.I. . . 41 49 N 71 24 W
Provo, Utah . . . . . 40 14 N 111 39 W
Pueblo, Colo. . . . 38 15 N 104 36 W
Pullman, Wash. . . 46 44 N 117 10 W
Quincy, Ill. . . . . 39 56 N 91 23 W
Racine, Wis. . . . . 42 44 N 87 48 W
Raleigh, N.C. . . . . 35 46 N 78 38 W
Rapid City, S.D. . 44 05 N 103 14 W
Rawlins, Wyo. . . . 41 47 N 107 14 W
Red Bluff, Calif. . . 40 11 N 122 15 W
Redding, Calif. . . 40 35 N 122 24 W
Redfield, S.D. . . . 44 53 N 98 31 W
Reno, Nev. . . . . 39 31 N 119 48 W
Rice Lake, Wis. . . 45 30 N 91 44 W
Richfield, Utah . . 38 46 N 112 05 W
Richmond, Ind. . . 39 50 N 84 54 W
Richmond, Va. . . 37 33 N 77 27 W
Rio Grande City, Tex. . . . . 26 23 N 98 49 W
Riverside, Calif. . . 33 59 N 117 22 W
Riverton, Wyo. . . . 43 02 N 108 23 W
Roanoke, Va. . . . 37 16 N 79 56 W
Rochester, Minn. . 44 01 N 92 28 W
Rochester, N.Y. . . 43 10 N 77 37 W
Rockford, Ill. . . . . 42 16 N 89 06 W
Rock Hill, S.C. . . . 34 56 N 81 01 W
Rock Island, Ill. . . 41 30 N 90 34 W
Rock Springs, Wyo. . . . . 41 35 N 109 12 W
Rolla, Mo. . . . . 37 57 N 91 46 W
Rome, Ga. . . . . 34 15 N 85 09 W
Roseburg, Ore. . . 43 13 N 123 20 W
Roswell, N.M. . . . 33 24 N 104 32 W
Rugby, N.D. . . . . 48 22 N 100 00 W
Russellville, Ark. . 35 16 N 93 08 W
Sacramento, Calif. . . . . 38 35 N 121 29 W
Saginaw, Mich. . . 43 26 N 83 56 W
St. Albans, Vt. . . . 44 49 N 73 05 W
St. Augustine, Fla. . . . . 29 54 N 81 19 W
St. Cloud, Minn. . 45 34 N 94 10 W
St. George, Utah . 37 06 N 113 35 W
St. Joseph, Mo. . . 39 46 N 94 50 W
St. Louis, Mo. . . . 38 37 N 90 11 W
St. Maries, Idaho .47 19 N 116 35 W
St. Paul, Minn. . . . 44 57 N 93 06 W
St. Petersburg, Fla. . . . . 27 46 N 82 39 W
Ste. Genevieve, Mo. . . . . 37 59 N 90 03 W
Salem, Ore. . . . . 44 56 N 123 02 W
Salina, Kan. . . . . 38 50 N 97 37 W
Salinas, Calif. . . . 36 40 N 121 39 W
Salmon, Idaho . . . 45 11 N 113 54 W
Salt Lake City, Utah . . . . . 40 45 N 111 53 W
San Angelo, Tex. . 31 28 N 100 26 W
San Antonio, Tex. . . . . 29 25 N 98 30 W
San Bernardino, Calif. . . . . 34 07 N 117 19 W
San Clemente, Calif. . . . . 33 25 N 117 36 W
San Diego, Calif. . 32 43 N 117 09 W
San Francisco, Calif. . . . . 37 47 N 122 25 W
San Jose, Calif. . . 37 20 N 121 53 W
San Luis Obispo, Calif. . . . . 35 17 N 120 40 W
Sanderson, Tex. . 30 09 N 102 24 W
Santa Ana, Calif. . 33 46 N 117 52 W
Santa Barbara, Calif. . . . . 34 25 N 119 42 W
Santa Cruz, Calif. . . . . 36 58 N 122 01 W
Santa Fe, N.M. . . 35 41 N 105 57 W
Santa Maria, Calif. . . . . 34 57 N 120 26 W
Santa Rosa, Calif. . . . . 38 26 N 122 43 W
Sarasota, Fla. . . . 27 20 N 82 32 W
Sault Sainte Marie, Mich. . . . 46 30 N 84 21 W
Savannah, Ga. . . 32 05 N 81 06 W
Schenectady, N.Y. . . . . 42 50 N 73 57 W
Scott City, Kan. . . 38 29 N 100 54 W
Scottsbluff, Neb. . 41 52 N 103 40 W
Scottsdale, Ariz. . 33 29 N 111 56 W
Scranton, Pa. . . . 41 25 N 75 40 W
Searcy, Ark. . . . . 35 15 N 91 44 W
Seattle, Wash. . . . 47 36 N 122 20 W
Sebring, Fla. . . . . 27 30 N 81 27 W
Seguin, Tex. . . . . 29 34 N 97 58 W
Selawik, Alaska . . 66 36 N 160 00 W
Seldovia, Alaska . 59 26 N 151 43 W
Selma, Ala. . . . . 32 25 N 87 01 W
Sharpsburg, Md. . 39 28 N 77 45 W
Sheboygan, Wis. . 43 45 N 87 42 W
Shelbyville, Tenn. . . . . 35 29 N 86 28 W
Sheridan, Wyo. . . 44 48 N 106 58 W
Show Low, Ariz. . 34 15 N 110 02 W
Shreveport, La. . . 32 31 N 93 45 W
Sierra Vista, Ariz. . . . . 31 33 N 110 18 W
Silver City, N.M. . 32 46 N 108 17 W
Sioux City, Ia. . . . 42 30 N 96 24 W
Sioux Falls, S.D. . 43 33 N 96 44 W
Skagway, Alaska . 59 28 N 135 19 W
Snyder, Tex. . . . . 32 44 N 100 55 W
Socorro, N.M. . . . 34 04 N 106 54 W
Somerset, Ky. . . . 37 05 N 84 36 W
South Bend, Ind. . 41 41 N 86 15 W
Sparks, Nev. . . . . 39 32 N 119 45 W
Spartanburg, S.C. . . . . 34 56 N 81 57 W
Spearfish, S.D. . . 44 30 N 103 52 W
Spencer, Ia. . . . . 43 09 N 95 10 W
Spokane, Wash. . 47 40 N 117 24 W
Springfield, Ill. . . . 39 48 N 89 38 W
Springfield, Mass. . . . . 42 06 N 72 35 W
Springfield, Mo. . . 37 13 N 93 17 W
State College, Pa. . . . . 40 48 N 77 52 W
Steamboat Springs, Colo. . . 40 29 N 106 50 W
Sterling, Colo. . . . 40 37 N 103 13 W
Steubenville, Ohio . . . . . 40 22 N 80 37 W
Stevens Point, Wis. . . . . 44 31 N 89 34 W
Stillwater, Minn. . . 45 03 N 92 49 W
Sulphur Springs, Tex. . . . . 33 08 N 95 36 W
Sumter, S.C. . . . . 33 55 N 80 21 W
Sun Valley, Idaho . . . . . 43 42 N 114 21 W
Superior, Wis. . . . 46 44 N 92 06 W
Syracuse, N.Y. . . 43 03 N 76 09 W
Tacoma, Wash. . . 47 14 N 122 26 W
Tallahassee, Fla. . 30 27 N 84 17 W
Tampa, Fla. . . . . 27 57 N 82 27 W
Tempe, Ariz. . . . . 33 25 N 111 56 W
Temple, Tex. . . . . 31 06 N 97 21 W
Terre Haute, Ind. . 39 28 N 87 25 W
Texarkana, Ark. . . 33 26 N 94 03 W
Thermopolis, Wyo. . . . . 43 39 N 108 13 W
Thief River Falls, Minn. . . . . 48 07 N 96 10 W
Tifton, Ga. . . . . 31 27 N 83 31 W
Titusville, Fla. . . . 28 37 N 80 49 W
Toledo, Ohio . . . . 41 39 N 83 33 W
Tonopah, Nev. . . . 38 04 N 117 14 W
Topeka, Kan. . . . . 39 03 N 95 40 W
Traverse City, Mich. . . . . 44 46 N 85 38 W
Trenton, N.J. . . . . 40 14 N 74 46 W
Trinidad, Colo. . . . 37 10 N 104 31 W
Troy, Ala. . . . . 31 48 N 85 58 W
Troy, N.Y. . . . . 42 44 N 73 41 W
Tucson, Ariz. . . . . 32 13 N 110 58 W
Tulsa, Okla. . . . . 36 10 N 95 55 W
Tupelo, Miss. . . . . 34 16 N 88 43 W
Tuscaloosa, Ala. . 33 12 N 87 34 W
Twin Falls, Idaho . 42 34 N 114 28 W
Tyler, Tex. . . . . 32 21 N 95 18 W
Ukiah, Calif. . . . . 39 09 N 123 12 W
Utica, N.Y. . . . . 43 06 N 75 14 W
Uvalde, Tex. . . . . 29 13 N 99 47 W
Valdez, Alaska . . 61 07 N 146 16 W
Valdosta, Ga. . . . 30 50 N 83 17 W
Valentine, Neb. . . 42 52 N 100 33 W
Vandalia, Ill. . . . . 38 58 N 89 06 W
Vernal, Utah . . . . 40 27 N 109 31 W
Vero Beach, Fla. . . . . 27 38 N 80 24 W
Vicksburg, Miss. . 32 21 N 90 53 W
Victoria, Tex. . . . . 28 48 N 97 00 W
Vincennes, Ind. . . 38 41 N 87 32 W
Virginia, Minn. . . . 47 31 N 92 32 W
Virginia Beach, Va. . . . . 36 51 N 75 59 W
Waco, Tex. . . . . 31 33 N 97 09 W
Wahpeton, N.D. . . 46 15 N 96 36 W
Wailuku, Hawaii . . 20 53 N 156 30 W
Walla Walla, Wash. . . . . 46 04 N 118 20 W
Waltham, Mass. . 42 23 N 71 14 W
Warner Robins, Ga. . . . . 32 37 N 83 36 W
Warren, Pa. . . . . 41 51 N 79 08 W
Washington, D.C. . . . . 38 54 N 77 02 W
Waterbury, Conn. . . . . 41 33 N 73 03 W
Waterloo, Ia. . . . . 42 30 N 92 21 W
Watertown, N.Y. . 43 59 N 75 55 W
Waterville, Me. . . 44 33 N 69 38 W
Waukegan, Wis. . 43 01 N 88 14 W
Waukesha, Wis. . 44 58 N 89 38 W
Waycross, Ga. . . 31 13 N 82 21 W
Wayne, Neb. . . . . 42 14 N 97 01 W
Weiser, Idaho . . . 44 45 N 116 58 W
West Palm Beach, Fla. . . . 26 43 N 80 03 W
West Plains, Mo. . 36 44 N 91 51 W
Wheeling, W.Va. . 40 04 N 80 43 W
Wichita, Kan. . . . . 37 42 N 97 20 W
Wichita Falls, Tex. . . . . 33 54 N 98 30 W
Wilkes-Barre, Pa. . 41 15 N 75 53 W
Williamsburg, Va. . . . . 37 16 N 76 43 W
Williamsport, Pa. . 41 15 N 77 00 W
Williston, N.D. . . . 48 09 N 103 37 W
Willmar, Minn. . . . 45 07 N 95 03 W
Wilmington, N.C. . 34 14 N 77 55 W
Winfield, Kan. . . . 37 15 N 96 59 W
Winnemucca, Nev. . . . . 40 58 N 117 44 W
Winslow, Ariz. . . . 35 02 N 110 42 W
Winston-Salem, N.C. . . . . 36 06 N 80 14 W
Worcester, Mass. . . . . 42 16 N 71 48 W
Worthington, Minn. . . . . 43 37 N 95 36 W
Wrangell, Alaska . . . . . 56 28 N 132 23 W
Yakima, Wash. . . 46 36 N 120 31 W
Yankton, S.D. . . . 42 53 N 97 23 W
Yazoo City, Miss. . . . . 32 51 N 90 25 W
York, Pa. . . . . 39 58 N 76 44 W
Youngstown, Ohio . . . . . 41 06 N 80 39 W
Yuba City, Calif. . 39 08 N 121 37 W
Yuma, Ariz. . . . . 32 43 N 114 37 W
Zanesville, Ohio . 39 56 N 82 01 W

Physical features and points of interest

Absaroka Range . 44 30 N 110 00 W
Adak Island . . . . . 51 45 N 176 40 W
Adirondack Mountains . . . . . 44 00 N 74 00 W
Alabama, *river* . . . 31 08 N 87 57 W
Alaska, Gulf of . . 58 00 N 145 00 W
Alaska Range . . . 59 45 N 156 00 W
Aleutian Islands . . 52 55 N 172 28 W
Alexander Archipelago, *islands* . . . . . 58 20 N 136 00 W
American Falls Reservoir . . . . . 42 47 N 112 52 W
Amistad Reservoir . . . . . 29 30 N 101 05 W
Appalachian Mountains . . . . . 38 00 N 80 00 W
Arctic Ocean . . . . 74 00 N 148 00 W
Arkansas, *river* . . 33 47 N 91 04 W
Atka Island . . . . . 52 07 N 174 30 W
Atlantic Ocean . . 35 00 N 72 00 W
Attu Island . . . . . 52 55 N 172 55 W
Baffin Bay . . . . . 27 15 N 97 31 W
Bering Sea . . . . . 60 00 N 175 00 W
Bering Strait . . . . 66 00 N 169 00 W
Big Bend National Park . . . 29 15 N 103 30 W
Big Black, *river* . . 32 03 N 91 04 W
Big Sandy, *river* . 34 19 N 113 32 W
Big Sioux, *river* . . 42 29 N 96 27 W
Big Stone Lake . . 45 18 N 96 27 W
Bighorn, *river* . . . 46 10 N 107 28 W
Bighorn Mountains . . . . . 44 30 N 107 15 W
Bitterroot Range . 46 00 N 114 30 W
Black, *river* . . . . . 35 38 N 91 20 W
Black Hills . . . . . 44 33 N 104 29 W
Blackfeet Indian Reservation . . . . 48 30 N 113 00 W
Black Mesa, *peak* . . . . . 37 05 N 103 10 W
Blanco, Cape . . . 42 50 N 124 34 W
Blue Mountains . . 44 43 N 118 32 W
Boeuf, *river* . . . . . 33 13 N 91 16 W
Borah Peak . . . . . 44 08 N 113 48 W
Boston Mountains . . . . . 35 45 N 93 30 W
Boundary Peak . . 37 51 N 118 21 W
Brazos, *river* . . . . 28 53 N 95 23 W
Bristol Bay . . . . . 58 00 N 159 00 W
Broad, *river* . . . . . 34 01 N 81 04 W
Brooks Range . . . 68 00 N 154 00 W
California Aqueduct . . . . . 35 34 N 119 40 W
Canadian, *river* . . 35 28 N 95 03 W
Canaveral, Cape . 28 28 N 80 32 W
Canyonlands National Park . . . 38 15 N 110 00 W
Carson Sink, *dry lake bed* . . . . . 39 56 N 118 19 W
Cedar, *river* . . . . . 41 17 N 91 21 W
Champlain, Lake . 44 33 N 73 21 W
Chattahoochee, *river* . . . . . 30 54 N 84 57 W
Chesuncook, Lake . . . . . 46 00 N 69 20 W
Cheyenne River Indian Reservation . . . . 45 05 N 101 20 W
Chippewa, *river* . . 44 56 N 95 44 W
Cimarron, *river* . . 36 08 N 96 31 W
Clark Fork, *river* . . 48 09 N 116 15 W
Clarks Hill Lake, *reservoir* . . . . . 33 40 N 82 12 W
Clinch, *river* . . . . . 35 53 N 84 29 W
Coast Ranges . . . 41 00 N 123 30 W
Cod, Cape . . . . . 41 45 N 70 00 W
Colorado, *river* . . 40 28 N 105 50 W
Colorado, *river* . . 28 36 N 95 58 W
Colorado Desert . 33 00 N 116 00 W
Colorado Plateau . 36 30 N 108 00 W
Colorado River Indian Reservation . . . . 34 00 N 114 25 W
Columbia, *river* . . 46 15 N 124 03 W
Colville Indian Reservation . . . . 48 15 N 119 00 W
Conchas Lake, *reservoir* . . . . . 35 23 N 104 11 W
Coosa, *river* . . . . 32 30 N 86 16 W
Crater Lake National Park . . . 42 49 N 122 08 W

Crow Indian Reservation . . . . 45 30 N 108 00 W
Cumberland Plateau . . . . . 35 57 N 84 41 W
Davis Mountains . 30 35 N 104 00 W
Death Valley . . . . 36 15 N 116 50 W
Denali National Park and Preserve . . . . . 63 15 N 150 30 W
Disappointment, Cape . . . . . 46 18 N 124 03 W
Dixon Entrance . . 54 25 N 132 30 W
Duck Valley Indian Reservation . . . . 42 00 N 116 10 W
Edisto, *river* . . . . . 32 29 N 80 21 W
Edwards Plateau . 30 45 N 101 25 W
Elbert, Mount . . . 39 07 N 106 27 W
Elephant Butte Reservoir . . . . . 33 09 N 107 11 W
Erie, Lake . . . . . 42 00 N 81 00 W
Erie Canal . . . . . 43 07 N 76 26 W
Eufaula Lake, *reservoir* . . . . . 35 15 N 95 35 W
Everglades National Park . . . 25 30 N 81 00 W
Falcon Reservoir . 26 34 N 99 10 W
Fear, Cape . . . . . 33 50 N 77 57 W
Finger Lakes . . . . 43 00 N 76 45 W
Flaming Gorge Reservoir . . . . . 40 55 N 109 25 W
Flathead Indian Reservation . . . . 47 30 N 114 30 W
Flathead Lake . . . 47 51 N 114 08 W
Florida, Straits of . . . . . 25 00 N 79 45 W
Florida Keys, *islands* . . . . . 24 45 N 81 00 W
Fort Apache Indian Reservation . . . . 33 35 N 110 10 W
Fort Belknap Indian Reservation . . . . 48 15 N 108 30 W
Fort Berthold Indian Reservation . . . . 47 40 N 102 25 W
Fort Hall Indian Reservation . . . . 43 00 N 112 30 W
Fort Peck Indian Reservation . . . . 48 15 N 105 30 W
Fort Peck Lake . . 47 45 N 106 30 W
Fort Totten Indian Reservation . . . . 47 53 N 98 50 W
Francis Case, Lake, *reservoir* . . 43 04 N 98 34 W
Franklin D. Roosevelt Lake, *reservoir* . . . . . 47 57 N 118 59 W
Friant Kern Canal . . . . . 36 12 N 119 03 W
Front Range . . . . 40 48 N 106 40 W
Gallatin, *river* . . . . 45 56 N 111 30 W
Gannett Peak . . . 43 11 N 109 39 W
Gates of the Arctic National Park and Preserve . . . . . 67 45 N 153 30 W
George, Lake . . . 29 17 N 81 36 W
Gila, *river* . . . . . 32 43 N 114 33 W
Glacier National Park . . . . . 48 40 N 114 00 W
Glen Canyon Dam . . . . . 36 57 N 111 29 W
Glendo Reservoir . . . . . 42 29 N 104 57 W
Grand, *river* . . . . . 39 23 N 93 07 W
Grand, *river* . . . . . 45 40 N 100 45 W
Grand Canyon . . . 36 07 N 114 00 W
Grand Canyon National Park . . . 36 03 N 112 08 W
Grand Coulee Dam . . . . . 47 57 N 118 59 W
Grand Teton . . . . 43 44 N 110 48 W
Grand Teton National Park . . . 43 45 N 110 45 W
Great Basin . . . . . 39 00 N 116 30 W
Great Dismal Swamp . . . . . 36 35 N 76 30 W
Great Divide Basin . . . . . 42 00 N 108 00 W
Great Plains . . . . 40 00 N 102 00 W
Great Salt Lake . . 41 10 N 112 30 W
Great Salt Lake Desert . . . . . 40 39 N 113 32 W
Great Smoky Mountains . . . . . 35 30 N 83 00 W
Green, *river* . . . . . 37 54 N 87 30 W
Green, *river* . . . . . 38 11 N 109 53 W
Green Bay . . . . . 45 02 N 87 28 W
Guadalupe Peak . 31 50 N 104 52 W
Harney Basin . . . 43 15 N 119 00 W
Harry S. Truman Reservoir . . . . . 38 16 N 93 24 W
Hatteras, Cape . . 35 14 N 75 32 W
Hawaii, *island* . . . 19 30 N 155 30 W
Hawaii Volcanoes National Park . . . 19 23 N 155 17 W
Hells Canyon . . . . 45 22 N 116 38 W
Holston, *river* . . . . 35 58 N 83 51 W
Hood, Mount . . . . 45 23 N 121 42 W
Hoover Dam . . . . 36 01 N 114 44 W
Hopi Indian Reservation . . . . 35 47 N 110 30 W
Hudson, *river* . . . 40 42 N 74 02 W
Humboldt, *river* . . 39 59 N 118 32 W
Huron, Lake . . . . 44 30 N 82 15 W
Illinois, *river* . . . . . 38 58 N 90 28 W
James, *river* . . . . 42 52 N 97 18 W
Jicarilla Apache Indian Reservation . . . . 36 40 N 107 00 W
John Day, *river* . . 45 44 N 120 39 W
John Day Lock and Dam . . . . . 45 43 N 120 42 W
Juan de Fuca, Strait of . . . . . 48 12 N 123 34 W
Kaskaskia, *river* . . 37 58 N 89 57 W
Katmai National Park and Preserve . . . . . 58 30 N 155 00 W
Kauai, *island* . . . . 22 03 N 159 30 W
Kentucky Lake . . 37 01 N 88 16 W
Kings Canyon National Park . . . 36 48 N 118 30 W
Kiska Island . . . . 51 59 N 177 30 W
Klamath, *river* . . . 41 33 N 124 05 W
Klamath Mountains . . . . . 41 40 N 123 20 W
Kodiak Island . . . 57 20 N 153 22 W
Kootenai, *river* . . . 45 40 N 115 20 W
Lanai, *island* . . . . 20 50 N 156 55 W
Laramie Mountains . . . . . 41 31 N 105 30 W
Leech Lake Indian Reservation . . . . 47 30 N 94 27 W
Lewis Range . . . . 48 30 N 113 30 W
Lewis Smith Lake, *reservoir* . . 33 56 N 87 06 W
Licking, *river* . . . . 39 06 N 84 30 W
Livingston, Lake, *reservoir* . . . . . 30 40 N 95 10 W
Llano Estacado . . 33 00 N 102 30 W
Long Island . . . . . 40 49 N 73 04 W
Loup, *river* . . . . . 41 24 N 97 19 W
Lower Red Lake . 47 58 N 95 00 W
Madison, *river* . . . 45 56 N 111 31 W
Martin, Lake, *reservoir* . . . . . 32 41 N 85 55 W
Matagorda Bay . . 28 32 N 96 20 W
Maui, *island* . . . . 20 48 N 156 20 W
Maumee, *river* . . . 41 42 N 83 28 W
Mauna Kea, *volcano* . . . . . 19 50 N 155 28 W
Mauna Loa, *volcano* . . . . . 19 30 N 155 35 W
McKinley, Mount . 63 04 N 151 00 W
Mead, Lake . . . . . 36 01 N 114 44 W
Medicine Bow Mountains . . . . . 40 56 N 106 11 W
Memphremagog, Lake . . . . . 45 00 N 72 15 W
Mesabi Range . . . 47 30 N 93 00 W
Mescalero Apache Indian Reservation . . . . 33 15 N 105 30 W
Mexico, Gulf of . . 27 30 N 90 00 W
Michigan, Lake . . 44 00 N 87 00 W
Middle Loup, *river* . . . . . 41 17 N 98 24 W
Milk, *river* . . . . . 48 05 N 106 15 W
Mille Lacs Lake . . 46 15 N 93 39 W
Mississippi, *river* . 29 09 N 89 15 W
Missouri, *river* . . . 38 49 N 90 07 W
Mitchell, Mount . . 35 50 N 82 15 W
Mojave Desert . . . 34 57 N 117 12 W
Molokai, *island* . . 21 08 N 157 00 W
Mount Rainier National Park . . . 46 50 N 121 45 W
Mount St. Helens National Volcanic Monument . . . . . 46 15 N 122 10 W
Musselshell, *river* . 47 21 N 107 57 W
Navajo Indian Reservation . . . . 36 30 N 109 00 W
Neosho, *river* . . . . 36 48 N 95 18 W
Nez Perce Indian Reservation . . . . 46 30 N 116 30 W
Niihau, *island* . . . 21 54 N 160 09 W
Niobrara, *river* . . . 42 46 N 98 03 W
North Anna, *river* . 37 48 N 77 24 W
North Canadian, *river* . . . . . 35 16 N 95 31 W
North Cascades National Park . . . 48 30 N 121 00 W
North Loup, *river* . 41 16 N 98 24 W
North Platte, *river* . . . . . 41 07 N 100 42 W
Northern Cheyenne Indian Reservation . . . . 45 31 N 106 45 W
Oahe, Lake, *reservoir* . . . . . 45 38 N 100 35 W
Oahu, *island* . . . . 21 28 N 157 58 W
Ohio, *river* . . . . . 36 59 N 89 08 W
Okeechobee, Lake . . . . . 26 56 N 80 49 W
Okefenokee Swamp . . . . . 30 40 N 82 20 W
Olympic Mountains . . . . . 48 02 N 123 41 W
Olympic National Park . . . . . 47 48 N 123 30 W
Ontario, Lake . . . 43 45 N 78 00 W
Osage, *river* . . . . 38 36 N 92 57 W
Osage Indian Reservation . . . . 36 30 N 96 30 W
Ouachita, *river* . . 34 08 N 96 36 W
Ouachita Mountains . . . . . 34 30 N 94 30 W
Owyhee, *river* . . . 43 46 N 117 02 W
Ozark Plateau . . . 36 30 N 92 30 W
Ozarks, Lake of the, *reservoir* . . . 38 12 N 92 38 W
Pacific Ocean . . . 35 00 N 123 00 W
Papago Indian Reservation . . . . 32 08 N 112 05 W
Pearl, *river* . . . . . 30 11 N 89 32 W
Pecos, *river* . . . . . 29 42 N 101 22 W
Pee Dee, *river* . . . 33 22 N 79 16 W
Pend Oreille, Lake, *reservoir* . . 48 10 N 116 20 W
Pine Mountain . . . 37 04 N 82 49 W
Pine Ridge Indian Reservation . . . . 43 25 N 102 21 W
Platte, *river* . . . . . 41 04 N 95 53 W
Porcupine, *river* . . 66 35 N 145 19 W
Powder, *river* . . . . 46 44 N 105 06 W
Powder, *river* . . . . 44 45 N 117 10 W
Powell, Lake, *reservoir* . . . . . 36 56 N 111 29 W
Puget Sound . . . . 47 50 N 122 26 W
Pyramid Lake Indian Reservation . . . . 40 20 N 119 35 W
Rainier, Mount . . . 46 51 N 121 46 W
Rainy Lake . . . . . 48 38 N 93 00 W
Red, *river* . . . . . 31 01 N 91 44 W
Republican, *river* . 39 04 N 96 48 W
Reyes, Point . . . . 37 59 N 123 01 W
Rio Grande, *river* . . . . . 25 57 N 97 10 W
Roanoke, *river* . . . 35 57 N 76 42 W
Rock, *river* . . . . . 41 29 N 90 37 W
Rocky Mountain National Park . . . 40 19 N 105 42 W
Rocky Mountains . 40 00 N 108 00 W
Rosebud Indian Reservation . . . . 43 25 N 100 28 W
Sabine, *river* . . . . 29 59 N 93 47 W
Sacramento Mountains . . . . . 33 12 N 105 41 W
St. Clair, Lake . . . 42 27 N 82 39 W
St. Croix, *river* . . . 45 04 N 67 05 W
St. Croix, *river* . . . 44 45 N 92 48 W
St. Elias, Mount . . 60 18 N 140 56 W
St. Francis, *river* . 34 38 N 90 36 W
St. Helens, Mount . . . . . 46 12 N 122 12 W
St. John, *river* . . . 47 05 N 67 48 W
St. Lawrence, *river* . . . . . 44 30 N 75 30 W
St. Lawrence Island . . . . . 63 30 N 170 30 W
Sakakawea, Lake *reservoir* . . . . . 47 30 N 101 26 W
Saline, *river* . . . . . 38 52 N 97 30 W
Salmon, *river* . . . . 45 51 N 116 47 W
Sam Rayburn Reservoir . . . . . 31 04 N 94 05 W
San Carlos Indian Reservation . . . . 33 23 N 110 06 W
San Juan, *river* . . 37 16 N 110 26 W
San Juan Mountains . . . . . 37 29 N 106 49 W
Sangre de Cristo Mountains . . . . . 37 00 N 105 30 W
Santee, *river* . . . . 33 07 N 79 17 W
Sardis Lake, *reservoir* . . . . . 34 25 N 89 48 W
Savannah, *river* . . 32 02 N 80 53 W
Sequoia National Park . . . . . 36 34 N 118 46 W
Seward Peninsula . . . . . 65 00 N 164 00 W
Shasta, Mount . . . 41 25 N 122 12 W
Shishaldin Volcano . . . . . 54 45 N 163 58 W
Sierra Nevada, *mountains* . . . . . 37 42 N 119 19 W
Sisseton Indian Reservation . . . . 45 40 N 97 02 W
Smoky Hill, *river* . 39 04 N 96 48 W
Snake, *river* . . . . . 46 12 N 119 02 W
Sonoran Desert . . 32 55 N 112 40 W
South Platte, *river* . . . . . 41 07 N 100 42 W
Spokane, *river* . . . 47 54 N 118 20 W
Standing Rock Indian Reservation . . . . 45 50 N 101 10 W
Stikine, *river* . . . . 56 31 N 132 24 W
Sunflower, Mount . . . . . 39 04 N 102 01 W
Superior, Lake . . . 47 30 N 88 00 W
Suwannee, *river* . 29 18 N 83 09 W
Table Rock Lake, *reservoir* . . . . . 36 36 N 93 19 W
Tahoe, Lake . . . . 39 06 N 120 02 W
Tallapoosa, *river* . 32 30 N 86 16 W
Tanana, *river* . . . . 65 10 N 151 58 W
Teton Range . . . . 44 00 N 111 00 W
Texoma, Lake, *reservoir* . . . . . 33 50 N 96 40 W
Timms Hill . . . . . 45 27 N 90 11 W
Toledo Bend Reservoir . . . . . 31 11 N 93 34 W
Tombigbee, *river* . 31 08 N 87 57 W
Tongue, *river* . . . . 46 25 N 105 52 W
Trinidad Head, *point* . . . . . 41 03 N 124 09 W
Uinta Mountains . 40 53 N 109 18 W
Uintah and Ouray Indian Reservation . . . . 40 15 N 110 15 W
Umnak Island . . . 53 15 N 168 20 W
Unimak Island . . . 54 45 N 164 00 W
Upper Red Lake . 48 08 N 94 45 W
Venetie Indian Reservation . . . . 66 01 N 146 25 W
Wabash, *river* . . . 37 46 N 88 02 W
Waialeale, Mount . 22 04 N 159 30 W
Walker River Indian Reservation . . . . 39 00 N 118 40 W
Warm Springs Indian Reservation . . . . 45 00 N 121 25 W
Wasatch Range . 41 04 N 111 51 W
Wheeler Peak . . . 38 59 N 114 19 W
White, *river* . . . . . 33 57 N 91 05 W
White, *river* . . . . . 43 42 N 99 27 W
White Butte, *peak* . . . . . 46 23 N 103 18 W
White Earth Indian Reservation . . . . 47 18 N 95 50 W
White Mountains . 37 34 N 118 14 W
Whitefish Bay . . . 46 29 N 84 49 W
Whitney, Mount . . 36 35 N 118 18 W
Wind River Indian Reservation . . . . 43 30 N 109 00 W

Wind River Range	43 00 N	109 30 W
Winnebago, Lake	44 00 N	88 26 W
Wolf, *river*	44 11 N	88 48 W
Woods, Lake of the	49 15 N	94 45 W
Yellowstone, *river*	47 59 N	103 59 W
Yellowstone National Park	44 30 N	110 30 W
Yosemite National Park	37 45 N	119 35 W
Yukon, *river*	62 32 N	163 54 W
Zion National Park	37 20 N	113 00 W
Zuni Indian Reservation	35 15 N	108 20 W

decades and helps explain the huge concentration of heavy industry along the lower Great Lakes.

The Great Plains

The western flanks of the Interior Lowlands are the Great Plains, a territory of awesome bulk that spans the full distance between Canada and Mexico in a swath nearly 500 miles wide. The Great Plains were built by successive layers of poorly cemented sand, silt, and gravel—debris laid down by parallel east-flowing streams from the Rocky Mountains. Seen from the east, the surface of the Great Plains rises inexorably from about 2,000 feet near Omaha, Neb., to more than 6,000 feet at Cheyenne, Wyo., but the climb is so gradual that popular legend holds the Great Plains to be flat. True flatness is rare, although the High Plains of western Texas, Oklahoma, Kansas, and eastern Colorado come close. More commonly, the land is broadly rolling, and parts of the northern plains are sharply dissected into badlands.

The main mineral wealth of the Interior Lowlands derives from fossil fuels. Coal occurs in structural basins protected from erosion—high-quality bituminous in the Appalachian, Illinois, and western Kentucky basins; and subbituminous and lignite in the eastern and northwestern Great Plains. Petroleum and natural gas have been found in nearly every state between the Appalachians and the Rockies, but the Midcontinent Fields of western Texas and the Texas Panhandle, Oklahoma, and Kansas surpass all others. Aside from small deposits of lead and zinc, metallic minerals are of little importance.

The Appalachian Mountain system. The Appalachians dominate the eastern United States and separate the Eastern Seaboard from the interior with a belt of subdued uplands that extends nearly 1,500 miles from northeastern Alabama to the Canadian border. They are old, complex mountains, the eroded stumps of much greater ranges. Present topography results from erosion that has carved weak rocks away, leaving a skeleton of resistant rocks behind as highlands. Geologic differences are thus faithfully reflected in topography. In the Appalachians these differences are sharply demarcated and neatly arranged, so that all the major subdivisions except New England lie in strips parallel to the Atlantic and to one another.

The core of the Appalachians is a belt of complex metamorphic and igneous rocks that stretches all the way from Alabama to New Hampshire. The western side of this belt forms the long slender rampart of the Blue Ridge Mountains, containing the highest elevations in the Appalachians (Mount Mitchell, N.C., 6,684 feet [2,037 metres]) and some of its most handsome mountain scenery. On its eastern, or seaward, side the Blue Ridge descends in an abrupt and sometimes spectacular escarpment to the Piedmont, a well-drained, rolling land—never quite hills, but never quite a plain. Before the settlement of the Midwest the Piedmont was the most productive agricultural region in the United States, and several Pennsylvania counties still consistently report some of the highest farm yields per acre in the entire country.

The Piedmont

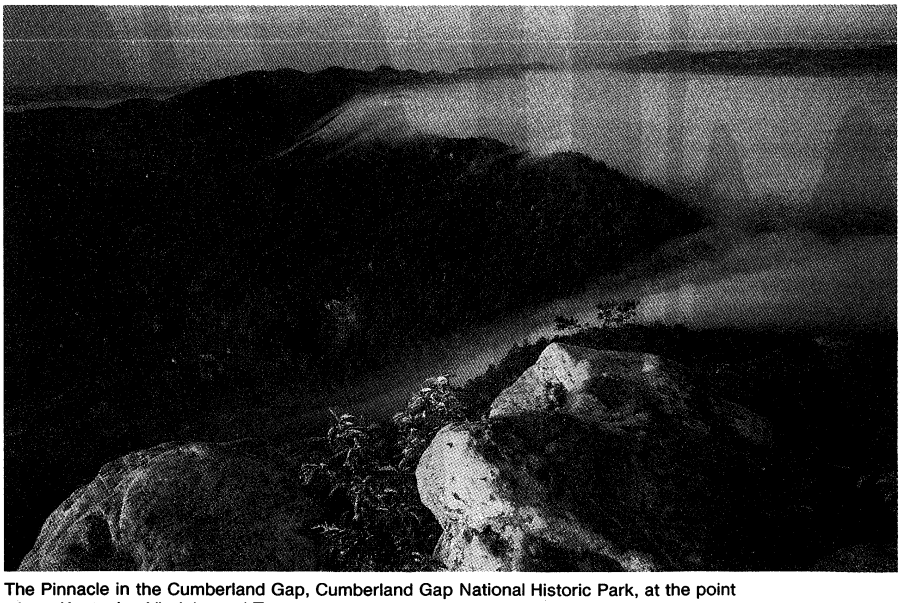
West of the crystalline zone, away from the axis of primary geologic deformation, sedimentary rocks have escaped metamorphism but are compressed into tight folds. Erosion has carved the upturned edges of these folded rocks into the remarkable Ridge and Valley country of the western Appalachians. Long linear ridges characteristically stand about 1,000 feet from base to crest and run for tens of miles, paralleled by broad open valleys of comparable length. In Pennsylvania, ridges run unbroken for great distances, occasionally turning abruptly in a zigzag pattern; by contrast, the southern ridges are broken by faults and form short, parallel segments that are lined up like magnetized iron filings. By far the largest valley—and one of the most important routes in North America—is the Great Valley, an extraordinary trench of shale and limestone that runs nearly the entire length of the Appalachians. It provides a lowland passage from the middle Hudson valley to Harrisburg, Pa., and on southward, where it forms the Shenandoah and Cumberland valleys, and has been one of the main paths through the Appalachians since pioneer times. In New England it is floored with slates and marbles and forms the Valley of Vermont, one of the few fertile areas in an otherwise mountainous region.

Topography much like that of the Ridge and Valley is found in the Ouachita Mountains of western Arkansas and eastern Oklahoma, an area generally thought to be a detached continuation of Appalachian geologic structure, the intervening section buried beneath the sediments of the lower Mississippi valley.

The New England mountain systems

The once-glaciated New England section of the Appalachians is divided from the rest of the chain by an indentation of the Atlantic. Although almost completely underlain by crystalline rocks, New England is laid out in north–south bands, reminiscent of the southern Appalachians. The rolling, rocky hills of southeastern New England are not dissimilar to the Piedmont, while, farther northwest, the rugged and lofty White Mountains are a New England

D. Muench—H. Armstrong Roberts



The Pinnacle in the Cumberland Gap, Cumberland Gap National Historic Park, at the point where Kentucky, Virginia, and Tennessee meet.

analogue to the Blue Ridge. (Mount Washington, N.H., at 6,288 feet [1917 metres], is the highest peak in the northeastern United States.) The westernmost ranges—the Taconics, Berkshires, and Green Mountains—show a strong north–south lineation like the Ridge and Valley. Unlike the rest of the Appalachians, however, glaciation has scoured the crystalline rocks much like those of the Canadian Shield, so that New England is best known for its picturesque landscape, not for its fertile soil.

Typical of diverse geologic regions, the Appalachians contain a great variety of minerals. Only a few occur in quantities large enough for sustained exploitation, notably iron in Pennsylvania's Blue Ridge and Piedmont and the famous granites, marbles, and slates of northern New England. In Pennsylvania the Ridge and Valley region contains one of the world's largest deposits of anthracite coal, once the basis of a thriving mining economy; many of the mines are now shut, oil and gas having replaced coal as the major fuel used to heat homes.

The Atlantic Plain. The eastern and southeastern fringes of the United States are part of the outermost margins of the continental platform, repeatedly invaded by the sea and veneered with layer after layer of young, poorly consolidated sediments. Part of this platform now lies slightly above sea level and forms a nearly flat and often swampy coastal plain, which stretches from Cape Cod, Mass., to beyond the Mexican border. Most of the platform, however, is still submerged, so that a band of shallow water, the continental shelf, parallels the Atlantic and Gulf coasts, in some places reaching 250 miles out to sea.

Coastal uplifts and depressions

The Atlantic Plain slopes so gently that even slight crustal upwarping can shift the coastline far out to sea at the expense of the continental shelf. The peninsula of Florida is just such an upwarp; nowhere in its 400-mile length does the land rise more than 350 feet above sea level; much of the southern and coastal areas rise less than 10 feet and are poorly drained and dangerously exposed to Atlantic storms. Downwarps can result in extensive flooding. North of New York City, for example, the weight of glacial ice depressed most of the Coastal Plain beneath the sea, and the Atlantic now beats directly against New England's rock-ribbed coasts. Cape Cod, Long Island (N.Y.), and a few offshore islands are all that remain of New England's drowned Coastal Plain. Another downwarp lies perpendicular to the Gulf coast and guides the course of the lower Mississippi. The river, however, has filled with alluvium what otherwise would be an arm of the Gulf, forming a great inland salient of the Coastal Plain called the Mississippi Embayment.

South of New York the Coastal Plain gradually widens, but ocean water has invaded the lower valleys of most of the coastal rivers and has turned them into estuaries. The greatest of these is Chesapeake Bay, merely the flooded lower valley of the Susquehanna River and its tributaries, but there are hundreds of others. Offshore a line of sandbars and barrier beaches stretches intermittently the length of the Coastal Plain, hampering entry of shipping into the estuaries but providing the eastern United States with a playground that is more than 1,000 miles long.

Poor soils are the rule on the Coastal Plain, though rare exceptions have formed some of America's most famous agricultural regions—for example, the citrus country of central Florida's limestone uplands and the Cotton Belt of the Old South, once centred on the alluvial plain of the Mississippi and belts of chalky black soils of eastern Texas, Alabama, and Mississippi. The Atlantic Plain's greatest natural wealth derives from petroleum and natural gas trapped in domal structures that dot the Gulf Coast of eastern Texas and Louisiana. Onshore and offshore drilling have revealed colossal reserves of oil and natural gas.

The Western Cordillera. West of the Great Plains the United States seems to become a craggy land whose skyline is rarely without mountains—totally different from the open plains and rounded hills of the East. On a map the alignment of the two main chains—the Rocky Mountains on the east, the Pacific ranges on the west—tempts one to assume a geologic and hence topographic homogeneity. Nothing could be farther from the truth, for each chain is divided into widely disparate sections.

Diversity of the Rocky Mountain chains

The Rockies are typically diverse. The Southern Rockies are composed of a disconnected series of lofty elongated upwarps, their cores made of granitic basement rocks, stripped of sediments, and heavily glaciated at high elevations. In New Mexico and along the western flanks of the Colorado ranges, widespread volcanism and deformation of colourful sedimentary rocks have produced rugged and picturesque country, but the characteristic central Colorado or southern Wyoming range is impressively austere rather than spectacular. The Front Range west of Denver is prototypical, rising abruptly from its base at about 6,000 feet to rolling alpine meadows between 11,000 and 12,000 feet. Peaks appear as low hills perched on this high-level surface, so that Colorado, for example, boasts 53 mountains over 14,000 feet but not one over 14,500 feet.

The Middle Rockies cover most of west central Wyoming. Most of the ranges resemble the granitic upwarps of Colorado, but thrust faulting and volcanism have produced varied and spectacular country to the west, some of which is included in Grand Teton and Yellowstone national parks. Much of the subregion, however, is not mountainous at all but consists of extensive intermontane basins and plains—largely floored with enormous volumes of sedimentary waste eroded from the mountains themselves. Whole ranges have been buried, producing the greatest gap in the Cordilleran system, the Wyoming Basin—resembling in geologic structure and topography an intermontane peninsula of the Great Plains. As a result, the Rockies have never posed an important barrier to east–west transportation in the United States; all major routes, from the Oregon Trail to interstate highways, funnel through the basin, essentially circumventing the main ranges of the Rockies.

The Northern Rockies contain the most varied mountain landscapes of the Cordillera, reflecting a corresponding geologic complexity. The region's backbone is a mighty series of batholiths—huge masses of molten rock that slowly cooled below the surface and were later uplifted. The batholiths are eroded into rugged granitic ranges, which, in central Idaho, compose the most extensive wilderness country in the coterminous United States. East of the batholiths and opposite the Great Plains, sediments have been folded and thrust-faulted into a series of linear north–south ranges, a southern extension of the spectacular Canadian Rockies. Although elevations run 2,000 to 3,000 feet lower than the Colorado Rockies (most of the Idaho Rockies lie well below 10,000 feet), increased rainfall and northern latitude have encouraged glaciation—there as elsewhere a sculptor of handsome alpine landscape.

The Pacific mountain chains

The western branch of the Cordillera directly abuts the Pacific Ocean. This coastal chain, like its Rocky Mountain cousins on the eastern flank of the Cordillera, conceals bewildering complexity behind a facade of apparent simplicity. At first glance the chain consists merely of two lines of mountains with a discontinuous trough between them. Immediately behind the coast is a line of hills and low mountains—the Pacific Coast Ranges. Farther inland, averaging 150 miles from the coast, the line of the Sierra Nevada and the Cascade Range includes the highest elevations in the coterminous United States. Between these two unequal mountain lines is a discontinuous trench, the Troughs of the Coastal Margin.

The apparent simplicity disappears under the most cursory examination. The Pacific Coast Ranges actually contain five distinct sections, each of different geologic origin and each with its own distinctive topography. The Transverse Ranges of southern California are a crowded assemblage of islandlike faulted ranges, with peak elevations of more than 10,000 feet but sufficiently separated by plains and low passes so that travel through them is easy. From Point Conception to the Oregon border, however, the main California Coast Ranges are entirely different, resembling the Appalachian Ridge and Valley region, with low linear ranges that result from erosion of faulted and folded rocks. Major faults run parallel to the low ridges, and the greatest—the notorious San Andreas Fault—was responsible for the earthquake that all but destroyed San Francisco in 1906. Along the California–Oregon border, everything changes again. In this region, the wildly rugged

Klamath Mountains represent a western salient of interior structure reminiscent of the Idaho Rockies and the northern Sierra Nevada. In western Oregon and southwestern Washington the Coast Ranges are also different—a gentle, hilly land carved by streams from a broad arch of marine deposits interbedded with tabular lavas. In the northernmost part of the Coast Ranges and the remote northwest, a domal upwarp has produced the Olympic Mountains; its serrated peaks tower nearly 8,000 feet above Puget Sound and the Pacific, and the heavy precipitation on its upper slopes supports the largest active glaciers in the United States outside of Alaska.

East of these Pacific Coast Ranges the Troughs of the Coastal Margin contain the only extensive lowland plains of the Pacific margin—California's Central Valley, Oregon's Willamette River valley, and the half-drowned basin of Puget Sound in Washington. Parts of an inland trench that extends for great distances along the east coast of the Pacific, similar valleys occur in such diverse areas as Chile and the Alaska panhandle. These valleys are blessed with superior soils, easily irrigated, and very accessible from the Pacific. They have enticed settlers for more than a century and have become the main centres of population and economic activity for much of the U.S. West Coast.

The Cascades and Sierras

Still farther east rise the two highest mountain chains in the coterminous United States—the Cascades and the Sierra Nevada. Aside from elevation, geographic continuity, and spectacular scenery, however, the two ranges differ in almost every important respect. Except for its northern section, where sedimentary and metamorphic rocks occur, the Sierra Nevada is largely made of granite, part of the same batholithic chain that creates the Idaho Rockies. The range is grossly asymmetrical, the result of massive faulting that has gently tilted the western slopes toward the Central Valley but has uplifted the eastern side to confront the interior with an escarpment nearly two miles high. At high elevation glaciers have scoured the granites to a gleaming white, while on the west the ice has carved spectacular valleys such as the Yosemite. The loftiest peak in the Sierras is Mount Whitney, which at 14,494 feet (4,418 metres) is the highest mountain in the coterminous states. The upfaulting that produced Mount Whitney is accompanied by downfaulting that formed nearby Death Valley, at 282 feet (86 metres) below sea level the lowest point in North America.

The Cascades are made largely of volcanic rock; those in northern Washington contain granite like the Sierras, but the rest are formed from relatively recent lava outpourings of dun-coloured basalt and andesite. The Cascades are in effect two ranges. The lower, older range is a long belt of upwarped lava, rising unspectacularly to elevations between 6,000 and 8,000 feet. Perched above the "low Cascades" is a chain of lofty volcanoes that punctuate the horizon with magnificent glacier-clad peaks. The highest is Mount Rainier, which at 14,410 feet (4,392 metres) is all the more dramatic for rising from near sea level. Most of these volcanoes are quiescent, but they are far from extinct. Mount Lassen in northern California erupted violently in 1914, as did Mount St. Helens in the state of Washington in 1980. Most of the other high Cascade volcanoes exhibit some sign of seismic activity.

The Western Intermontane Region. The Cordillera's two main chains enclose a vast intermontane region of arid basins, plateaus, and isolated mountain ranges that stretches from the Mexican border nearly to Canada and extends 600 miles from east to west. This enormous territory contains three huge subregions, each with a distinctive geologic history and its own striking topography.

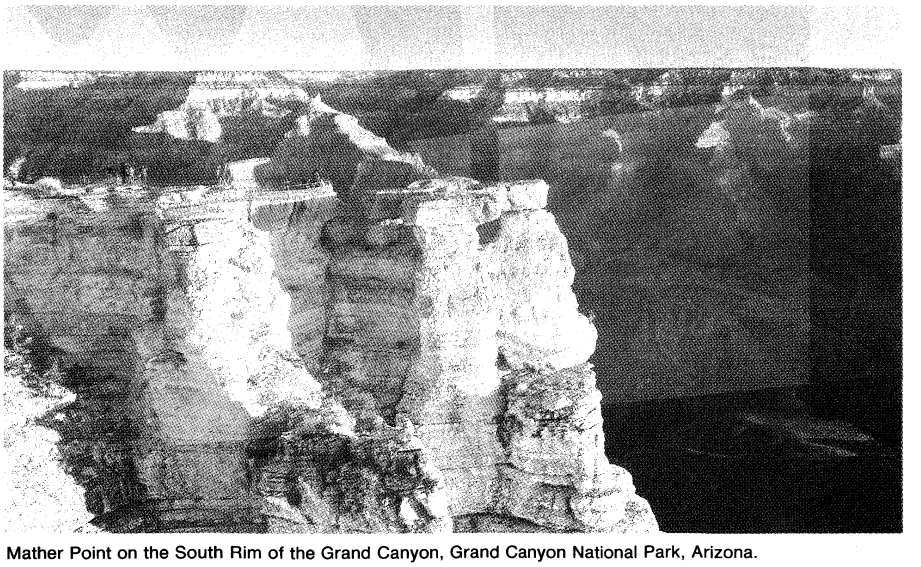
The Colorado Plateau

The Colorado Plateau, nestled against the western flanks of the Southern Rockies, is an extraordinary island of geologic stability set in the turbulent sea of Cordilleran tectonic activity. Stability was not absolute, of course, so that parts of the plateau are warped and injected with volcanics, but in general the landscape results from the erosion by streams of nearly flat-lying sedimentary rocks. The result is a mosaic of angular mesas, buttes, and steplike canyons intricately cut from rocks that often are vividly coloured. Large areas of the plateau are so improbably picturesque that they have been set aside as national preserves. The Grand Canyon of the Colorado River is the most famous of several dozen such areas.

West of the plateau and abutting the Sierra Nevada's eastern escarpment lies the arid Basin and Range subregion, among the most remarkable topographic provinces of the United States. The Basin and Range extends from southern Oregon and Idaho into northern Mexico. Rocks of great complexity have been broken by faulting, and the resulting blocks have tumbled, eroded, and been partly buried by lava and alluvial debris accumulating in the desert basins. The eroded blocks form mountain ranges that are characteristically dozens of miles long, several thousand feet from base to crest, with peak elevations that rarely rise to more than 10,000 feet, and almost always aligned roughly north–south. The basin floors are typically alluvium and sometimes salt marshes or alkali flats.

The third intermontane region, the Columbia Basin, is literally the last, for in some parts its rocks are still being formed. Its entire area is underlain by innumerable tabular lava flows that have flooded the basin between the Cascades and Northern Rockies to undetermined depths. The volume of lava must be measured in thousands of cubic miles, for the flows blanket large parts of Washington, Oregon, and Idaho and in southern Idaho have drowned the flanks of the Northern Rocky Mountains in a basaltic sea. Where the lavas are fresh, as in southern Idaho, the surface is often nearly flat, but more often the floors have been trenched by rivers—conspicuously the Columbia and the Snake—or by glacial floodwaters that

E. Cooper—H. Armstrong Roberts



Mather Point on the South Rim of the Grand Canyon, Grand Canyon National Park, Arizona.

have carved an intricate system of braided canyons in the remarkable Channeled Scablands of eastern Washington. In surface form the eroded lava often resembles the topography of the Colorado Plateau, but the gaudy colours of the Colorado are replaced here by the sombre black and rusty brown of weathered basalt.

Minerals of the Cordillera

Most large mountain systems are sources of varied mineral wealth, and the American Cordillera is no exception. Metallic minerals have been taken from most crystalline regions and have furnished the United States with both romance and wealth—the Sierra Nevada gold that provoked the 1849 gold rush, the fabulous silver lodes of western Nevada's Basin and Range, and gold strikes all along the Rocky Mountain chain. Industrial metals, however, are now far more important; copper and lead are among the base metals, and the more exotic molybdenum, vanadium, and cadmium are mainly useful in alloys.

In the Cordillera, as elsewhere, the greatest wealth stems from fuels. Most major basins contain oil and natural gas, conspicuously the Wyoming Basin, the Central Valley of California, and the Los Angeles Basin. The Colorado Plateau, however, has yielded some of the most interesting discoveries—considerable deposits of uranium and colossal occurrences of oil shale. Oil from the shale, however, probably cannot be economically removed without widespread strip-mining and correspondingly large-scale damage to the environment. Wide exploitation of low-sulfur bituminous coal has been initiated in the Four Corners area of the Colorado Plateau, and open-pit mining has already devastated parts of this once-pristine country as completely as it has West Virginia.

DRAINAGE

As befits a nation of continental proportions, the United States has an extraordinary network of rivers and lakes, including some of the largest and most useful in the world. In the humid East they provide an enormous mileage of cheap inland transportation; westward, most rivers and streams are unnavigable but are heavily used for irrigation and power generation. Both East and West, however, traditionally have used lakes and streams as public sewers, and despite efforts to clean them up, most large waterways are laden with vast, poisonous volumes of industrial, agricultural, and human wastes.

The Eastern systems. Chief among U.S. rivers is the Mississippi, which, with its great tributaries, the Ohio and the Missouri, drains most of the midcontinent. The Mississippi is navigable to Minneapolis nearly 1,200 miles by air from the Gulf of Mexico; and along with the Great Lakes–St. Lawrence system it forms the world's greatest network of inland waterways. The Mississippi's eastern branches, chiefly the Ohio and the Tennessee, are also navigable for great distances. From the west, however, many of its numerous Great Plains tributaries are too seasonal and choked with sandbars to be used for shipping. The Missouri, for example, though longer than the Mississippi itself, was essentially without navigation until the mid-20th century, when a combination of dams, locks, and dredging opened the river to barge traffic.

The Mississippi and Great Lakes waters

The Great Lakes–St. Lawrence system, the other half of the midcontinental inland waterway, is connected to the Mississippi–Ohio via Chicago by canals and the Illinois River. The five Great Lakes (four of which are shared with Canada) constitute by far the largest freshwater lake group in the world and carry a larger tonnage of shipping than any other. The three main barriers to navigation—the St. Marys Rapids, at Sault Sainte Marie; Niagara Falls; and the rapids of the St. Lawrence—are all bypassed by locks, whose 27-foot draft lets ocean vessels penetrate 1,300 miles into the continent, as far as Duluth, Minnesota, and Chicago.

The third group of Eastern rivers drains the coastal strip along the Atlantic Ocean and the Gulf of Mexico. Except for the Rio Grande, which rises west of the Rockies and flows about 1,900 circuitous miles to the Gulf, few of these coastal rivers measure more than 300 miles, and most flow in an almost straight line to the sea. Except in glaciated New England and in arid southwestern Texas, most of the larger coastal streams are navigable for some distance.

The Pacific systems. West of the Rockies, nearly all of the rivers are strongly influenced by aridity. In the deserts and steppes of the intermontane basins, most of the scanty runoff disappears into interior basins, only one of which, the Great Salt Lake, holds any substantial volume of water. Aside from a few minor coastal streams, only three large river systems manage to reach the sea—the Columbia, the Colorado, and the San Joaquin–Sacramento system of California's Central Valley. All three of these river systems are exotic: that is, they flow for considerable distances across dry lands from which they receive little water. Both the Columbia and the Colorado have carved awesome gorges, the former through the sombre lavas of the Cascades and the Columbia Basin, the latter through the brilliantly coloured rocks of the Colorado Plateau. These gorges lend themselves to easy damming, and the once-wild Columbia has been turned into a stairway of placid lakes whose waters irrigate the arid plateaus of eastern Washington and power one of the world's largest hydroelectric networks. The Colorado is less extensively developed, and proposals for new dam construction have met fierce opposition from those who want to preserve the spectacular natural beauty of the river's canyon lands.

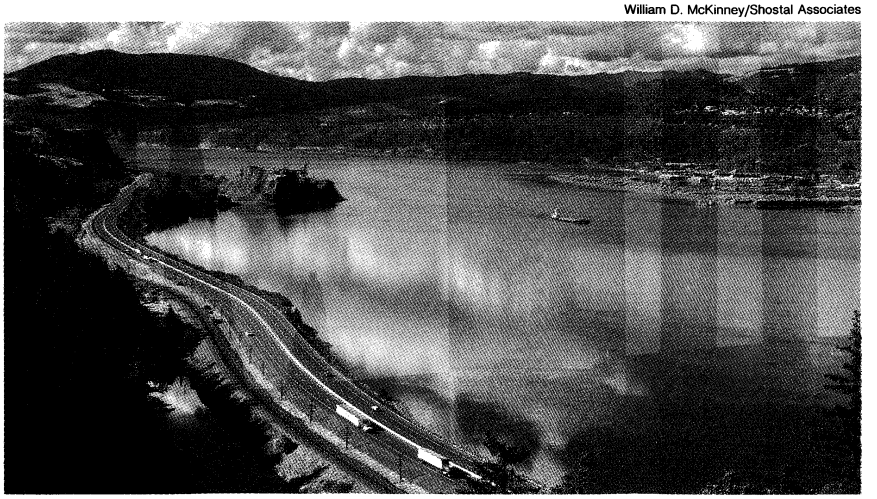
Columbia and Colorado rivers

CLIMATE

Climate affects human habitats both directly and indirectly through its influence on vegetation, soils, and wildlife. In the United States, however, the natural environment has been altered drastically by nearly four centuries of European settlement, as well as thousands of years of Indian occupancy.

Wherever land is abandoned, however, "wild" conditions return rapidly, achieving over the long run a dynamic

William D. McKinney/Shostal Associates



Columbia River, on the Oregon–Washington boundary.

equilibrium among soils, vegetation, and the inexorable strictures of climate. Thus, though Americans have created an artificial environment of continental proportions, the United States still can be divided into a mosaic of bioclimatic regions, each of them distinguished by peculiar climatic conditions and each with a potential vegetation and soil that eventually would return in the absence of humans. The main exception to this generalization applies to fauna, so drastically altered that it is almost impossible to know what sort of animal geography would redevelop in the areas of the United States if humans were removed from the scene.

Climatic controls. The pattern of U.S. climates is largely set by the location of the coterminous United States almost entirely in the middle latitudes, by its position with respect to the continental landmass and its fringing oceans, and by the nation's gross pattern of mountains and lowlands. Each of these geographic controls operates to determine the character of air masses and their changing behaviour from season to season.

Climatic impact of latitude, landforms, and oceans

The coterminous United States lies entirely between the tropic of Cancer and 50° N latitude, a position that confines Arctic climates to the high mountaintops and genuine tropics to a small part of southern Florida. By no means, however, is the climate literally temperate, for the middle latitudes are notorious for extreme variations of temperature and precipitation.

The great size of the North American landmass tends to reinforce these extremes. Since land heats and cools more rapidly than bodies of water, places distant from an ocean tend to have continental climates; that is, they alternate between extremes of hot summers and cold winters, in contrast to the marine climates, which are more equable. Most U.S. climates are markedly continental, the more so because the Cordillera effectively confines the moderating Pacific influence to a narrow strip along the West Coast. Extremes of continentality occur near the centre of the country, and in North Dakota temperatures have ranged between a summer high record of 121° F (49° C) and a winter low of −60° F (−51° C). Moreover, the general eastward drift of air over the United States carries continental temperatures all the way to the Atlantic coast. Bismarck, N.D., for example, has a great annual temperature range. Boston, on the Atlantic but largely exempt from its influence, has a lesser but still-continental range, while San Francisco, which is under strong Pacific influence, has only a small summer–winter differential.

In addition to confining Pacific temperatures to the coastal margin, the Pacific Coast Ranges are high enough to make a local rain shadow in their lee, although the main barrier is the great rampart formed by the Sierra Nevada and Cascade ranges. Rainy on their western slopes and barren on the east, this mountain crest forms one of the sharpest climatic divides in the United States.

The rain shadow continues east to the Rockies, leaving the entire Intermontane Region either arid or semiarid, except where isolated ranges manage to capture leftover moisture at high altitudes. East of the Rockies the westerly drift brings mainly dry air, and as a result, the Great Plains are semiarid. Still farther east, humidity increases owing to the frequent incursion from the south of warm, moist, and unstable air from the Gulf of Mexico, which produces more precipitation in the United States than the Pacific and Atlantic oceans combined.

Although the landforms of the Interior Lowlands have been termed dull, there is nothing dull about their weather conditions. Air from the Gulf of Mexico can flow northward across the Great Plains, uninterrupted by topographical barriers, but continental Canadian air flows south by the same route, and, since these two air masses differ in every important respect, the collisions often produce disturbances of monumental violence. Plainsmen and Midwesterners are accustomed to sudden displays of furious weather—tornadoes, blizzards, hailstorms, precipitous drops and rises in temperature, and a host of other spectacular meteorological displays, sometimes dangerous but seldom boring.

The change of seasons. Most of the United States is marked by sharp differences between winter and summer. In winter, when temperature contrasts between land and water are greatest, huge masses of frigid, dry Canadian air periodically spread far south over the midcontinent, bringing cold, sparkling weather to the interior and generating great cyclonic storms where their leading edges confront the shrunken mass of warm Gulf air to the south. Although such cyclonic activity occurs throughout the year, it is most frequent and intense during the winter, parading eastward out of the Great Plains to bring the Eastern states practically all their winter precipitation. Winter temperatures differ widely, depending largely on latitude. Thus, New Orleans, La., at 30° N latitude, and International Falls, Minn., at 49° N, have respective January temperature averages of 55° F (13° C) and 3° F (−16° C). In the north, therefore, precipitation often comes as snow, often driven by furious winds; farther south, cold rain alternates with sleet and occasional snow. Southern Florida is the only dependably warm part of the East, though "polar outbursts" have been known to bring temperatures below 0° F (−18° C) as far south as Tallahassee. The main uniformity of Eastern weather in wintertime is the expectation of frequent change.

Nationwide seasonal variations

Winter climate on the West Coast is very different. A great spiraling mass of relatively warm, moist air spreads south from the Aleutian Islands of Alaska, its semipermanent front producing gloomy overcast and drizzles that hang over the Pacific Northwest all winter long, occasionally reaching southern California, which receives nearly all of its rain at this time of year. This Pacific air brings mild temperatures along the length of the coast; the average January day in Seattle, Wash., ranges between 33° and 44° F (1° and 7° C) and in Los Angeles between 45° and 64° F (7° and 18° C). In southern California, however, rains are separated by long spells of fair weather, and the whole region is a winter haven for those seeking refuge from less agreeable weather in other parts of the country. The Intermontane Region is similar to the Pacific Coast, but with much less rainfall and a considerably wider range of temperatures.

During the summer there is a reversal of the air masses, and east of the Rockies the change resembles the summer monsoon of Southeast Asia. As the midcontinent heats up, the cold Canadian air mass weakens and retreats, pushed north by an aggressive mass of warm, moist air from the Gulf. The great winter temperature differential between North and South disappears as the hot, soggy blanket spreads from the Gulf coast to the Canadian border. Heat and humidity are naturally most oppressive in the South, but there is little comfort in the more northern latitudes. In Houston, Texas, the temperature on a typical July day reaches 93° F (34° C), with relative humidity averaging near 75 percent, but Minneapolis, Minn., more than 1,000 miles north, is only slightly cooler and less humid.

Effects of summer Gulf air

Since the Gulf air is unstable as well as wet, convectional and frontal summer thunderstorms are endemic east of the Rockies, accounting for a majority of total summer rain. These storms usually drench small areas with short-lived, sometimes violent downpours, so that crops in one Midwestern county may prosper, those in another shrivel in drought, and those in yet another be flattened by hailstones. Relief from the humid heat comes in the northern Midwest from occasional outbursts of cool Canadian air; small but more consistent relief is found downwind from the Great Lakes and at high elevations in the Appalachians. East of the Rockies, however, U.S. summers are distinctly uncomfortable, and air conditioning is viewed as a desirable amenity in most areas.

Again, the Pacific regime is different. The moist Aleutian air retreats northward, to be replaced by mild, stable air from over the subtropical but cool waters of the Pacific, and except in the mountains the Pacific Coast is nearly rainless though often foggy. In the meanwhile, a small but potent mass of dry hot air raises temperatures to blistering levels over much of the intermontane Southwest. In Yuma, Ariz., for example, the normal temperature in July reaches 107° F (42° C), while nearby Death Valley, Calif., holds the national record, 134° F (57° C). During its summer peak this scorching air mass spreads from the Pacific margin as far as Texas on the east and Idaho to the north,

turning the whole interior basin into a summer desert.

Over most of the United States, as in most continental climates, spring and autumn are agreeable but disappointingly brief. Autumn is particularly idyllic in the East, with a romantic Indian summer of ripening corn and brilliantly coloured foliage and of mild days and frosty nights. The shift in dominance between marine and continental air masses, however, spawns furious weather in some regions. Along the Atlantic and Gulf coasts, for example, autumn is the season for hurricanes—the American equivalent of typhoons of the Asian Pacific—which rage northward from the warm tropics to create havoc along the Gulf and Atlantic coasts as far north as New England. The Mississippi valley holds the dubious distinction of recording more tornadoes than any other area on Earth. These violent and often deadly storms usually occur over relatively small areas and are confined largely to spring and early summer.

The bioclimatic regions. Three first-order bioclimatic zones encompass most of the coterminous United States—regions in which climatic conditions are similar enough to dictate similar conditions of mature (zonal) soil and potential climax vegetation (*i.e.,* the assemblage of plants that would grow and reproduce indefinitely given stable climate and average conditions of soil and drainage). These are the Humid East, the Humid Pacific Coast, and the Dry West. In addition, the boundary zone between the Humid East and the Dry West is so large and important that it constitutes a separate region, the Humid–Arid Transition. Finally, because the Western Cordillera contains an intricate mosaic of climatic types, largely determined by local elevation and exposure, it is useful to distinguish the Western Mountain Climate. The first three zones, however, are very diverse and require further breakdown, producing a total of 10 main bioclimatic regions.

Indistinctness of climatic boundaries

For two reasons, the boundaries of these bioclimatic regions are much less distinct than boundaries of landform regions. First, climate varies from year to year, especially in boundary zones, whereas landforms obviously do not. Second, regions of climate, vegetation, and soils coincide generally but sometimes not precisely. Boundaries, therefore, should be interpreted as zonal and transitional, and rarely should be considered as sharp lines in the landscape.

For all of their indistinct boundaries, however, these bioclimatic regions have strong and easily recognized identities. Such regional identity is strongly reinforced when a particular area falls entirely within a single bioclimatic region and at the same time a single landform region. The result—as in the Piedmont South, the central Midwest, or the western Great Plains—is a landscape with an unmistakable regional personality.

The Humid East. The largest and in some ways the most important of the bioclimatic zones, the Humid East was where the Europeans first settled, tamed the land, and adapted to American conditions. In early times almost all of this territory was forested, a fact of central importance in American history that profoundly influenced both soils and wildlife. As in most of the world's humid lands, soluble minerals have been leached from the earth, leaving a great family of soils called pedalfers, rich in relatively insoluble iron and aluminum compounds.

Both forests and soils, however, differ considerably within this vast region. Since rainfall is ample and summers are warm everywhere, the main differences result from the length and severity of winters, which determine the length of the growing season. Winter, obviously, differs according to latitude, so that the Humid East is sliced into four great east–west bands of soils and vegetation, with progressively more amenable winters as one travels southward. These changes occur very gradually, however, and the boundaries therefore are extremely subtle.

Climatic subregions of the Eastern United States

The Sub-Boreal Forest Region is the northernmost of these bands. It is only a small and discontinuous part of the United States, representing the tattered southern fringe of the vast Canadian taiga—a scrubby forest dominated by evergreen needle-leaf species that can endure the ferocious winters and reproduce during the short, erratic summers. Average growing seasons are less than 120 days, though localities in Michigan's Upper Peninsula have recorded frost-free periods lasting as long as 161 days and as short as 76 days. Soils of this region that survived the scour of glaciation are miserably thin podzols—heavily leached, highly acid, and often interrupted by extensive stretches of bog. Most attempts at farming in the region long since have been abandoned.

Farther south lies the Humid Microthermal Zone of milder winters and longer summers. Large broadleaf trees begin to predominate over the evergreens, producing a mixed forest of greater floristic variety and economic value that is famous for its brilliant autumn colours. As the forest grows richer in species, sterile podzols give way to more productive gray-brown podzolic soils, stained and fertilized with humus. Although winters are warmer than in the Sub-Boreal zone, and although the Great Lakes help temper the bitterest cold, January temperatures ordinarily average below freezing, and a winter without a few days of subzero temperatures is uncommon. Everywhere, the ground is solidly frozen and snow covered for several months of the year.

Still farther south are the Humid Subtropics. The region's northern boundary is one of the country's most significant climatic lines: the approximate northern limit of a growing season of 180–200 days, the outer margin of cotton growing, and, hence, of the Old South. Most of the South lies in the Piedmont and Coastal Plain, for higher elevations in the Appalachians cause a peninsula of Northern mixed forest to extend as far south as northern Georgia. The red-brown podzolic soil, once moderately fertile, has been severely damaged by overcropping and burning. Thus much of the region that once sustained a rich, broadleaf-forest flora now supports poor piney woods. Throughout the South, summers are hot, muggy, long, and disagreeable; Dixie's "frosty mornings" bring a welcome respite in winter.

The southern margins of Florida contain the only real tropics in the coterminous United States; it is an area in which frost is almost unknown. Hot, rainy summers alternate with warm and somewhat drier winters, with a secondary rainfall peak during the autumn hurricane season—altogether a typical monsoonal regime. Soils and vegetation are mostly immature, however, since southern Florida rises so slightly above sea level that substantial areas, such as the Everglades, are swampy and often brackish. Peat and sand frequently masquerade as soil, and much of the vegetation is either salt-loving mangrove or sawgrass prairie.

The Humid Pacific Coast. The western humid region differs from its eastern counterpart in so many ways as to be a world apart. Much smaller, it is crammed into a narrow littoral belt to the windward of the Sierra–Cascade summit, dominated by mild Pacific air, and chopped by irregular topography into an intricate mosaic of climatic and biotic habitats. Throughout the region rainfall is extremely seasonal, falling mostly in the winter half of the year. Summers are droughty everywhere, but the main regional differences come from the length of drought—from about two months in humid Seattle, Wash., to nearly five months in semiarid San Diego, Calif.

Western Washington, Oregon, and northern California lie within a zone that climatologists call Marine West Coast. Winters are raw, overcast, and drizzly—not unlike northwestern Europe—with subfreezing temperatures restricted mainly to the mountains, upon which enormous snow accumulations produce local alpine glaciers. Summers, by contrast, are brilliantly cloudless, cool, and frequently foggy along the West Coast and somewhat warmer in the inland valleys. This mild marine climate produces some of the world's greatest forests of enormous straight-boled evergreen trees that furnish the United States with much of its commercial timber. Mature soils are typical of humid midlatitude forestlands, a moderately leached gray-brown podzol.

California's Mediterranean climate

Toward the south, with diminishing coastal rain the moist marine climate gradually gives way to California's tiny but much-publicized Mediterranean regime. Although mountainous topography introduces a bewildering variety of local environments, scanty winter rains are quite inadequate to compensate for the long summer drought, and

much of the region has a distinctly arid character. For much of the year, cool, stable Pacific air dominates the West Coast, bringing San Francisco its famous fogs and Los Angeles its infamous smoggy temperature inversions. Inland, however, summer temperatures reach blistering levels, so that in July, while Los Angeles expects a normal daily maximum of 83° F (28° C), Fresno expects 100° F (38° C) and is climatically a desert. As might be expected, Mediterranean California contains a huge variety of vegetal habitats, but the commonest perhaps is the chaparral, a drought-resistant, scrubby woodland of twisted hard-leafed trees, picturesque but of little economic value. Chaparral is a pyrophytic (fire-loving) vegetation—*i.e.*, under natural conditions its growth and form depend on regular burning. These fires constitute a major environmental hazard in the suburban hills above Los Angeles and San Francisco Bay, especially in autumn, when hot dry Santa Ana winds from the interior regularly convert brush fires into infernos. Soils are similarly varied, but most of them are light in colour and rich in soluble minerals, qualities typical of subarid soils.

The Dry West. In the United States, to speak of dry areas is to speak of the West. It covers an enormous region beyond the dependable reach of moist oceanic air, occupying the entire Intermontane area and sprawling from Canada to Mexico across the western part of the Great Plains. To Americans nurtured in the Humid East, this vast territory across the path of all transcontinental travelers has been harder to tame than any other—and no region has so gripped the national imagination as this fierce and dangerous land.

In the Dry West nothing matters more than water. Thus, though temperatures may differ radically from place to place, the really important regional differences depend overwhelmingly on the degree of aridity, whether an area is extremely dry and hence desert or semiarid and therefore steppe.

The Great American Desert

Americans of the 19th century were preoccupied by the myth of a Great American Desert, which supposedly occupied more than one-third of the entire country. True desert, however, is confined to the Southwest, with patchy outliers elsewhere, all without exception located in the lowland rain shadows of the Cordillera. Vegetation in these desert areas varies between nothing at all (a rare circumstance confined mainly to salt flats and sand dunes) to a low cover of scattered woody scrub and short-lived annuals that burst into flamboyant bloom after rains. Soils are usually thin, light-coloured, and very rich with mineral salts. In some areas wind erosion has removed fine-grained material, leaving behind desert pavement, a barren veneer of broken rock.

Most of the West, however, lies in the semiarid region, in which rainfall is scanty but adequate to support a thin cover of short bunchgrass, commonly alternating with scrubby brush. Here, as in the desert, soils fall into the large family of the pedocals, rich in calcium and other soluble minerals, but in the slightly wetter environments of the West, they are enriched with humus from decomposed grass roots. Under the proper type of management, these chestnut-coloured steppe soils have the potential to be very fertile.

Weather in the West resembles that of other dry regions of the world, often extreme, violent, and reliably unreliable. Rainfall, for example, obeys a cruel natural law: as total precipitation decreases, it becomes more undependable. John Steinbeck's novel *The Grapes of Wrath* describes the problems of a family enticed to the arid frontier of Oklahoma during a wet period only to be driven out by the savage drought of the 1930s that turned the western Great Plains into the great American Dust Bowl. Temperatures in the West also fluctuate convulsively within short periods, and high winds are infamous throughout the region.

The Humid–Arid Transition. East of the Rockies all climatic boundaries are gradational. None, however, is so important or so imperceptibly subtle as the boundary zone that separates the Humid East from the Dry West and that alternates unpredictably between arid and humid conditions from year to year. Stretching approximately from Texas to North Dakota in an ill-defined band between the 95th and 100th meridians, this transitional region deserves separate recognition, partly because of its great size, and partly because of the fine balance between surplus and deficit rainfall, which produces a unique and valuable combination of soils, flora, and fauna. The native vegetation, insofar as it can be reconstructed, was prairie, the legendary sea of tall, deep-rooted grass now almost entirely tilled and planted to grains. Soils, often of loessial derivation, include the enormously productive chernozem (black earth) in the north, with reddish prairie soils of nearly equal fertility in the south. Throughout the region temperatures are severely continental, with bitterly cold winters in the north and scorching summers everywhere.

The prairie

The western edge of the prairie fades gradually into the shortgrass steppe of the High Plains, the change a function of diminishing rainfall. The eastern edge, however, represents one of the few major discordances between a climatic and biotic boundary in the United States, for the grassland penetrates the eastern forest in a great salient across humid Illinois and Indiana. Many scholars believe this part of the prairie was artificially induced by repeated burning and consequent destruction of the forest margins by Indians.

The Western mountains. Throughout the Cordillera and Intermontane regions, irregular topography shatters the grand bioclimatic pattern into an intricate mosaic of tiny regions that differ drastically according to elevation and exposure. No small- or medium-scale map can accurately record such complexity, and mountainous parts of the West are said, noncommitally, to have a "mountain climate." Lowlands are usually dry, but increasing elevation brings lower temperature, decreased evaporation, and—if a slope faces prevailing winds—greater precipitation. Soils vary wildly from place to place, but vegetation is fairly predictable. From the desert or steppe of intermontane valleys, a climber typically ascends into parklike savanna, then through an orderly sequence of increasingly humid and boreal forests until, if the range is high enough, one reaches the timberline and Arctic tundra. The very highest peaks are snow-capped, although permanent glaciers rarely occur outside the cool humid highlands of the Pacific Northwest. (P.F.L.)

PLANT LIFE

The dominant features of the vegetation are indicated by the terms forest, grassland, desert, and alpine tundra.

A coniferous forest of white and red pine, hemlock, spruce, jack pine, and balsam fir extends interruptedly in a narrow strip near the Canadian border from Maine to Minnesota and southward along the Appalachian Mountains. There may be found smaller stands of tamarack, spruce, paper birch, willow, alder, and aspen or poplar. Southward, a transition zone of mixed conifers and deciduous trees gives way to a hardwood forest of broadleaved trees. This forest, with varying mixtures of maple, oak, ash, locust, linden, sweet gum, walnut, hickory, sycamore, beech, and the more southerly tulip tree, once extended uninterruptedly from New England to Missouri and eastern Texas. Pines are prominent on the Atlantic and Gulf coastal plain and adjacent uplands, often occurring in nearly pure stands called pine barrens. Pitch, longleaf, slash, shortleaf, Virginia, and loblolly pines are commonest. Hickory and various oaks combine to form a significant part of this forest, with magnolia, white cedar, and ash often seen. In the frequent swamps, bald cypress, tupelo, and white cedar predominate. Pines, palmettos, and live oaks are replaced at the southern tip of Florida by the more tropical royal and thatch palms, figs, satinwood, and mangrove.

Principal grasslands

The grasslands occur principally in the Great Plains area and extend westward into the intermontane basins and benchlands of the Rocky Mountains. Numerous grasses such as buffalo, grama, side oat, bunch, needle, and wheat grass, together with many kinds of herbs, make up the plant cover. Coniferous forests cover the lesser mountains and high plateaus of the Rockies, Cascades, and Sierra Nevada. Ponderosa (yellow) pine, Douglas fir, western red cedar, western larch, white pine, lodgepole pine, several spruces, western hemlock, grand fir, red fir, and the lofty

redwood are the principal trees of these forests. The densest growth occurs west of the Cascade and Coast ranges in Washington, Oregon, and northern California, where the trees are often 100 feet or more in height. There the forest floor is so dark that only ferns, mosses, and a few shade-loving shrubs and herbs may be found.

The alpine tundra, located in the coterminous United States only in the mountains above the limit of trees, consists principally of small plants that bloom brilliantly for a short season. Sagebrush is the most common plant of the arid basins and semideserts west of the Rocky Mountains, but juniper, nut pine, and mountain mahogany are often found on the slopes and low ridges. The desert, extending from southeastern California to Texas, is noted for the many species of cactus, some of which grow to the height of trees, and for the Joshua tree and other yuccas, creosote bush, mesquite, and acacias.

The United States is rich in the variety of its native forest trees, some of which, as the species of sequoia, are the most massive known. More than 1,000 species and varieties have been described, of which almost 200 are of economic value, either because of the timber and other useful products that they yield or by reason of their importance in forestry.

Besides the native flowering plants, estimated at between 20,000 to 25,000 species, many hundreds of species introduced from other regions—chiefly Europe, Asia, and tropical America—have become naturalized. A large proportion of these are common annual weeds of fields, pastures, and roadsides. In some districts these naturalized "aliens" constitute 50 percent or more of the total plant population. (P.H.O./R.C.R./Ed.)

ANIMAL LIFE

With most of North America, the United States lies in the Nearctic faunistic realm, a region containing an assemblage of species similar to Eurasia and North Africa but sharply different from the tropical and subtropical zones to the south. Main regional differences correspond roughly with primary climatic and vegetal patterns. Thus, for example, the animal communities of the Dry West differ sharply from those of the Humid East and from those of the Pacific Coast. Because animals tend to range over wider areas than plants, faunal regions are generally coarser than vegetal regions and harder to delineate sharply.

Disruption of natural patterns of animal life

The animal geography of the United States, however, is far from a natural pattern, for European settlement produced a series of environmental changes that grossly altered the distribution of animal communities. First, many species were hunted to extinction or near extinction, most conspicuously, perhaps, the American bison, which ranged by the millions nearly from coast to coast but now rarely lives outside of zoos and wildlife preserves. Second, habitats were upset or destroyed throughout most of the country—forests cut, grasslands plowed and overgrazed, and migration paths interrupted by fences, railroads, and highways. Third, certain introduced species found hospitable niches and, like the English sparrow, spread over huge areas, often preempting the habitats of native animals. Fourth, though their effects are not fully understood, chemical biocides such as DDT were used for so long and in such volume that they are believed at least partly responsible for catastrophic mortality rates among large mammals and birds, especially predators high on the food chain. Fifth, there has been a gradual northward migration of certain tropical and subtropical insects, birds, and mammals, perhaps encouraged by gradual climatic warming. In consequence, many native animals have been reduced to tiny fractions of their former ranges or exterminated completely, while other animals, both native and introduced, have found the new anthropocentric environment well suited to their needs, with explosive effects on their populations. The coyote, opossum, armadillo, and several species of deer are among the animals that now occupy much larger ranges than they once did. (P.F.L.)

Arrangement of the account of the distribution of the fauna according to the climatic and vegetal regions has the merit that it can be compared further with the distribution of insects and of other invertebrates, some of which may be expected to fall into the same patterns as the vertebrates, while others, with different modes or different ages of dispersal, have geographic patterns of their own.

The transcontinental zone of coniferous forest at the north, the taiga, and the tundra zone into which it merges at the northern limit of tree growth are strikingly paralleled by similar vertical zones in the Rockies, and on Mount Washington in the east, where the area above the timberline and below the snow line is often inhabited with tundra animals like the ptarmigan and the white *Parnassius* butterflies, while the spruce and other conifers below the timberline form a belt sharply set off from the grassland or hardwood forest or desert at still lower altitudes.

A whole series of important types of animals spread beyond the limits of such regions or zones, sometimes over most of the continent. Aquatic animals, in particular, may live equally in forest and plains, in the Gulf states, and at the Canadian border. Such widespread animals include the white-tailed (Virginia) deer and black bear, the puma (though only in the remotest parts of its former range) and bobcat, the river otter (though now rare in inland areas south of the Great Lakes) and mink, and the beaver and muskrat. The distinctive coyote ranges over all of western North America and eastward as far as Maine. The snapping turtle ranges from the Atlantic coast to the Rocky Mountains.

Widespread fauna

In the northern coniferous forest zone, or taiga, the relations of animals with European or Eurasian representatives are numerous, and this zone is also essentially circumpolar. The relations are less close than in the Arctic forms, but the moose, beaver, hare, red fox, otter, wolverine, and wolf are recognizably related to Eurasian animals. Even some fishes, like the whitefishes (Coregonidae), the yellow perch, and the pike, exhibit this kind of Old World–New World relation. A distinctively North American animal in this taiga assemblage is the Canadian porcupine.

The hardwood forest area of the eastern and the southeastern pinelands compose the most important of the faunal regions within the United States. A great variety of fishes, amphibians, and reptiles of this region have related forms in East Asia, and this pattern of representation is likewise found in the flora. This area is rich in catfishes, minnows, and suckers. The curious ganoid fishes, the bowfin and the gar, are ancient types. The spoonbill cat, a remarkable type of sturgeon in the lower Mississippi, is represented elsewhere in the world only in the Yangtze in China. The Appalachian region is headquarters for the salamanders of the world, with no less than seven of the eight families of this large group of amphibians represented; no other continent has more than three of the eight families together. The eellike sirens and amphiumas (congo snakes) are confined to the southeastern states. The lungless salamanders of the family Plethodontidae exhibit a remarkable variety of genera and a number of species centring in the Appalachians. There is a great variety of frogs, and these include tree frogs whose main development is South American and Australian. The emydid freshwater turtles of the southeast parallel those of East Asia to a remarkable degree, though the genus *Clemmys* is the only one represented in both regions. Much the same is true of the water snakes, pit vipers, rat snakes, and green snakes, though still others are peculiarly American. The familiar alligator is a form with an Asiatic relative, the only other living true alligator being a species in central China.

In its mammals and birds the southeastern fauna is less sharply distinguished from the life to the north and west and is less directly related to that of East Asia. The forest is the home of the white-tailed deer, the black bear, the gray fox, the raccoon, and the common opossum. The wild turkey and the extinct hosts of the passenger pigeon were characteristic. There is a remarkable variety of woodpeckers. The birdlife in general tends to differ from that of Eurasia in the presence of birds, like the tanagers, American orioles, and hummingbirds, that belong to South American families. Small mammals abound with types of the worldwide rodent family Cricetidae, and with distinctive moles and shrews.

American bison

Most distinctive of the grassland animals proper is the American bison, whose nearly extinct European relative,

the wisent, is a forest dweller. The most distinctive of the American hoofed animals is the pronghorn, or prongbuck, which represents a family intermediate between the deer and the true antelopes in that it sheds its horns like a deer but retains the bony horn cores. The pronghorn is perhaps primarily a desert mammal, but it formerly ranged widely into the shortgrass plains. Everywhere in open country in the West there are conspicuous and distinctive rodents. The burrowing pocket gopher is peculiarly American, rarely seen making its presence known by pushed-out mounds of earth. The ground squirrels of the genus *Citellus* are related to those of Central Asia, and resemble them in habit; in North America the gregarious prairie dog is a closely related form. The American badger, not especially related to the badger of Europe, has its headquarters in the grasslands. The prairie chicken is a bird distinctive of the plains region, which is invaded everywhere by birds from both the east and the west.

The Southwestern deserts are a paradise for reptiles. Distinctive lizards such as the poisonous Gila monster abound, and the rattlesnakes, of which only a few species are found elsewhere in the United States, are common there. Desert reptile species often range to the Pacific Coast and northward into the Great Basin. Noteworthy mammals are the graceful bipedal kangaroo rat (almost exlusively nocturnal), the ring-tailed cat, a relative of the raccoon, and the piglike peccary.

The Rocky Mountains and other western ranges afford distinctive habitats for rock- and cliff-dwelling hoofed animals and rodents. The small pikas, related to the rabbit, inhabit talus areas at high altitudes as they do in the mountain ranges of East Asia. Marmots live in the Rockies as in the Alps. Every western range formerly had its own race of mountain sheep. At the north the Rocky Mountain goat lives at high altitudes—it is more properly a goat antelope, related to the takin of the mountains of western China. The dipper, remarkable for its habit of feeding in swift-flowing streams, though otherwise a bird without special aquatic adaptations, is a Rocky Mountain form with relatives in Asia and Europe.

In the Pacific region the extremely distinctive primitive tailed frog *Ascaphus,* which inhabits icy mountain brooks, represents a family by itself, perhaps more nearly related to the frogs of New Zealand than to more familiar types. The Cascades and Sierras form centres for salamanders of the families Ambystomoidae and Plethodontidae second only to the Appalachians, and there are also distinctive newts. The burrowing lizards, of the well-defined family Anniellidae, are found only in a limited area in coastal California. The only family of birds distinctive of North America, that of the wren-tits, Chamaeidae, is found in the chaparral of California. The mountain beaver, or sewellel (which is not at all beaverlike), is likewise a type peculiar to North America, confined to the Cascades and Sierras, and there are distinct kinds of moles in the Pacific area.

Mountain beaver

The mammals of the two coasts are strikingly different, though true seals (the harbour seal and the harp seal) are found on both. The sea lions, with longer necks and with projecting ears, are found only in the Pacific—the California sea lion, the more northern Steller's sea lion, and the fur seal. On the East Coast the larger rivers of Florida are inhabited by the Florida manatee, or sea cow, a close relative of the more widespread and more distinctively marine West Indian species. (K.P.S./Ed.)

SETTLEMENT PATTERNS

Although the land that now constitutes the United States was occupied and much affected by diverse Indian cultures over many millennia, these pre-European settlement patterns have had virtually no impact upon the contemporary nation—except locally, as in parts of New Mexico. A benign habitat permitted a huge contiguous tract of settled land to materialize across nearly all the eastern half of the United States and within substantial patches of the West. The vastness of the land, the scarcity of labour, and the abundance of migratory opportunities in a land replete with raw physical resources contributed to exceptional human mobility and a quick succession of ephemeral forms of land use and settlement. Human endeavours have greatly transformed the landscape, but such efforts have been largely destructive. Most of the pre-European landscape in the United States was so swiftly and radically altered that it is difficult to conjecture intelligently about its earlier appearance.

The overall impression of the settled portion of the American landscape, rural or urban, is one of disorder and incoherence, even in areas of strict geometric survey. The individual landscape unit is seldom in visual harmony with its neighbour, so that, however sound in design or construction the single structure may be, the general effect is untidy. These attributes have been intensified by the acute individualism of the American, vigorous speculation in land and other commodities, a strongly utilitarian attitude toward the land and the treasures above and below it, and government policy and law. The landscape is also remarkable for its extensive transportation facilities, which have greatly influenced the configuration of the land.

Another special characteristic of American settlement, one that became obvious only by the mid-20th century, is the convergence of rural and urban modes of life. The farmsteads—and rural folk in general—have become increasingly urbanized, and agricultural operations have become more automated, while the metropolis grows more gelatinous, unfocused, and pseudo-bucolic along its margins.

Rural settlement. Patterns of rural settlement indicate much about the history, economy, society, and minds of those who created them as well as about the land itself. The essential design of rural activity in the United States bears a strong family resemblance to that of other neo-European lands, such as Canada, Australia, New Zealand, South Africa, Argentina, or tsarist Siberia—places that have undergone rapid occupation and exploitation by immigrants intent upon short-term development and enrichment. In all such areas, under novel social and political conditions and with a relative abundance of territory and physical resources, ideas and institutions derived from a relatively stable medieval or early modern Europe have undergone major transformation. Further, these are non-peasant countrysides, alike in having failed to achieve the intimate symbiosis of people and habitat, the humanized rural landscapes characteristic of many relatively dense, stable, earthbound communities in parts of Asia, Africa, Europe, and Latin America.

Comparative rural patterns

Early models of land allocation. From the beginning the prevalent official policy of the British (except between 1763 and 1776) and then of the U.S. government was to promote agricultural and other settlement—to push the frontier westward as fast as physical and economic conditions permitted. The British crown's grants of large, often vaguely specified tracts to individual proprietors or companies enabled the grantees to draw settlers by the sale or lease of land at attractive prices or even by outright gift.

Of the numerous attempts at group colonization, the most notable effort was the theocratic and collectivist New England town that flourished, especially in Massachusetts, Connecticut, and New Hampshire, during the first century of settlement. The town, the basic unit of government and comparable in area to townships in other states, allotted both rural and village parcels to single families by group decision. Contrary to earlier scholarly belief, in all but a few cases settlement was spatially dispersed in the socially cohesive towns, at least until about 1800. The relatively concentrated latter-day villages persist today as amoeba-like entities straggling along converging roads, neither fully rural nor agglomerated in form. The only latter-day settlement experiment of notable magnitude to achieve enduring success was a series of Mormon settlements in the Great Basin region of Utah and adjacent states, with their tightly concentrated farm villages reminiscent of the New England model. Other efforts have been made along ethnic, religious, or political lines, but success has been at best brief and fragile.

Creating the national domain. With the coming of independence and after complex negotiations, the original 13 states surrendered to the new national government nearly all their claims to the unsettled western lands beyond their boundaries. Some tracts, however, were reserved for

disposal to particular groups. Thus, the Western Reserve of northeastern Ohio gave preferential treatment to natives of Connecticut, while the military tracts in Ohio and Indiana were used as bonus payments to veterans of the American Revolution.

A federally administered national domain was created, to which the great bulk of the territory acquired in 1803 in the Louisiana Purchase and later beyond the Mississippi and in 1819 in Florida was consigned. The only major exceptions were the public lands of Texas, which were left within that state's jurisdiction; such earlier French and Spanish land grants as were confirmed, often after tortuous litigation; and some Indian lands. In sharp contrast to the slipshod methods of colonial land survey and disposal, the federal land managers expeditiously surveyed, numbered, and mapped their territory in advance of settlement, beginning with Ohio in the 1780s, then sold or deeded it to settlers under inviting terms at a number of regional land offices.

The establishment of townships

The design universally followed in the new survey system (except within the French, Spanish, and Indian grants) was a simple, efficient rectangular scheme. Townships were laid out as blocks, each six by six miles in size, oriented with the compass directions. Thirty-six sections, each one square mile, or 640 acres (260 hectares), in size, were designated within each township; and public roads were established along section lines and, where needed, along half-section lines. At irregular intervals, offsets in survey lines and roads were introduced to allow for the Earth's curvature. Individual property lines were coincident with, or parallel to, survey lines, and this pervasive rectangularity generally carried over into the geometry of fields and fences or into the townsites later superimposed upon the basic rural survey.

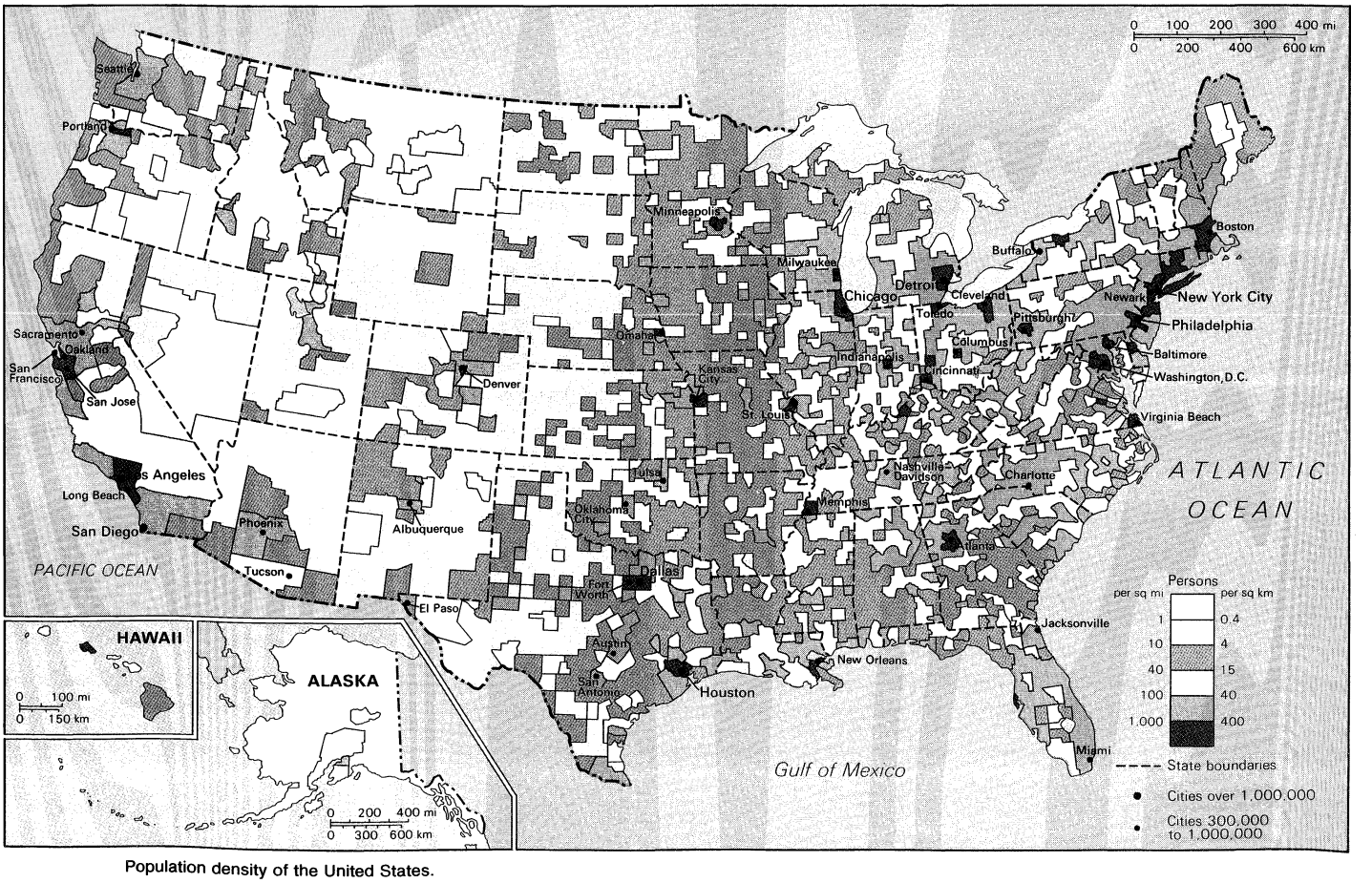
This all-encompassing checkerboard pattern is best appreciated from an airplane window over Iowa or Kansas. There, one sees few streams or other natural features and few diagonal highways or railroads interrupting the overwhelming squareness of the landscape. A systematic rectangular layout, rather less rigorous in form, also appears in much of Texas and in those portions of Maine, western New York and Pennsylvania, and southern Georgia that were settled after the 1780s.

Distribution of rural lands. Since its formation, Congress has enacted a series of complex schemes for distribution of the national domain. The most famous of these plans was the Homestead Act of 1862, which offered title to 160 acres to individual settlers, subject only to residence for a certain period of time and to the making of minimal improvements to the land thus acquired. The legal provisions of such acts have varied with time as the nature of farming technology and of the remaining lands have changed, but their general effect has been to perpetuate the Jeffersonian ideal of a republic in which yeoman farmers own and till self-sufficient properties.

The program was successful in providing private owners with relatively choice lands, aside from parcels reserved for schools and various township and municipal uses. More than one-third of the national territory, however, is still owned by federal and state governments, with much of this land in forest and wildlife preserves. A large proportion of this land is in the West and is unsuited for intensive agriculture or grazing because of the roughness, dryness, or salinity of the terrain; much of it is leased out for light grazing or for timber cutting.

Patterns of farm life. During the classic period of American rural life, around 1900, the typical American lived or worked on a farm or was economically dependent upon farmers. In contrast to rural life in many other parts of the world, the farm family lived on an isolated farmstead some distance from town and often from farm neighbours; its property averaged less than one-quarter square mile. This farmstead varied in form and content with local tradition and economy. In particular, barn types were localized—for example, the tobacco barns of the South, the great dairy barns of Wisconsin, or the general-purpose forebay barns of southeastern Pennsylvania—as were modes of fencing. In general, however, the farmstead contained

The farm in about 1900



Population density of the United States.

dwelling, barn, storage and sheds for small livestock and equipment, a small orchard, and a kitchen garden. A woodlot might be found in the least-accessible or least-fertile part of the farm.

Successions of such farms were connected with one another and with the towns by means of a dense, usually rectangular lattice of roads, largely unimproved at the time. The hamlets, villages, and smaller cities were arrayed at relatively regular intervals, with size and affluence determined in large part by the presence and quality of rail service or status as the county seat. But, among people who have been historically rural, individualistic, and antiurban in bias, many services normally located in urban places might be found in rustic settings. Thus, much retail business was transacted by means of itinerant peddlers, while small shops for the fabrication, distribution, or repair of various items were often located in isolated farmsteads, as were many post offices.

Social activity also tended to be widely dispersed among numerous rural churches, schools, or grange halls; and the climactic event of the year might well be the county fair, political rally, or religious encampment—again on a rural site. Not the least symptomatic sign of the strong tendency toward spatial isolation are the countless family burial plots or community cemeteries so liberally distributed across the countryside.

Town morphologies in the rural United States

Regional small-town patterns. There has been much regional variation among smaller villages and hamlets, but such phenomena have received relatively little attention from students of American culture or geography. The distinctive New England village, of course, is generally recognized and cherished: it consists of a loose clustering of white frame buildings, including a church (usually Congregationalist or Unitarian), town hall, shops, and stately homes with tall shade trees around the central green, or commons—a grassy expanse that may contain a bandstand and monuments or flowers. Derivative village forms were later carried westward to sections of the northern Midwest.

Less widely known but equally distinctive is the town morphology characteristic of the Midland, or Pennsylvanian, culture area and most fully developed in southeastern and central Pennsylvania and Piedmont Maryland. It differs totally from the New England model in density, building materials, and general appearance. Closely packed, often contiguous buildings—mostly brick, but sometimes stone, frame, or stucco—abut directly on a sidewalk, which is often paved with brick and usually thickly planted with maple, sycamore, or other shade trees. Such towns are characteristically linear in plan, have dwellings intermingled with other types of buildings, have only one or two principal streets, and may radiate outward from a central square lined with commercial and governmental structures.

The most characteristic U.S. small town is the one whose pattern evolved in the Midwest. Its simple scheme is usually based on the grid plan. Functions are rigidly segregated spatially, with the central business district, consisting of closely packed two- or three-story brick buildings, limited exclusively to commercial and administrative activity. The residences, generally set well back within spacious lots, are peripheral in location, as are most rail facilities, factories, and warehouses.

Even the modest urbanization of the small town came late to the South. Most urban functions long were spatially dispersed—almost totally so in the early Chesapeake Bay country or North Carolina—or were performed entirely by the larger plantations dominating the economic life of much of the region. When city and town began to materialize in the 19th and 20th centuries, they tended to follow the Midwestern model in layout.

Mormon villages

Although quite limited in geographic area, the characteristic villages of the Mormon and Hispanic-American districts are of considerable interest. The Mormon settlement uncompromisingly followed the ecclesiastically imposed grid plan composed of square blocks, each with perhaps only four very large house lots, and the block surrounded by extremely wide streets. Those villages in New Mexico in which population and culture were derived from Old Mexico were often built according to the standard Latin-American plan. The distinctive feature is a central plaza dominated by a Roman Catholic church and encircled by low stone or adobe buildings.

The rural–urban transition. *Weakening of the agrarian ideal.* The United States has had little success in achiev-

Grant Heilman

Agricultural fields in a checkerboard pattern, near Alden, Kan.

ing or maintaining the ideal of the family farm. Through purchase, inheritance, leasing, and other means, some of dubious legality, smaller properties have been merged into much larger entities. By the late 1980s, for example, when the average farm size had surpassed 460 acres, farms containing 2,000 or more acres accounted for almost half of all farmland and 20 percent of the cropland harvested, even though they comprised less than 3 percent of all farms. At the other extreme were those 60 percent of all farms that contained fewer than 180 acres and reported less than 15 percent of cropland harvested. This trend toward fewer but larger farms has continued.

The huge, heavily capitalized "neoplantation," essentially a factory in the field, is especially conspicuous in parts of California, Arizona, and the Mississippi Delta, but examples can be found in any state. There are also many smaller but intensive operations that call for large investments and advanced managerial skills. This trend toward large-scale, capital-intensive farm enterprise has been paralleled by a sharp drop in rural farm population—a slump from the all-time high of some 32,000,000 in the early 20th century to about 5,000,000 in the late 1980s; but even in 1940, when farm folk still numbered more than 30,000,000, nearly 40 percent of farm operators were tenants, and another 10 percent were only partial owners.

The mythology of small-town America

As the agrarian population has dwindled, so too has its immediate impact lessened, though less swiftly, in economic and political matters. The rural United States, however, has been the source of many of the nation's values and images. The United States has become a highly urbanized, technologically advanced society far removed in daily life from cracker barrel, barnyard, corral, or logging camp. Although Americans have gravitated, sometimes reluctantly, to the big city, in the daydreams and assumptions that guide many sociopolitical decisions, the memory of a rapidly vanishing agrarian America is well noted. This is revealed not only in the works of contemporary novelists, poets, and painters but also throughout the popular arts: in movies, television, soap operas, folklore, country music, political oratory, and in much leisure activity.

Impact of the motor vehicle. Since about 1920 more genuine change has occurred in American rural life than during the preceding three centuries of European settlement in North America. Although the basic explanation is the profound social and technological transformations engulfing most of the world, the most immediate agent of change has been the internal-combustion engine. The automobile, truck, bus, and paved highway have more than supplanted a moribund passenger and freight railroad system. While many local rail depots have been boarded up and scores of secondary lines have been abandoned, hundreds of thousands of miles of old dirt roads have been paved, and a vast system of interstate highways has been constructed to connect major cities in a single nonstop network. The net result has been a shrinking of travel time and an increase in miles traveled for the individual driver, rural or urban.

Small towns in the United States have undergone a number of changes. Before 1970 towns near highways and urban centres generally prospered; while in the less-fortunate towns, where the residents lingered on for the sake of relatively cheap housing, downtown businesses often became extinct. From the late 1960s until about 1981 the rural and small-town population grew at a faster rate than the metropolitan population, the so-called metro–nonmetro turnaround, thus reversing more than a century of relatively greater urban growth. Subsequent evidence, however, suggests an approach toward equilibrium between the urban and rural sectors.

As Americans have become increasingly mobile, the visual aspect of rural America has altered drastically. The highway has become the central route, and many of the functions once confined to the local town or city now stretch for many miles along major roads.

Development of a metropolitan society

Reversal of the classic rural dominance. The metropolitanization of life in the United States has not been limited to city, suburb, or exurb; it now involves most of the rural area and population. The result has been the decline of local crafts and regional peculiarities, quite visibly in such items as farm implements, fencing, silos, and housing and in commodities such as clothing or bread. In many ways, the countryside is now economically dependent on the city.

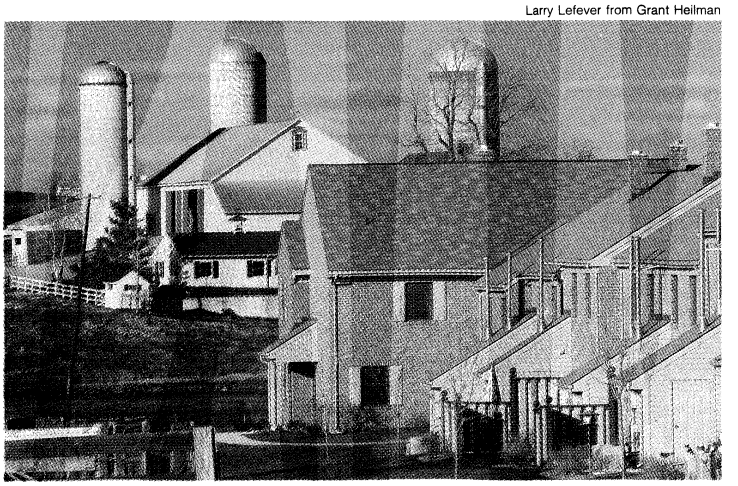
The city dweller is the dominant consumer for products other than those of field, quarry, or lumber mill; and city location tends to determine patterns of rural economy rather than the reverse. During weekends and the vacation seasons, swarms of city folk stream out to second homes in the countryside and to campgrounds, ski runs, beaches, boating areas, or hunting and fishing tracts. For many large rural areas, recreation is the principal source of income and employment; and such areas as northern New England and upstate New York have become playgrounds and sylvan refuges for many urban residents.

The larger cities reach far into the countryside for their vital supplies of water and energy. There is an increasing reliance upon distant coalfields to provide fuel for electrical power plants, and cities have gone far afield in seeking out rural disposal sites for their ever-growing volumes of garbage.

The majority of the rural population now lives within daily commuting range of a sizable city. This enables many farm residents to operate their farms while, at the same time, working part- or full-time at a city job, and it thus helps to prevent the drastic decline in rural population that has occurred in remoter parts of the country. Similarly, many small towns within the shadow of a metropolis, with fewer and fewer farmers to service, have become dormitory satellites, serving residents from nearby cities and suburbs.

Urban settlement. The United States has moved from a predominantly rural settlement into an urban society. In so doing, it has followed the general path that other advanced nations have traveled and one along which

Larry Lefever from Grant Heilman



Residential area encroaching on farmland, Lancaster county, Pa.

developing nations have begun to hasten. About three-fourths of the population live clustered within officially designated urban places and urbanized areas, which account for less than 2 percent of the national territory. At least another 15 percent live in dispersed residences that are actually urban in economic or social orientation.

Classic patterns of siting and growth. Although more than 95 percent of the population was rural during the colonial period and for the first years of independence, cities were crucial elements in the settlement system from the earliest days. Boston; New Amsterdam (New York City); Jamestown, Va.; Charleston, S.C.; and Philadelphia were founded at the same time as the colonies they served. Like nearly all other North American colonial towns of consequence, they were ocean ports. Until at least the beginning of the 20th century the historical geography of U.S. cities was intimately related with that of successive transportation systems. The location of successful cities with respect to the areas they served, as well as their internal structure, was determined largely by the nature of these systems.

Evolving functions of cities

The colonial cities acted as funnels for the collection and shipment of farm and forest products and other raw materials from the interior to trading partners in Europe, the Caribbean, or Africa and for the return flow of manufactured goods and other locally scarce items, as well as immigrants. Such cities were essentially marts and warehouses, and only minimal attention was given to social, military, educational, or religious functions. The inadequacy and high cost of overland traffic dictated sites along major ocean embayments or river estuaries; the only pre-1800 nonports worthy of notice were Lancaster and York, both in Pennsylvania, and Williamsburg, Va. With the populating of the interior and the spread of a system of canals and improved roads, such new cities as Pittsburgh, Pa.; Cincinnati, Ohio; Buffalo, N.Y.; and St. Louis, Mo., mushroomed at junctures between various routes or at which modes of transport were changed. Older ocean ports, such as New Castle, Del.; Newport, R.I.; Charleston, S.C.; Savannah, Ga.; and Portland, Maine, whose locations prevented them from serving large hinterlands, tended to stagnate.

From about 1850 to 1920 the success of new cities and the further growth of older ones in large part were dependent on their location within the new steam railroad system and on their ability to dominate a large tributary territory. Such waterside rail hubs as Buffalo; Toledo, Ohio; Chicago; and San Francisco gained population and wealth rapidly, while such offspring of the rail era as Atlanta, Ga.; Indianapolis, Ind.; Minneapolis, Minn.; Fort Worth, Texas; and Tacoma, Wash., also grew dramatically. Much of the rapid industrialization of the 19th and early 20th centuries occurred in places already favoured by water or rail transport systems; but in some instances, such as in the cities of northeastern Pennsylvania's anthracite region, some New England mill towns, and the textile centres of the Carolina and Virginia Piedmont, manufacturing brought about rapid urbanization and the consequent attraction of transport facilities. The extraction of gold, silver, copper, coal, iron, and, in the 20th century, gas and oil led to rather ephemeral centres—unless these places were able to capitalize on local or regional advantages other than minerals.

A strong early start, whatever the inital economic base may have been, was often the key factor in competition among cities. With sufficient early momentum, urban capital and population tended to expand almost automatically. The point is illustrated perfectly by the larger cities of the northeastern seaboard, from Portland, Maine, through Baltimore, Md. The nearby physical wealth is poor to mediocre, and they are now far off-centre on the national map; but a prosperous mercantile beginning, good land and sea connections with distant places, and a rich local accumulation of talent, capital, and initiative were sufficient to bring about the growth of one of the world's largest concentrations of industry, commerce, and people.

New factors in municipal development. The pre-1900 development of the American city was almost completely a chronicle of the economics of the production, collection, and distribution of physical commodities and basic services dictated by geography, but there have been striking deviations from this pattern. The physical determinants of urban location and growth have given way to social factors. Increasingly, the most successful cities are oriented toward the more advanced modes for the production and consumption of services, specifically the knowledge, managerial, and recreational industries. The largest cities have become more dependent upon corporate headquarters, communications, and the manipulation of information for their sustenance. Washington, D.C., is the most obvious example of a metropolis in which government and ancillary activities have been the spur for vigorous growth; but almost all of the state capitals have displayed a similar demographic and economic vitality. Further, urban centres that contain a major college or university often have enjoyed remarkable expansion.

Social factors in city growth

With the coming of relative affluence and abundant leisure to the population and a decrease of labour input in industrial processes, a new breed of cities has sprouted across the land: those that cater to the pleasure-seeker, vacationer, and the retired—for example, the young, flourishing cities of Florida or Nevada and many locations in California, Arizona, and Colorado.

The automobile as a means of personal transportation was developed about the time of World War I, and the American city was catapulted into a radically new period, both quantitatively and qualitatively, in the further evolution of physical form and function. The size, density, and internal structure of the city were previously constrained by the limitations of the pedestrian and early mass-transit systems. Only the well-to-do could afford horse and carriage or a secluded villa in the countryside. Cities were relatively small and compact, with a single clearly defined centre, and they grew by accretion along their edges, without any significant spatial hiatuses except where commuter railroads linked outlying towns to the largest of metropolises. Workers living beyond the immediate vicinity of their work had to locate within reach of the few horse-drawn omnibuses or the later electric street railways.

Effect of the automobile

The universality of the automobile, even among the less affluent, and the parallel proliferation of service facilities and highways greatly loosened and fragmented the American city, which spread over surrounding rural lands. Older, formerly autonomous towns grew swiftly. Many towns became satellites of the larger city or were absorbed. Many suburbs and subdivisions arose with single-family homes on lots larger than had been possible for the ordinary householder in the city. These communities were almost totally dependent on the highway for the flow of commuters, goods, and services, and many were located in splendid isolation, separated by tracts of farmland, brush, or forest from other such developments. At the major interchanges of the limited-access highways, a new form of agglomerated settlement sprang up. In a further elaboration of this trend, many larger cities have been girdled by a set of mushrooming complexes. These creations of private enterprise embody a novel concept of urban existence: a metropolitan module no longer reliant on the central city or its downtown. Usually anchored on a cluster of shopping malls and office parks, these "hypersuburbs," whose residents and employees circulate freely within the outer metropolitan ring, offer virtually all of the social and economic facilities needed for the modern life-style.

The new look of the metropolitan area. The outcome has been a broad, ragged, semiurbanized belt of land surrounding each city, large or small, and quite often blending imperceptibly into the suburban-exurban halo encircling a neighbouring metropolitan centre. There is a great similarity in the makeup and general appearance of all such tracts: the planless intermixture of scraps of the rural landscape with the fragments of the scattered metropolis; the randomly distributed subdivisions or single homes; the vast shopping centres, the large commercial cemeteries, drive-in theatres, junkyards, and golf courses and other recreational enterprises; and the regional or metropolitan airport, often with its own cluster of factories, warehouses, or travel-oriented businesses. The traditional city—unitary, concentric in form, with a single

well-defined middle—has been replaced by a relatively amorphous, polycentric metropolitan sprawl.

The inner cities

The inner city of a large U.S. metropolitan area displays some traits that are common to the larger centres of all advanced nations. A central business district, almost always the oldest section of the city, is surrounded by a succession of roughly circular zones, each distinctive in economic and social-ethnic character. The symmetry of this scheme is distorted by the irregularities of surface and drainage or the effects of radial highways and railroads. Land is most costly, and hence land use is most intensive, toward the centre. Major business, financial and governmental offices, department stores, and specialty shops dominate the downtown which is usually fringed by a band of factories and warehouses. The outer parts of the city, like the suburbs, are mainly residential.

With some exceptions—*e.g.,* large apartment complexes in downtown Chicago—people do not reside in the downtown areas, and there is a steady downward gradient in population density per unit area (and more open land and single-family residences) as one moves from the inner city toward the open country. Conversely, there is a general rise in income and social status with increasing distance from the core. The sharply defined immigrant neighbourhoods of the 19th century generally persist in a somewhat diluted form, though specific ethnic groups may have shifted their location. Later migrant groups, notably Southern blacks and Latin-Americans, generally dominate the more run-down neighbourhoods of the inner cities.

Individual and collective character of cities. American cities, more so than the small-town or agrarian landscape, tend to be the product of a particular period rather than of location. The relatively venerable centres of the Eastern Seaboard—Boston; Philadelphia; Baltimore, Md.; Albany, N.Y.; Chester, Pa.; Alexandria, Va.; or Georgetown (a district of Washington, D.C.), for example—are virtual replicas of the fashionable European models of their early period rather than the fruition of a regional culture, unlike New Orleans and Santa Fe, N.M., which reflect other times and regions. The townscapes of Pittsburgh; Detroit, Mich.; Chicago; and Denver, Colo., depict national modes of thought and the technological development of their formative years, just as Dallas, Texas; Las Vegas, Nev.; San Diego, Calif.; Tucson, Ariz.; and Albuquerque, N.M., proclaim contemporary values and gadgetry more than any local distinctiveness. When strong-minded city founders instituted a highly individual plan and their successors managed to preserve it—as, for example, in Savannah, Ga.; Washington, D.C.; and Salt Lake City, Utah—or when there is a happy combination of a spectacular site and appreciative residents—as in San Francisco or Seattle, Wash.—a genuine individuality does seem to emerge. Such an identity also may develop where immigration has been highly selective, as in such places as Miami, Fla.; Phoenix, Ariz.; and Los Angeles.

As a group, U.S. cities differ from cities in other countries in both type and degree. The national political structure, the social inclinations of the people, and the strong outward surge of urban development have led to the political fragmentation of metropolises that socially and economically are relatively cohesive units. The fact that a single metropolitan area may sprawl across numerous incorporated towns and cities, several townships, and two or more counties and states has a major impact upon both its appearance and the way it functions. Not the least of these effects is a dearth of overall physical and social planning (or its ineffectuality when attempted), and the rather chaotic, inharmonious appearance of both inner-city and peripheral zones painfully reflects the absence of any effective collective action concerning such matters.

Political fragmentation and lack of planning

The American city is a place of sharp transitions. Construction, demolition, and reconstruction go on almost ceaselessly, though increasing thought has been given to preserving monuments and buildings. From present evidence, it would be impossible to guess that New York City and Albany date from the 1620s or that Detroit was founded in 1701. Preservation and restoration do occur, but often only when it makes sense in terms of tourist revenue. Physical and social blight has reached epidemic proportions in the slum areas of the inner city; but, despite the wholesale razing of such areas and the subsequent urban-renewal projects (sometimes as apartment or commercial developments for the affluent), the belief has become widespread that the ills of the U.S. city are incurable, especially with the increasing flight of capital, tax revenue, and the more highly educated, affluent elements of the population to suburban areas and the spatial and political polarization of whites and nonwhites.

In the central sections of U.S. cities, there is little sense of history or continuity; instead, one finds evidence of the dominance of the engineering mentality and of the credo that the business of the city is business. Commercial and administrative activities are paramount, and usually there is little room for church buildings or for parks or other nonprofit enterprises. The role of the cathedral, so central in the medieval European city, is filled by a U.S. invention serving both utilitarian and symbolic purposes, the skyscraper. Some cities have felt the need for other bold secular monuments; hence the Gateway Arch looming over St. Louis, Seattle's Space Needle, and Houston's Astrodome. Future archaeologists may well conclude from their excavations that American society was ruled by an oligarchy of highway engineers, architects, and bulldozer operators. The great expressways converging upon, or looping, the downtown area and the huge amount of space devoted to parking lots and garages are even more impressive than the massive surgery executed upon U.S. cities a century ago to hack out room for railroad terminals and marshaling yards.

Within many urban sites there has been such radical physical transformation of shoreline, drainage systems, and land surface as would be difficult to match elsewhere in the world. Thus, in their physical lineaments, Manhattan and inner Boston bear scant resemblance to the landscapes seen by their initial settlers. The surface of downtown Chicago has been raised several feet above its former swamp level, the city's lakefront extensively reshaped, and the flow of the Chicago River reversed. Los Angeles, notorious for its disregard of the environment, has its concrete arroyo bottoms, terraced hillsides and landslides, and its own artificial microclimate.

The supercities. The unprecedented outward sprawl of American urban settlement has created some novel settlement forms, for the quantitative change has been so great as to induce qualitative transformation. The conurbation—a territorial coalescence of two or more sizable cities whose peripheral zones have grown together—may have first appeared in early 19th-century Europe. There are major examples in Great Britain, the Low Countries, and West Germany, as well as in Japan.

Megalopolis: concept and forms

Nothing elsewhere, however, rivals in size and complexity the aptly named megalopolis, that supercity stretching along the Atlantic from Portland, Maine, past Richmond, Va. Other large conurbations include, in the Great Lakes region, one centred on Chicago and containing large slices of Illinois, Wisconsin, and Indiana; another based in Detroit, embracing large parts of Michigan and Ohio and reaching into Canada; and a third stretching from Buffalo through Cleveland and back to Pittsburgh. All three are reaching toward one another and may form another megalopolis that, in turn, may soon be grafted onto the seaboard megalopolis by a corridor through central New York state.

Another example of a growing megalopolis is the huge southern California conurbation reaching from Santa Barbara, through a dominating Los Angeles, to the Mexican border. The solid strip of urban territory that lines the eastern shore of Puget Sound is a smaller counterpart. Quite exceptional in form is the slender linear multicity occupying Florida's Atlantic coastline, from Jacksonville to Miami, and the loose swarm of medium-sized cities clustering along the Southern Piedmont, from south central Virginia to Birmingham, Ala.; also of note are the Texas cities of Dallas–Fort Worth, Houston, and San Antonio, which have formed a rapidly growing—though discontinuous—urbanized triangle.

One of the few predictions that seem safe in so dynamic and innovative a land as the United States is that, un-

less severe and painful controls are placed on land use, the shape of the urban environment will be increasingly megalopolitan: a small set of great constellations of polycentric urban zones, each complexly interlocked socially and physically with its neighbours.

TRADITIONAL REGIONS OF THE UNITED STATES

Degree of difference among regions

The differences among America's traditional regions, or culture areas, tend to be slight and shallow as compared with such areas in most older, more stable countries. The muted, often subtle nature of interregional differences can be ascribed to the relative newness of American settlement, a perpetually high degree of mobility, a superb communications system, and the galloping centralization of economy and government. It might even be argued that some of these regions are quaint vestiges of a vanishing past, of interest only to antiquarians.

Yet, in spite of the nationwide standardization in many areas of American thought and behaviour, the lingering effects of the older culture areas do remain potent. In the case of the South, for example, the differences helped to precipitate the gravest political crisis and bloodiest military conflict in the nation's history. More than a century after the Civil War, the South remains a powerful entity in political, economic, and social terms, and its peculiar status is recognized in religious, educational, athletic, and literary circles.

Even more intriguing is the appearance of a series of essentially 20th-century regions. Southern California is the largest and perhaps the most distinctive region, and its special culture has attracted large numbers of immigrants to the state. Similar trends are visible in southern Florida; in Texas, whose mystique has captured the national imagination; and to a certain degree in the more ebullient regions of New Mexico and Arizona as well. At the metropolitan level, it is difficult to believe that such distinctive cities as San Francisco, Las Vegas, Dallas, Tucson, and Seattle have become like all other American cities. A detailed examination, however, would show significant if sometimes subtle interregional differences in terms of language, religion, diet, folklore, folk architecture and handicrafts, political behaviour, social etiquette, and a number of other cultural categories.

The hierarchy of culture areas. A multitiered hierarchy of culture areas might be postulated for the United States; but the most interesting levels are, first, the nation as a whole and, second, the five to 10 large subnational regions, each embracing several states or major portions thereof.

Relation of political and cultural boundaries

There is a remarkably close coincidence between the political United States and the cultural United States. Crossing into Mexico, the traveler passes across a cultural chasm. If the contrasts are less dramatic between the two sides of the U.S.-Canadian boundary, they are nonetheless real, especially to the Canadian. Erosion of the cultural barrier has been largely limited to the area that stretches from northern New York state to Aroostook county, Maine. There, a vigorous demographic and cultural immigration by French-Canadians has gone far toward eradicating international differences.

While the international boundaries act as a cultural container, the interstate boundaries are curiously irrelevant. Even when the state had a strong autonomous early existence—as happened with Massachusetts, Virginia, or Pennsylvania—subsequent economic and political forces have tended to wash away such initial identities. Actually, it could be argued that the political divisions of the 48 coterminous states are anachronistic in the context of contemporary socioeconomic and cultural forces. Partially convincing cases might be built for equating Utah and Texas with their respective culture areas because of exceptional historical and physical circumstances, or perhaps Oklahoma, given its very late European occupation and its dubious distinction as the territory to which exiled Indian tribes of the East were relegated. In most instances, however, the states either contain two or more distinctly different culture and political areas or fragments thereof or are part of a much larger single culture area. Thus sharp North–South dichotomies characterize California, Missouri, Illinois, Indiana, Ohio, and Florida, while Tennessee advertises that there are really three Tennessees. In Virginia the opposing cultural forces were so strong that actual fission took place in 1863 (with the admission to the Union of West Virginia) along one of those rare interstate boundaries that approximate a genuine cultural divide.

Much remains to be learned about the cause and effect relations between economic and culture areas in the United States. If the South or New England could at one time be correlated with a specific economic system, this is no longer easy to do. Cultural systems appear to respond more slowly to agents of change than do economic or urban systems. Thus the Manufacturing Belt, a core region for many social and economic activities, now spans parts of four traditional culture areas—New England, the Midland, the Midwest, and the northern fringes of the South. The great urban sprawl, from southern Maine to central Virginia, blithely ignores the cultural slopes that are still visible in its more rural tracts.

The cultural hearths. The culture areas of the United States are generally European in origin, the result of importing European colonists and ways of life and the subsequent adaptation of social groups to new habitats. The aboriginal cultures have had relatively little influence on the nation's modern culture. In the Southwestern and the indistinct Oklahoman subregions, the Indian element merits consideration only as one of several ingredients making up the regional mosaic. With some exceptions, the map of American culture areas in the East can be explained in terms of the genesis, development, and expansion of the three principal colonial cultural hearths along the Atlantic seaboard. Each was basically British in character, but their personalities remain distinct because of, first, different sets of social and political conditions during the critical period of first effective settlement and, second, local physical and economic circumstances. The cultural gradients between them tend to be much steeper and the boundaries more distinct than is true for the remainder of the nation.

New England. New England was the dominant region during the century of rapid expansion following the American Revolution and not merely in terms of demographic or economic expansion. In social and cultural life—in education, politics, theology, literature, science, architecture, and the more advanced forms of mechanical and social technology—the area exercised its primacy. New England was the leading source of ideas and styles for the nation from about 1780 to 1880; it furnishes an impressive example of the capacity of strongly motivated communities to rise above the constraints of a harsh environment.

The character of early New England

During its first two centuries, New England had an unusually homogeneous population. With some exceptions, the British immigrants shared the same nonconformist religious beliefs, language, social organization, and general outlook. A distinctive regional culture took form, most noticeably in terms of dialect, town morphology, and folk architecture. The personality of the people also took on a regional coloration both in folklore and in actuality; there is sound basis for the belief that the traditional New England Yankee is self-reliant, thrifty, inventive, and enterprising. The influx of immigrants that began in the 1830s diluted and altered the New England identity, but much of its early personality survived.

By virtue of location, wealth, and seniority, the Boston metropolitan area has become the cultural economic centre of New England. This sovereignty is shared to some degree, however, with two other old centres, the lower Connecticut Valley and the Narragansett Bay region of Rhode Island.

The early westward demographic and ideological expansion of New England was so influential that it is justifiable to call New York, northern New Jersey, northern Pennsylvania, and much of the Upper Midwest "New England Extended." Further, the energetic endeavours of New England whalers, merchants, and missionaries had a considerable impact on the cultures of Hawaii, various other Pacific isles, and several points in the Caribbean. New Englanders also were active in the Americanization of early Oregon and Washington, with results that are still visible. Later, the overland diffusion of New England natives and practices meant a recognizable New England

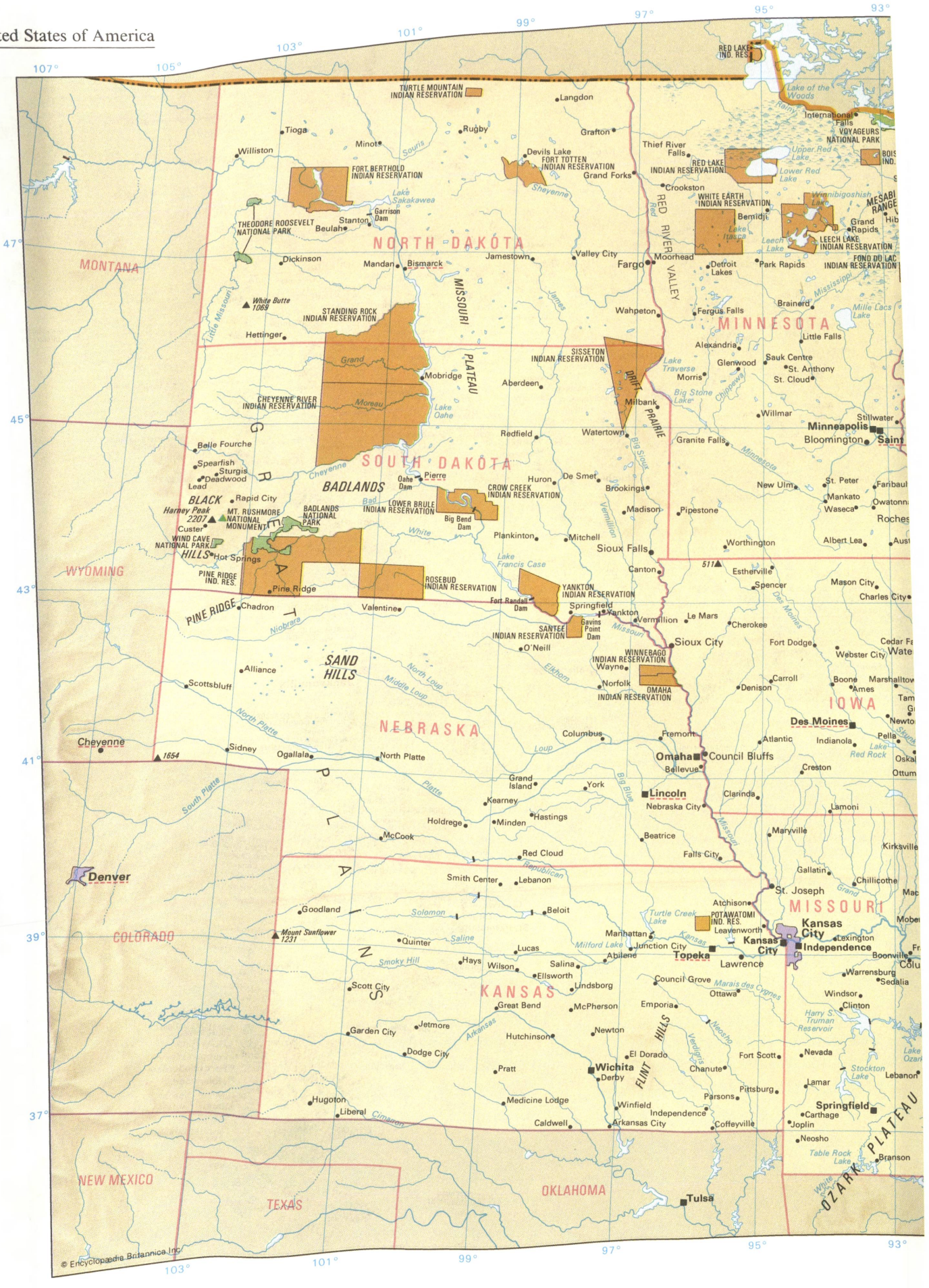
CANADA
Montreal
Quebec
St. Lawrence
Lake Memphremagog
Lake Champlain
Connecticut
Androscoggin
Kennebec
Penobscot
Allagash
St. John
Eagle Lake
Chamberlain Lake
Chesuncook Lake
Mount Katahdin 1606
Moosehead Lake
Kennebago Lake
Nicatous Lake
West Grand Lake
Passamaquoddy Bay
Grand Manan Channel
Lake Winnipesaukee
Sebago Lake
Casco Bay
Penobscot Bay
Muscongus Bay
Sunapee Lake
Merrimack
Quabbin Reservoir
Walden Pond
Massachusetts Bay
Cape Cod Bay
Cape Cod Canal
Buzzards Bay
Nantucket Sound
Block Island Sound
Rhode Island Sound
Narragansett Bay
Long Island Sound
Hudson
Housatonic
Thames
Quinebaug
Blackstone
Charles
Scituate
Ocatuqua
Mississquoi
Winooski
Lamoille
Otter Creek
White
Mad
Saco

80°
78°
76°
74°
72°
45°
43°
41°
39°
37°
CANADA
Georgian Bay
Lake Simcoe
Lake Ontario
Lake Erie
Ottawa
Montreal
Toronto
Hamilton
Buffalo
Rochester
Syracuse
Albany
Schenectady
Utica
Erie
Pittsburgh
Harrisburg
Philadelphia
Baltimore
Washington, D.C.
Alexandria
Silver Spring
Bethesda
Richmond
Norfolk
Wilmington
Trenton
Camden
Newark
Jersey City
Elizabeth
New York City
Yonkers
Paterson
Allentown
Bethlehem
Reading
Scranton
Binghamton
Williamsport
Altoona
Lancaster
York
Hagerstown
Cumberland
Frederick
Annapolis
Dover
Wilmington
Atlantic City
Ocean City
Cape May
Wildwood
Salisbury
Lewes
Rehoboth Beach
Bethany Beach
Selbyville
Seaford
Laurel
Cambridge
Easton
Chestertown
Crisfield
Princess Anne
Pocomoke
Chesapeake Bay
Delaware Bay
Delaware River
Susquehanna River
Potomac River
Rappahannock
James
Monongahela
Ohio
Allegheny
Genesee
Mohawk
Hudson
Connecticut
Housatonic
Thames
Long Island Sound
Lake Champlain
Lake Placid
Saranac Lake
Tupper Lake
Mount Marcy 1629
Indian Lake
Schroon Lake
Ticonderoga
Glens Falls
Saratoga Springs
Amsterdam
Gloversville
Cooperstown
Cobleskill
Catskill
Hudson
Kingston
Poughkeepsie
Newburgh
Middletown
Monticello
Port Jervis
Peekskill
Ossining
White Plains
Port Chester
New Rochelle
Stony Brook
Hampton Bays
Coram
Hicksville
Brentwood
Levittown
Freeport
Long Island
Fire Island National Seashore
Staten Island
Perth Amboy
Edison
Princeton
Trenton
Asbury Park
Neptune
Mantoloking
Seaside Park
Beach Haven
Ship Bottom
Brigantine National Wildlife Refuge
Atlantic City
Ocean City
Avalon
Cape May
Wildwood
Cape Henlopen
Rehoboth Beach
Bethany Beach
Selbyville
Assateague Island National Seashore
Ocean City
Sinepuxent Bay
Atlantic Ocean
Adirondack Mountains
Catskill Mts.
Pocono Mts.
Blue Mtn.
Tuscarora Mountain
South Mountain
Allegheny Mountain
Appalachian Mountains
Pennsylvania
New York
Vermont
New Hampshire
Massachusetts
Connecticut
Rhode Island
Maryland
Delaware
Virginia
West Virginia
North Carolina
Ohio
Delaware Water Gap National Recreation Area
St. Regis Mohawk Indian Reservation
Cattaraugus Indian Reservation
Tonawanda Indian Res.
Allegany Indian Reservation
Oneida Lake
Seneca Lake
Cayuga Lake
Chautauqua Lake
Finger Lakes
Erie Canal
Chesapeake and Ohio Canal
Deep Creek Lake
Backbone Mtn. 1024
Mount Davis 979
Niagara Falls
Carry Falls Reservoir
Stillwater Reservoir
Hinckley Reservoir
Sacandaga Reservoir
Pepacton Reservoir
Allegheny Reservoir
E. Br. Clarion River Lake
Glendale Lake
Hopatcong Lake
Francis Scott Key Bridge
Intracoastal Waterway
John H. Kerr Reservoir
Scale 1:4,480,000
1 inch equals approx. 70 miles
0 25 50 75 100 mi
0 25 50 75 100 125 150 km
Albers Equal-Area Projection
©Encyclopædia Britannica Inc.

Davenport
Chicago
Gary
Toledo
Lake Erie
Cleveland
Parma
Akron
Fort Wayne
Peoria
Springfield
Indianapolis
Springfield
Columbus
Cincinnati
Erlanger
Newport
Covington
Florence
St. Louis
Missouri
Louisville
Frankfort
Georgetown
Paris
Lexington
Versailles
Okolona
Mt. Sterling
Winchester
Richmond
Boonesboro
Danville
Berea
Harrodsburg
Bardstown
Brandenburg
Radcliff
Elizabethtown
Hodgenville
Leitchfield
Henderson
Owensboro
Madisonville
Greenville
Marion
Paducah
Mayfield
Clinton
Fulton
Murray
Hopkinsville
Bowling Green
Russellville
Glasgow
Somerset
London
Campbellsville
MAMMOTH CAVE NATIONAL PARK
Mt. Vernon
Corbin
Monticello
Franklin
Springfield
Clarksville
Union City
Paris
Dyersburg
Camden
Trenton
Nashville–Davidson
Waverly
Brentwood
Franklin
Gallatin
Jamestown
Lebanon
Cookeville
Smyrna
Murfreesboro
McMinnville
Shelbyville
Columbia
Tullahoma
Dayton
Athens
Cleveland
Chattanooga
Red Bank
Oak Ridge
Norris
Middlesboro
Cumberland Gap
Harlan
Pineville
Hazard
Clintwood
Norton
Abingdon
Bristol
Kingsport
Elizabethtown
Johnson City
Mountain City
Sparta
Mount Rogers 1746
Marion
Pulaski
Radford
Blacksburg
Christiansburg
Princeton
Bluefield
Welch
Pikeville
Oak Hill
Beckley
Logan
Williamson
Hinton
Lewisburg
White Sulphur Springs
Covington
Lexington
Huntington
Charleston
Dunbar
Summersville
Hillsboro
Spencer
Point Pleasant
Buckhannon
Weston
Philippi
Elkins
Weirton
Wellsburg
Wheeling
Moundsville
New Martinsville
Morgantown
Fairmont
Grafton
Clarksburg
Parkersburg
St. Marys
Pittsburgh
Berkeley Springs
Martinsburg
Keyser
Romney
Charles Town
Winchester
Harpers Ferry
Leesburg
Harrisburg
Trenton
Philadelphia
Wilmington
Baltimore
Dover
Annapolis
Herndon
Arlington
Annandale
Alexandria
Manassas
Warrenton
Washington, D.C.
Harrisonburg
Staunton
Waynesboro
Charlottesville
Fredericksburg
New Market
Spruce Knob 1482
SHENANDOAH NATIONAL PARK
Richmond
West Point
Chesterfield
Hopewell
Williamsburg
Yorktown
Poquoson
Petersburg
Colonial Heights
Farmville
Lynchburg
Roanoke
Salem
Rocky Mount
Martinsville
Danville
South Boston
Emporia
Hampton
Newport News
Norfolk
Portsmouth
Suffolk
Chesapeake
Virginia Beach
Chesapeake Bay Bridge–Tunnel
MACKAY ISLAND NWR
Elizabeth City
Edenton
Kitty Hawk
Nags Head
Manteo
ROANOKE ISLAND
Eden
Madison
Reidsville
Henderson
Halifax
Winston-Salem
Burlington
Hillsborough
Louisburg
Durham
Chapel Hill
Raleigh
Rocky Mount
Tarboro
Plymouth
Greensboro
High Point
Thomasville
Lexington
Statesville
Morganton
Hickory
Lenoir
Mount Mitchell 2037
Salisbury
Cary
Garner
Wilson
Greenville
Washington
Bath
Smithfield
Goldsboro
Kannapolis
Concord
Sanford
Lillington
Asheville
Gastonia
Charlotte
Albemarle
Pinehurst
Fayetteville
Clinton
Monroe
Rockingham
Laurinburg
Lumberton
Whiteville
Wilmington
Jacksonville
Burgaw
Morehead City
Beaufort
Havelock
Kinston
New Bern
Cape Hatteras
CAPE HATTERAS NATIONAL SEASHORE
OCRACOKE ISLAND
CAPE LOOKOUT NATIONAL SEASHORE
CEDAR ISLAND NATIONAL WILDLIFE RESERVE
Cape Lookout
Onslow Bay
Pamlico Sound
Albemarle Sound
Knoxville
Gatlinburg
GREAT SMOKY MOUNTAINS NATIONAL PARK
Clingmans Dome 2025
CHEROKEE INDIAN RESERVATION
Maryville
Murfreesboro
Shelbyville
Columbia
Tullahoma
Dayton
Athens
Cleveland
Chattanooga
Red Bank
Hendersonville
Rosman
Franklin
Shelby
Huntsville
Tennessee–Tombigbee Waterway
Memphis
Germantown
Millington
Somerville
Bolivar
Shiloh
Savannah
Waynesboro
Lawrenceburg
Pulaski
Brownsville
Jackson
HATCHIE NATIONAL WILDLIFE RESERVE
Columbia
Atlanta
Birmingham
Macon
Charleston
Columbia
Cape Fear
Intracoastal Waterway
Long Bay
Green Swamp
ATLANTIC OCEAN
Chesapeake–Delaware Canal
Delaware Bay
DELMARVA PENINSULA
Great Dismal Swamp
Lake Barkley
Kentucky Lake
Lake Cumberland
Dale Hollow Reservoir
Center Hill Reservoir
Watts Bar Lake
Chickamauga Lake
Nickajack Lake
Pickwick Lake
Guntersville Lake
Wheeler Lake
Wilson Lake
Norris Lake
Cherokee Lake
Douglas Lake
Watauga Lake
Holston Lake
Fort Loudoun Lake
South Holston Lake
Boone Lake
Fontana Lake
Jordan Lake
John H. Kerr Reservoir
Roanoke Rapids Lake
Gaston Lake
Lake Norman
Mountain Island Lake
Lake Wylie
Lake Moultrie
Santee Lake
Lake Marion
Lake Murray
Lake Greenwood
Lake Hartwell
Clark Hill Reservoir
Lake Oconee
Lake Sinclair
Lake Blackshear
Lake Seminole
Lake Eufaula
Lake Martin
Lake Jordan
Lake Weiss
Lake Wheeler
Lake Gaston
Lake Waccamaw
Lake Tahoma
Lake Tillery
Lake Blalock
Lake Robinson
Lake Russell
Lake Thurmond
Lake Strom Thurmond
Lake Keowee
Lake Jocassee
Lake Hartwell
Lake Lanier
Lake Allatoona
Lake Sidney Lanier
Lake Walter F. George
Lake Seminole
Lake Eufaula
Lake Martin
Lake Jordan
Lake Weiss
Lake Wheeler
Lake Gaston
Lake Waccamaw
Lake Tahoma
Lake Tillery
Lake Blalock
Lake Robinson
Lake Russell
Lake Thurmond
Lake Strom Thurmond
Lake Keowee
Lake Jocassee
Lake Hartwell
Lake Lanier
Lake Allatoona
Lake Sidney Lanier
Lake Walter F. George
Ohio
Mississippi
Missouri
Wabash
Kentucky
Green
Tennessee
Cumberland
Big Sandy
Kanawha
Guyandotte
Tug Fork
New
Greenbrier
Monongahela
Allegheny
Susquehanna
Delaware
Shenandoah
Potomac
James
Appomattox
Roanoke
Chowan
Neuse
Pamlico
Pee Dee
Catawba
Broad
Yadkin
Watauga
Holston
French Broad
Clinch
Powell
Duck
Elk
Big Black
Yazoo
Tombigbee
Pickwick
Cumberland Mountains
Cumberland Plateau
Black Mountain 1263
Pine Mountain
Allegheny Plateau
Allegheny Mountains
Blue Ridge
Appalachian Mountains
Appalachian Plateau
Piedmont
Coastal Plain
MICHIGAN
OHIO
INDIANA
ILLINOIS
IOWA
MISSOURI
KENTUCKY
WEST VIRGINIA
PENNSYLVANIA
NEW YORK
NEW JERSEY
MARYLAND
DELAWARE
VIRGINIA
NORTH CAROLINA
SOUTH CAROLINA
GEORGIA
ALABAMA
MISSISSIPPI
TENNESSEE
ARKANSAS
Scale 1:6,060,600
1 inch equals approx. 96 miles
0 25 50 75 100 mi
0 50 100 150 km
Albers Equal-Area Projection
© Encyclopædia Britannica Inc.
90°
88°
86°
84°
82°
80°
78°
76°
40°
38°
36°
34°

95°
93°
91°
89°
35°
33°
31°
29°
27°
25°
MISSOURI
PLATEAU
OKLAHOMA
ARKANSAS
MISSISSIPPI
LOUISIANA
TEXAS
BOSTON MTS.
OZARK MTS.
OUACHITA MTS.
Tulsa
Siloam Springs
Rogers
Eureka Springs
Fayetteville
Springdale
Harrison
Mountain Home
Pocahontas
Paragould
Blytheville
Jonesboro
Osceola
Trumann
Batesville
Mountain View
Greers Ferry Lake
Newport
Heber Springs
Van Buren
Fort Smith
Clarksville
Russellville
Dardanelle
Mount Magazine 839
Searcy
Conway
Wynne
West Memphis
Memphis
Forrest City
Jacksonville
North Little Rock
Little Rock
Marianna
Southaven
Hernando
Holly Springs
Mena
HOT SPRINGS NATIONAL PARK
Hot Springs
Benton
Malvern
Stuttgart
West Helena
Helena
WHITE RIVER NATIONAL WILDLIFE REFUGE
Arkadelphia
Pine Bluff
Clarksdale
Oxford
De Queen
Murfreesboro
Ashdown
Fordyce
Dumas
Cleveland
Hope
Camden
Warren
Monticello
Grenada
Texarkana
Smackover
Greenville
Leland
Indianola
Greenwood
Winona
Magnolia
El Dorado
Crossett
NOXUBEE NATIONAL WILDLIFE REFUGE
Kosciusko
Homer
Bastrop
Yazoo City
Bossier City
Minden
Ruston
Lake Providence
Shreveport
Monroe
West Monroe
Rayville
Canton
Mount Driskill 163
Tallulah
Jackson
Vicksburg
Clinton
Brandon
Pearl
Jonesboro
Winnsboro
Mansfield
Winnfield
Port Gibson
Natchitoches
Jena
Hazlehurst
Magee
Toledo Bend Reservoir
Vidalia
Natchez
Brookhaven
Alexandria
Pineville
McComb
Columbia
Leesville
Marksville
Bunkie
De Ridder
Ville Platte
Bogalusa
Eunice
Opelousas
Zachary
Amite
Baker
Hammond
Picayune
Baton Rouge
Gulfport
Lake Charles
Jennings
Sulphur
Lafayette
Crowley
Breaux Bridge
Bay St. Louis
Slidell
Beaumont
Houston
Pasadena
Abbeville
St. Martinville
New Iberia
Jeanerette
Franklin
Lake Pontchartrain
Lake Pontchartrain Bridge
Kenner
Metairie
New Orleans
Thibodaux
Morgan City
Houma
SABINE NATIONAL WILDLIFE REFUGE
BRETON NATIONAL WILDLIFE REFUGE
MARSH ISLAND
Atchafalaya Bay
Intracoastal Waterway
MISSISSIPPI DELTA
Scale 1:6,060,600
1 inch equals approx. 96 miles
0 25 50 75 100 mi
0 50 100 150 km
Albers Equal-Area Projection

© Encyclopædia Britannica Inc.

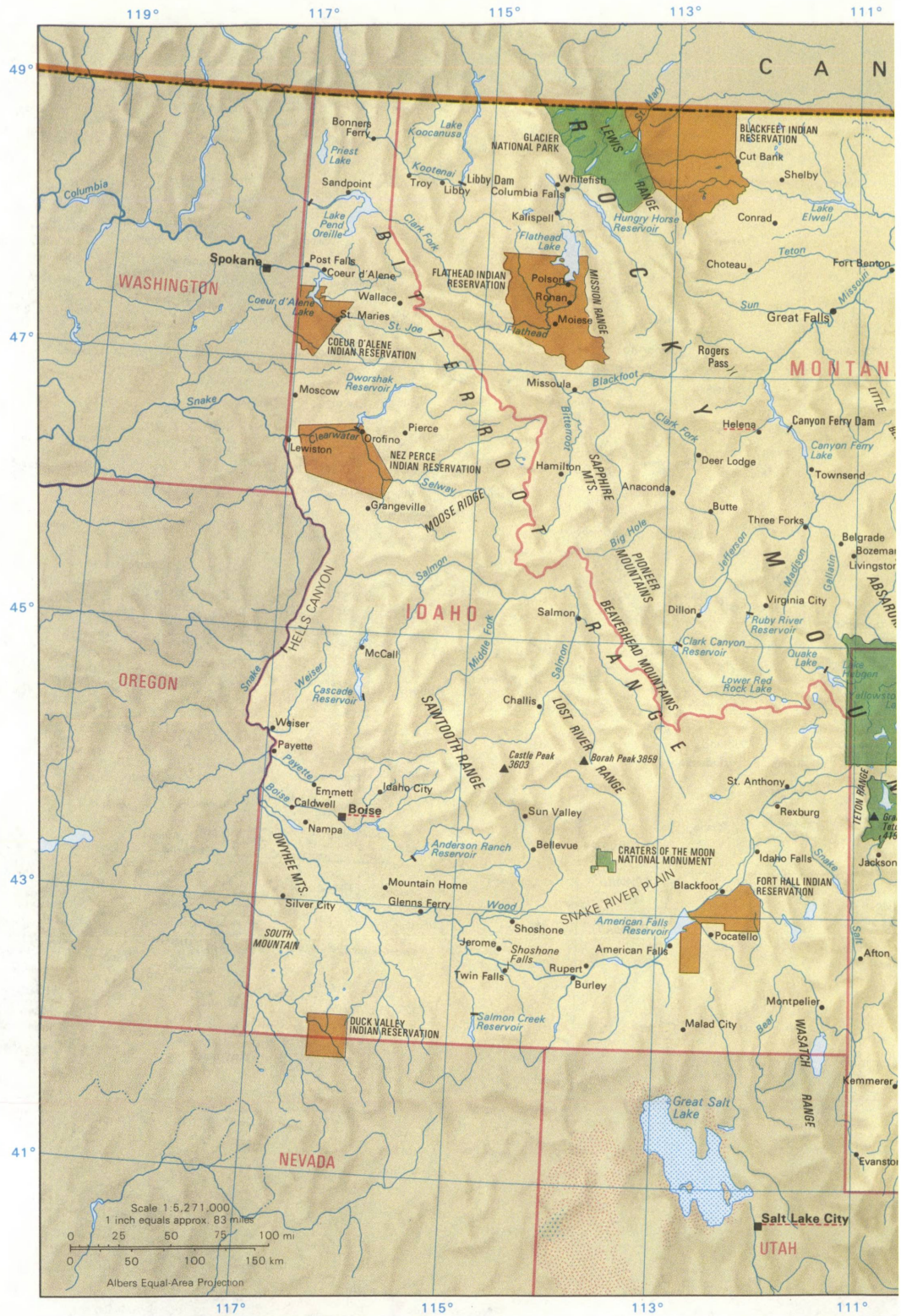


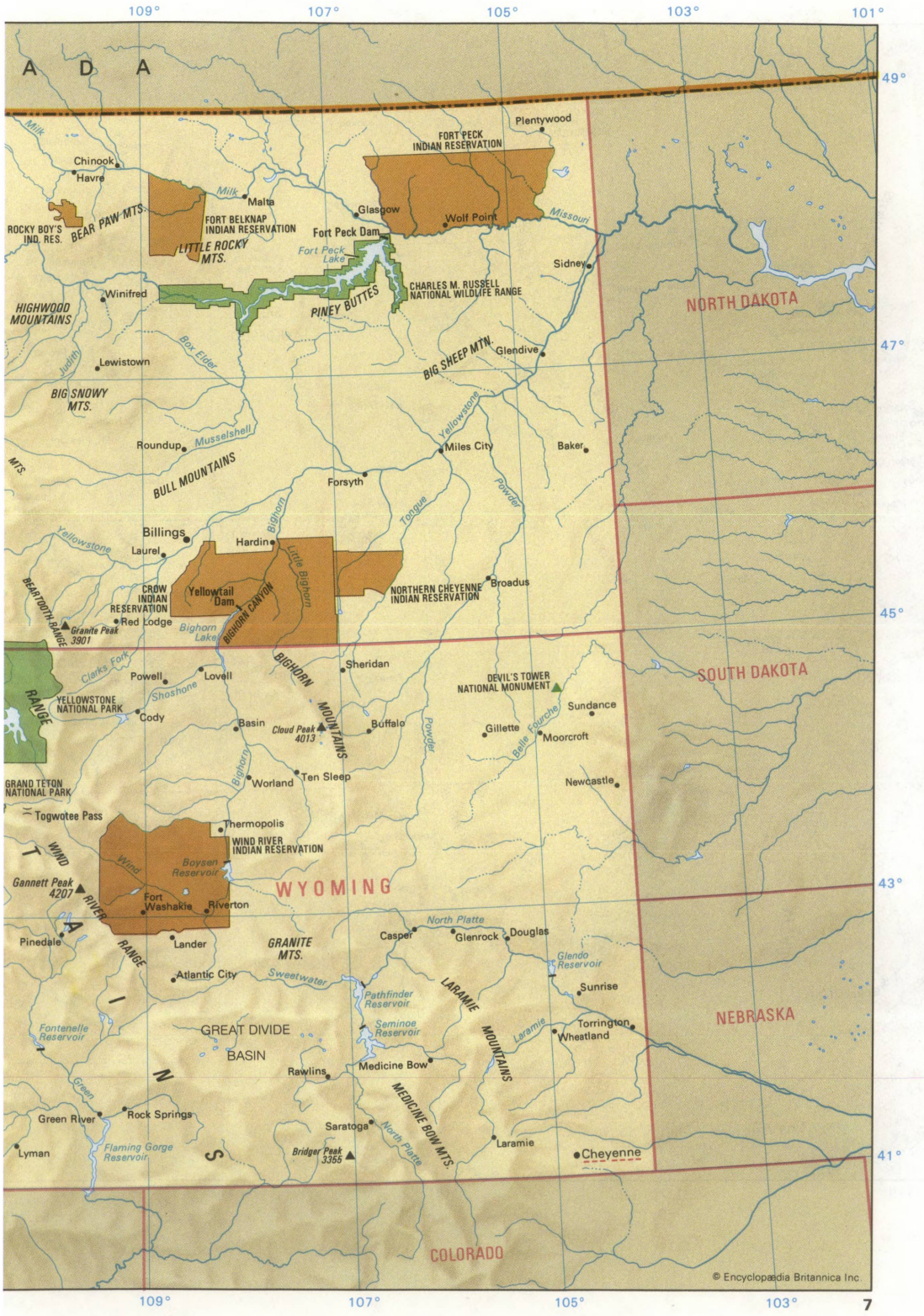
© Encyclopædia Britannica, Inc.

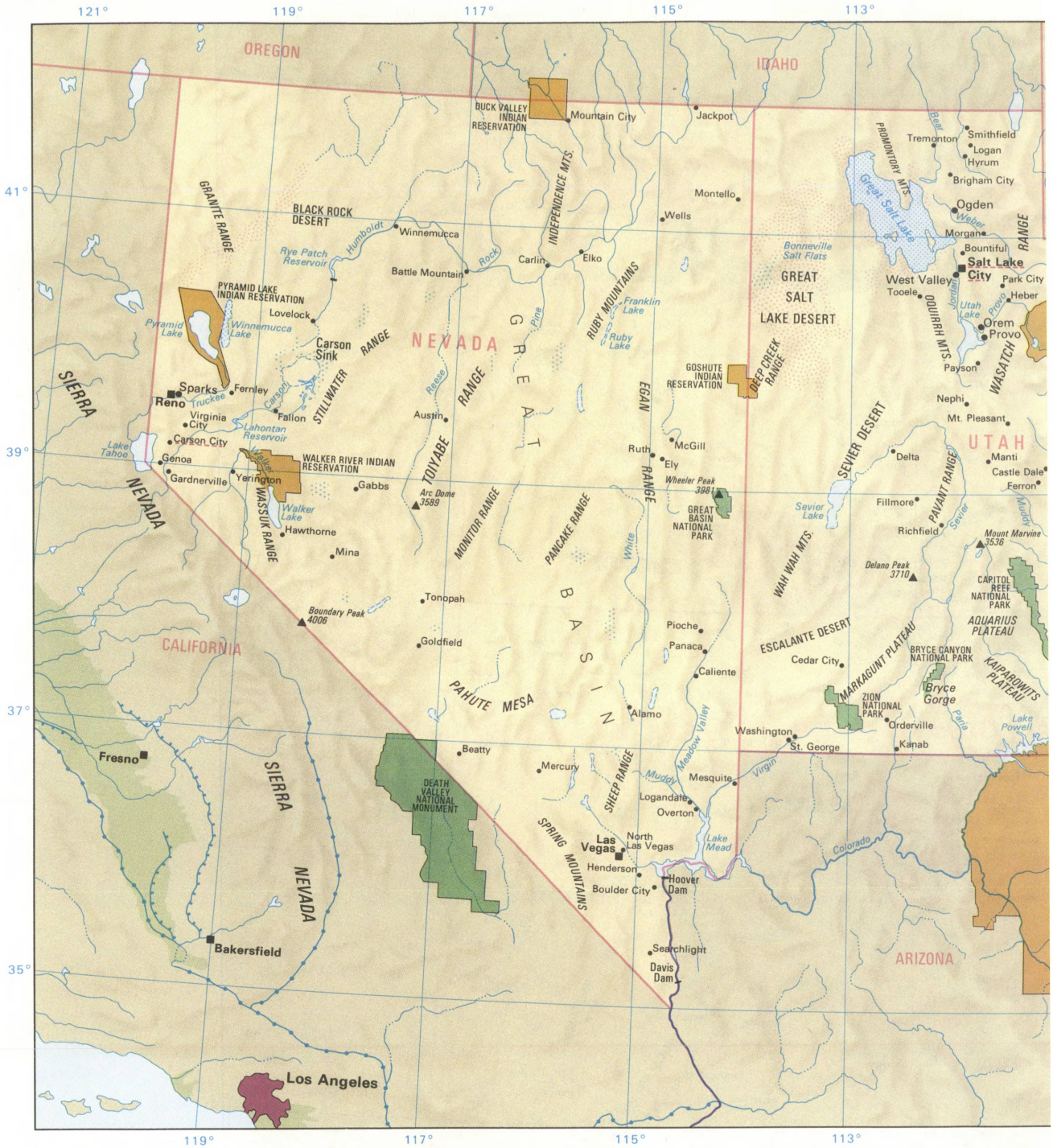
CANADA
Scale 1: 6,465,000
1 inch equals approx. 102 miles
0 25 50 75 100 125 mi
0 50 100 150 200 km
Albers Equal-Area Projection
91°
89°
87°
85°
83°
81°
79°
47°
45°
43°
41°
39°
37°
Lake Superior
Lake Michigan
Lake Huron
Lake Erie
Lake Ontario
MICHIGAN
WISCONSIN
ILLINOIS
INDIANA
OHIO
KENTUCKY
TENNESSEE
WEST VIRGINIA
PENNSYLVANIA
NEW YORK
VIRGINIA
NORTH CAROLINA
Thunder Bay
Duluth
Houghton
Marquette
Sault Ste. Marie
Mackinac Bridge
Mackinaw City
Cheboygan
Petoskey
Gaylord
Alpena
Escanaba
Iron Mountain
Rhinelander
Menominee
Green Bay
Appleton
Oshkosh
Fond du Lac
Sheboygan
Manitowoc
Two Rivers
Manistee
Ludington
Muskegon
Grand Rapids
Kalamazoo
Battle Creek
Jackson
Lansing
East Lansing
Flint
Saginaw
Bay City
Midland
Mount Pleasant
Big Rapids
Detroit
Warren
Pontiac
Livonia
Ann Arbor
Ypsilanti
Toledo
Cleveland
Akron
Youngstown
Canton
Mansfield
Columbus
Cincinnati
Dayton
Springfield
Lima
Fort Wayne
South Bend
Gary
Chicago
Milwaukee
Madison
Rockford
Peoria
Springfield
Champaign
Urbana
Bloomington
Decatur
Indianapolis
Terre Haute
Evansville
Louisville
Lexington
Nashville–Davidson
Knoxville
Charleston
Roanoke
Winston–Salem
Pittsburgh
Erie
Toronto
Hamilton
Buffalo
St. Louis
Cedar Rapids
Dubuque
La Crosse
Green Bay
Wausau
Eau Claire
Menominee
Iron Mountain
Escanaba
Houghton
Calumet
Hancock
Ironwood
Ashland
Rhinelander
Wisconsin Rapids
Stevens Point
Portage
Baraboo
Beaver Dam
Janesville
Beloit
Racine
Kenosha
Waukegan
Elgin
Aurora
Naperville
Joliet
Oak Park
Evanston
Gary
Hammond
Hobart
Valparaiso
Michigan City
Benton Harbor
St. Joseph
Kalamazoo
Grand Haven
Muskegon
Holland
Zeeland
Wyoming
Owosso
Port Huron
Sarnia
Windsor
Lake St. Clair
Erie
Conneaut
Ashtabula
Kirtland
Cleveland
Lorain
Elyria
Sandusky
Fremont
Bowling Green
Defiance
Fostoria
Findlay
Akron
Barberton
Cuyahoga Falls
Warren
Youngstown
Alliance
Wooster
Middleton
Ashland
Massillon
Canton
East Liverpool
Steubenville
Pittsburgh
Mansfield
Marion
Delaware
Newark
Zanesville
Cambridge
Quaker City
Coshocton
Mount Vernon
Sugar Creek
Dover
New Philadelphia
Martins Ferry
Marietta
Lancaster
Circleville
Chillicothe
Athens
Portsmouth
Gallipolis
Jackson
Hamilton
Middletown
Springfield
Xenia
Kettering
Germantown
Richmond
Greenfield
Connersville
Shelbyville
Greencastle
Crawfordsville
Lafayette
Frankfort
Elwood
Muncie
Anderson
Lebanon
New Castle
Greenville
Tipton
Kokomo
Peru
Logansport
Wabash
Huntington
Fort Wayne
Warsaw
Elkhart
Goshen
Middlebury
Nappanee
Mishawaka
Sturgis
Three Rivers
Kankakee
Joliet
Morris
Ottawa
La Salle
Peru
Streator
Pontiac
Bloomington
Normal
Rantoul
Champaign
Urbana
Danville
Decatur
Effingham
Mattoon
Charleston
Terre Haute
Martinsville
Columbus
Seymour
Bedford
Madison
Vincennes
Washington
Jasper
Clarksville
New Albany
Corydon
Louisville
Frankfort
Lexington
Evansville
New Harmony
Santa Claus
Olney
Salem
Centralia
Mount Vernon
Belleville
St. Louis
East St. Louis
Granite City
Edwardsville
Alton
Florissant
Canolia
Jefferson City
Springfield
Petersburg
Jacksonville
Quincy
Hannibal
Carlinville
Mexico
Fulton
Macomb
Canton
Peoria
Pekin
Galesburg
Monmouth
Burlington
Fairfield
Mt. Pleasant
Muscatine
Davenport
Rock Island
Moline
East Moline
Clinton
Marion
Cedar Rapids
Iowa City
Springdale
Oregon
De Kalb
Sterling
Dixon
Rockford
Belvidere
Freeport
Galena
Dubuque
Prairie du Chien
Madison
Stoughton
New Glarus
Spring Green
Wisconsin Dells
Baraboo
Decorah
Black River Falls
Tomah
Wisconsin Rapids
Stevens Point
Wausau
Eau Claire
Menomonie
Chippewa Falls
Rice Lake
Ashland
Ironwood
Houghton
Hancock
Calumet
Duluth
Superior
Ely
Virginia
Eagle Mountain 701
Thunder Bay
Sault Ste. Marie
Marquette
Escanaba
Iron Mountain
Rhinelander
Petoskey
Gaylord
Alpena
Cheboygan
Mackinaw City
Mackinac Island
St. Ignace
Traverse City
Sleeping Bear Dunes National Lakeshore
Interlochen
Cadillac
Houghton Lake
Saginaw Bay
Midland
Bay City
Saginaw
Frankenmuth
Isabella Ind. Res.
Mount Pleasant
Big Rapids
Ludington
Manistee
Muskegon
Grand Haven
Holland
Zeeland
Grand Rapids
Wyoming
Kalamazoo
Battle Creek
Jackson
Lansing
East Lansing
Owosso
Flint
Port Huron
Sarnia
Windsor
Lake St. Clair
Detroit
Warren
Pontiac
Livonia
Ann Arbor
Ypsilanti
Monroe
Adrian
Toledo
Sandusky
Lorain
Cleveland
Elyria
Akron
Barberton
Cuyahoga Falls
Warren
Youngstown
Conneaut
Ashtabula
Erie
Buffalo
Toronto
Hamilton
Niagara Falls
Lake Ontario
Lake Erie
Lake Huron
Lake Michigan
Lake Superior
Saginaw Bay
Houghton Lake
Grand Traverse Bay
Whitefish Bay
Keweenaw Bay
Green Bay
Lake Winnebago
Pentenwell Lake
Lake Wappapello
Lake of the Ozarks
Mississippi
Wisconsin
Fox
Rock
Illinois
Kaskaskia
Wabash
White
Ohio
Cumberland
Tennessee
Muskingum
Scioto
Miami
Maumee
St. Joseph
Kankakee
Des Plaines
Sangamon
Spoon
Meramec
Black
St. Francis
Arkansas
Missouri
Kansas
Nebraska
Iowa
Minnesota
North Dakota
South Dakota
Wisconsin
Michigan
Illinois
Indiana
Ohio
Kentucky
Tennessee
West Virginia
Pennsylvania
New York
Virginia
North Carolina

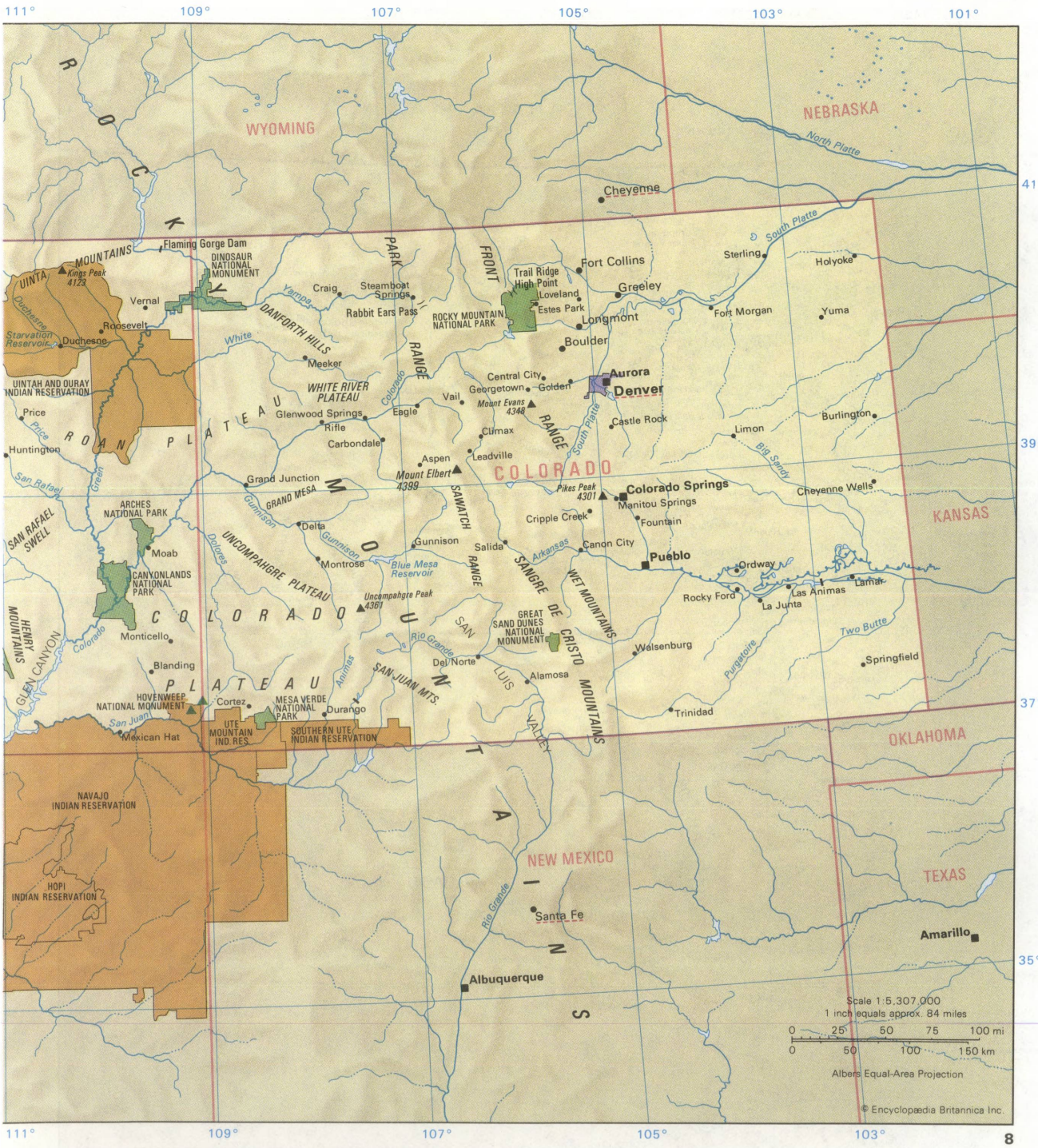
115°
113°
111°
109°
107°
105°
37°
35°
33°
31°
29°
27°
UTAH
COLORADO
NEVADA
CALIFORNIA
ARIZONA
NEW MEXICO
MEXICO
BAJA CALIFORNIA
Las Vegas
Lake Mead
Hoover Dam
Lake Mohave
Kingman
Bullhead
FORT MOJAVE IND. RES.
Lake Havasu
Lake Havasu City
COLORADO RIVER IND. RES.
FORT YUMA IND. RES.
Yuma
Somerton
Wellton
YUMA DESERT
ORGAN PIPE CACTUS NATIONAL MONUMENT
SONORAN DESERT
Gulf of California
Hermosillo
Ciudad Obregón
KAIBAB IND. RES.
Glen Canyon Dam
Page
NAVAJO INDIAN RESERVATION
Kayenta
Shiprock
Farmington
Aztec
Bloomfield
UTE MOUNTAIN IND. RES.
SOUTHERN UTE IND. RES.
Navajo Reservoir
Chama
Tierra Amarilla
Wheeler Peak 4011
TAOS IND. RES.
Taos
El Rito
Cerra Vista Peak 3839
Española
Los Alamos
Mora
JICARILLA APACHE IND. RES.
JEMEZ IND. RES.
COCHITI IND. RES.
Santa Fe
SANTO DOMINGO IND. RES.
Las Vegas
SAN FELIPE IND. RES.
ZIA IND. RES.
SANTA ANA IND. RES.
Bernalillo
Rio Rancho
Albuquerque
CANONCITO IND. RES.
ISLETA IND. RES.
Los Lunas
Belen
Estancia
Vaughn
Mountainair
ALAMO IND. RES.
ACOMA IND. RES.
LAGUNA IND. RES.
Milan
Grants
RAMAH NAVAJO IND. RES.
ZUNI IND. RES.
Gallup
Window Rock
NAVAJO IND. RES.
HOPI INDIAN RESERVATION
CANYON DE CHELLY NATIONAL MONUMENT
Tuba City
HAVASUPAI IND. RES.
Grand Canyon
GRAND CANYON NATIONAL PARK
HUALAPAI IND. RES.
Williams
Humphreys Peak 3851
Flagstaff
Kingman
Winslow
Holbrook
PETRIFIED FOREST NATIONAL PARK
Chino Valley
Cottonwood
Prescott
Sedona
Payson
Snowflake
St. Johns
Show Low
Eagar
FORT APACHE INDIAN RESERVATION
Baldy Peak 9553
Wickenburg
Phoenix
Sun City
Glendale
Scottsdale
Tempe
Mesa
Chandler
SALT RIVER IND. RES.
Globe
Superior
SAN CARLOS INDIAN RESERVATION
San Carlos Lake
GILA RIVER IND. RES.
Florence
Coolidge
Casa Grande
Gila Bend
MARICOPA IND. RES.
Clifton
Safford
Thatcher
Mount Graham 3267
Tucson
SAN XAVIER IND. RES.
PAPAGO INDIAN RESERVATION
Willcox
Benson
Tombstone
Tubac
Sierra Vista
Nogales
Bisbee
Douglas
Animas Peak 2601
Lordsburg
Deming
Silver City
Hatch
Las Cruces
University Park
Sunland Park
Anthony
El Paso
Ciudad Juárez
Truth or Consequences
Elephant Butte Reservoir
Carrizozo
Socorro
Magdalena
Ruidoso
MESCALERO APACHE IND. RES.
Alamogordo
WHITE SANDS NATIONAL MONUMENT
Roswell
Artesia
Dexter
Hagerman
PECOS PLAINS
Carlsbad
CARLSBAD CAVERNS NATIONAL PARK
Van Horn
Chihuahua
Rio Grande
Colorado
Virgin
Lake Powell
San Juan
Navajo Reservoir
Chama
Canadian
Pecos
Conchas Lake
Fort Sumner
Santa Rosa
Little Colorado
Verde
Salt
Gila
Black
Big Sandy
Bill Williams
Theodore Roosevelt Lake
Santa Cruz
San Pedro
Zuni
San Jose
Puerco
ROCKY MTS.
SANGRE DE CRISTO MTS.
MOGOLLON RIM
MOGOLLON MTS.
PELONCILLO MTS.
SAN ANDRES MTS.
SACRAMENTO MTS.
GUADALUPE MOUNTAINS NATIONAL P.
Guadalupe Peak 2667
BAYLOR MTS.
SIERRA VIEJA M.
PAINTED DESERT
COLORADO PLATEAU
© Encyclopædia Britannica, Inc.

KANSAS
MISSOURI
ARKANSAS
OKLAHOMA
TEXAS
LOUISIANA
HIGH PLAINS
LLANO ESTACADO
ESCARPMENT
ROCK CAP
WICHITA MTS.
ARBUCKLE MTS.
OUACHITA MOUNTAINS
EDWARDS PLATEAU
SANTIAGO MTS.
BIG BEND NATIONAL PARK
CHISOS MTS.
MATAGORDA PENINSULA
MATAGORDA ISLAND
ARANSAS NATIONAL WILDLIFE REFUGE
SAN JOSE ISLAND
MUSTANG ISLAND
PADRE ISLAND NATIONAL SEASHORE
PADRE ISLAND
Gulf of Mexico
Intracoastal Waterway
GALVESTON ISLAND
Amarillo
Lubbock
Abilene
Midland
Odessa
San Angelo
Wichita Falls
Oklahoma City
Tulsa
Norman
Edmond
Midwest City
Del City
Bethany
Lawton
Enid
Stillwater
Ponca City
Bartlesville
Pawhuska
Claremore
Owasso
Sand Springs
Broken Arrow
Muskogee
Bixby
Coweta
Tahlequah
Sallisaw
Poteau
McAlester
Ada
Ardmore
Durant
Hugo
Idabel
Paris
Sherman
Denison
Gainesville
Denton
McKinney
Plano
Garland
Richardson
Irving
Fort Worth
Arlington
Mesquite
Dallas
Greenville
Sulphur Springs
Mount Pleasant
Texarkana
Clarksville
Bonham
Cooper
Daingerfield
Longview
Marshall
Shreveport
Jefferson
Tyler
Kilgore
Henderson
Athens
Carthage
Jacksonville
Palestine
Rusk
Nacogdoches
Center
Lufkin
Crockett
Buffalo
Groesbeck
Waco
Hillsboro
Corsicana
Ennis
Waxahachie
Cleburne
Granbury
Stephenville
Glen Rose
Burleson
Lancaster
Kaufman
Canton
Eastland
Breckenridge
Mineral Wells
Weatherford
Albany
Anson
Snyder
Sweetwater
Colorado City
Big Spring
Andrews
Seminole
Hobbs
Lovington
Levelland
Brownfield
Tahoka
Post
Lamesa
Hamlin
Haskell
Graham
Seymour
Archer City
Henrietta
Vernon
Burkburnett
Quanah
Childress
Paducah
Crosbyton
Morton
Littlefield
Floydada
Plainview
Farwell
Muleshoe
Portales
Clovis
Bovina
Friona
Dimmitt
Tulia
Hereford
Canyon
Clarendon
Wellington
Memphis
Hollis
Altus
Granite
Hobart
Anadarko
Chickasha
Seminole
Wewoka
Holdenville
Wilburton
Stigler
Eufaula
Henryetta
Okemah
Shawnee
Tecumseh
Wetumka
Cushing
Sapulpa
Bristow
Kingfisher
Watonga
Guthrie
Chandler
Okmulgee
Broken Bow
Pine Creek Lake
Hugo Lake
Sardis Lake
Eufaula Lake
Robert S. Kerr Lake
Fort Gibson Lake
Grand Lake O the Cherokees
Oologah Lake
Keystone Lake
Arkansas
Kaw Lake
Newkirk
Blackwell
Medford
Alva
Cherokee
Fairview
Perry
Pawnee
Miami
Vinita
Jay
Pryor
Wagoner
Coweta
Fairview
Enid
Woodward
Fort Supply
Laverne
Beaver
Guymon
Boise City
Black Mesa 1516
Clayton
Perryton
Spearman
Dalhart
Dumas
Stinnett
Canadian
Borger
Pampa
Panhandle
Tucumcari
Lake Meredith
Canadian
Salt Fork
Cimarron
North Canadian
Canton Lake
Foss Lake
Elk City
Clinton
Weatherford
El Reno
North Fork
Washita
South Canadian
Lake Arrowhead
Lake Kemp
Lake Bridgeport
Lewisville Lake
Possum Kingdom Lake
Hubbard Creek Reservoir
Lake Tawakoni
Lake Fork Reservoir
Lake Texoma
Red
Brazos
Colorado
Leon
Lake Whitney
Trinity
Neches
Toledo Bend Reservoir
Sam Rayburn Reservoir
Lake Livingston
Lake Travis
Guadalupe
San Antonio
Nueces
Frio
Choke Canyon Lake
Amistad Reservoir
Rio Grande
Falcon Reservoir
Baffin Bay
Matagorda Bay
San Bernard
Colorado
Brazos
Sabine
Pecos
Devils
Black
Canyon
Lake Meredith
Lake Kemp
Lake Arrowhead
Lake Bridgeport
Lewisville Lake
Possum Kingdom Lake
Hubbard Creek Reservoir
Lake Tawakoni
Lake Fork Reservoir
Lake Texoma
Lake Whitney
Lake Livingston
Lake Travis
Choke Canyon Lake
Amistad Reservoir
Falcon Reservoir
Toledo Bend Reservoir
Sam Rayburn Reservoir
Wright Patman Lake
Lake Kemp
Lake Meredith
Fort Stockton
Alpine
Marfa
Pecos
Crane
Big Lake
Monahans
Kermit
Jal
Stanton
Midland
Odessa
Ballinger
Coleman
Santa Anna
Brownwood
Comanche
Goldthwaite
Gatesville
Marlin
Temple
Belton
Killeen
Copperas Cove
Lampasas
San Saba
Brady
Mason
Llano
Burnet
Cameron
Georgetown
Taylor
Round Rock
Austin
Johnson City
Fredericksburg
Junction
Sonora
Ozona
Eldorado
Kerrville
Boerne
New Braunfels
Seguin
Universal City
San Antonio
Leon Valley
Hondo
Uvalde
Del Rio
Eagle Pass
Crystal City
Carrizo Springs
Cotulla
Pearsall
Floresville
Lockhart
San Marcos
Bastrop
Giddings
Brenham
Hempstead
Washington
Conroe
Madisonville
Huntsville
Bryan
College Station
Caldwell
Kountze
Woodville
Livingston
Jasper
Vidor
Orange
Beaumont
Nederland
Port Arthur
Groves
Liberty
Baytown
Pasadena
Houston
Missouri City
Rosenberg
Columbus
La Grange
Bellville
Angleton
Lake Jackson
Freeport
Alvin
League City
Dickinson
Texas City
Galveston
Wharton
Hallettsville
El Campo
Cuero
Edna
Bay City
Victoria
Port Lavaca
Goliad
Karnes City
Refugio
George West
Beeville
Sinton
Portland
Robstown
Corpus Christi
Alice
San Diego
Freer
Kingsville
Hebbronville
Falfurrias
Laredo
Nuevo Laredo
Monclova
Monterrey
Raymondville
Rio Grande City
Edinburg
Mission
Reynosa
McAllen
San Benito
Harlingen
Brownsville
Matamoros
Scale 1:5,682,000
1 inch equals approx 90 miles
0 25 50 75 100 mi
0 50 100 150 Km
Albers Equal-Area Projection
103°
101°
99°
97°
95°
93°
37°
35°
33°
31°
29°
27°


119°
117°
115°
113°
111°
49°
47°
45°
43°
41°
C A N
WASHINGTON
OREGON
IDAHO
MONTAN
NEVADA
UTAH
Bonners Ferry
Priest Lake
Lake Koocanusa
Kootenai
Troy
Libby
Libby Dam
Columbia Falls
Whitefish
GLACIER NATIONAL PARK
BLACKFEET INDIAN RESERVATION
Cut Bank
Shelby
Sandpoint
Columbia
Lake Pend Oreille
Clark Fork
Kalispell
Hungry Horse Reservoir
Flathead Lake
Conrad
Lake Elwell
Spokane
Post Falls
Coeur d'Alene
Wallace
St. Maries
Coeur d'Alene Lake
FLATHEAD INDIAN RESERVATION
Polson
Ronan
Moiese
Flathead
Choteau
Teton
Fort Benton
Missouri
Sun
Great Falls
COEUR D'ALENE INDIAN RESERVATION
St. Joe
Rogers Pass
Missoula
Blackfoot
Dworshak Reservoir
Moscow
Snake
Bitterroot
Clark Fork
Helena
Canyon Ferry Dam
Canyon Ferry Lake
Townsend
Deer Lodge
Pierce
Orofino
Clearwater
Lewiston
NEZ PERCE INDIAN RESERVATION
Hamilton
SAPPHIRE MTS.
Anaconda
Butte
Three Forks
Selway
Grangeville
MOOSE RIDGE
Big Hole
PIONEER MOUNTAINS
Jefferson
Madison
Gallatin
Belgrade
Bozeman
Livingston
Salmon
HELLS CANYON
Salmon
Dillon
Virginia City
Ruby River Reservoir
Clark Canyon Reservoir
Quake Lake
Lake Hebgen
Lower Red Rock Lake
McCall
Middle Fork
Weiser
Cascade Reservoir
Challis
SAWTOOTH RANGE
LOST RIVER RANGE
Borah Peak 3859
Castle Peak 3603
Payette
Boise
Emmett
Idaho City
Caldwell
Boise
Nampa
Sun Valley
Rexburg
St. Anthony
TETON RANGE
Bellevue
Anderson Ranch Reservoir
CRATERS OF THE MOON NATIONAL MONUMENT
Idaho Falls
Jackson
Snake
SNAKE RIVER PLAIN
Blackfoot
FORT HALL INDIAN RESERVATION
OWYHEE MTS.
Silver City
Mountain Home
Glenns Ferry
Wood
American Falls Reservoir
Shoshone
Pocatello
SOUTH MOUNTAIN
Jerome
Shoshone Falls
American Falls
Afton
Twin Falls
Rupert
Burley
Montpelier
Malad City
Salmon Creek Reservoir
DUCK VALLEY INDIAN RESERVATION
WASATCH RANGE
Bear
Kemmerer
Great Salt Lake
Evanston
Salt Lake City
Scale 1:5,271,000
1 inch equals approx. 83 miles
0 25 50 75 100 mi
0 50 100 150 km
Albers Equal-Area Projection


109°
107°
105°
103°
101°
49°
47°
45°
43°
41°
A D A
Plentywood
FORT PECK INDIAN RESERVATION
Chinook
Havre
Milk
Malta
Glasgow
Wolf Point
Missouri
ROCKY BOY'S IND. RES.
BEAR PAW MTS.
FORT BELKNAP INDIAN RESERVATION
LITTLE ROCKY MTS.
Fort Peck Dam
Fort Peck Lake
Sidney
Winifred
CHARLES M. RUSSELL NATIONAL WILDLIFE RANGE
PINEY BUTTES
HIGHWOOD MOUNTAINS
NORTH DAKOTA
Judith
Box Elder
Lewistown
BIG SHEEP MTN.
Glendive
BIG SNOWY MTS.
Yellowstone
Roundup
Musselshell
Miles City
Baker
BULL MOUNTAINS
Forsyth
Tongue
Powder
Billings
Bighorn
Hardin
Laurel
Little Bighorn
NORTHERN CHEYENNE INDIAN RESERVATION
Broadus
BEARTOOTH RANGE
CROW INDIAN RESERVATION
Yellowtail Dam
BIGHORN CANYON
Red Lodge
Granite Peak 3901
Bighorn Lake
Clarks Fork
Powell
Lovell
Sheridan
BIGHORN MOUNTAINS
DEVIL'S TOWER NATIONAL MONUMENT
SOUTH DAKOTA
Shoshone
YELLOWSTONE NATIONAL PARK
Cody
Basin
Cloud Peak 4013
Buffalo
Gillette
Sundance
Belle Fourche
Moorcroft
Powder
Bighorn
Worland
Ten Sleep
Newcastle
GRAND TETON NATIONAL PARK
Togwotee Pass
Thermopolis
WIND RIVER INDIAN RESERVATION
WIND RIVER RANGE
Wind
Boysen Reservoir
WYOMING
Gannett Peak 4207
Fort Washakie
Riverton
North Platte
Casper
Glenrock
Douglas
Pinedale
Lander
GRANITE MTS.
Glendo Reservoir
Atlantic City
Sweetwater
Pathfinder Reservoir
LARAMIE MOUNTAINS
Sunrise
NEBRASKA
Fontenelle Reservoir
GREAT DIVIDE BASIN
Seminoe Reservoir
Laramie
Torrington
Wheatland
Rawlins
Medicine Bow
Green
Green River
Rock Springs
Saratoga
MEDICINE BOW MTS.
Laramie
Lyman
Flaming Gorge Reservoir
Bridger Peak 3355
North Platte
Cheyenne
COLORADO
© Encyclopædia Britannica Inc.
7


121°
119°
117°
115°
113°
41°
39°
37°
35°
OREGON
IDAHO
NEVADA
CALIFORNIA
UTAH
ARIZONA
BUCK VALLEY INDIAN RESERVATION
Mountain City
Jackpot
Tremonton
Smithfield
Logan
Hyrum
Brigham City
PROMONTORY MTS.
Great Salt Lake
Ogden
Weber
Morgan
Bountiful
Salt Lake City
West Valley
Tooele
Park City
Heber
Orem
Provo
Payson
WASATCH RANGE
OQUIRRH MTS.
Jordan
Utah Lake
Provo
Bonneville Salt Flats
GREAT SALT LAKE DESERT
Montello
Wells
INDEPENDENCE MTS.
BLACK ROCK DESERT
GRANITE RANGE
Humboldt
Winnemucca
Rye Patch Reservoir
Rock
Carlin
Elko
Battle Mountain
RUBY MOUNTAINS
Franklin Lake
Ruby Lake
Pine
PYRAMID LAKE INDIAN RESERVATION
Lovelock
Pyramid Lake
Winnemucca Lake
Carson Sink
Carson
Reese
STILLWATER
RANGE
TOIYABE RANGE
GREAT BASIN
EGAN RANGE
GOSHUTE INDIAN RESERVATION
DEEP CREEK RANGE
SIERRA
Sparks
Reno
Truckee
Fernley
Fallon
Virginia City
Carson City
Lahontan Reservoir
Lake Tahoe
Genoa
Gardnerville
Yerington
Walker
WALKER RIVER INDIAN RESERVATION
WASSUK RANGE
Walker Lake
Hawthorne
Gabbs
Mina
Austin
Arc Dome 3589
MONITOR RANGE
PANCAKE RANGE
White
Ruth
McGill
Ely
Wheeler Peak 3981
GREAT BASIN NATIONAL PARK
SEVIER DESERT
Delta
Manti
Castle Dale
Ferron
Fillmore
Sevier Lake
Sevier
Richfield
PAVANT RANGE
Mount Marvine 3536
Delano Peak 3710
WAH WAH MTS.
CAPITOL REEF NATIONAL PARK
AQUARIUS PLATEAU
Boundary Peak 4006
Tonopah
Goldfield
Pioche
Panaca
Caliente
ESCALANTE DESERT
Cedar City
BRYCE CANYON NATIONAL PARK
KAIPAROWITS PLATEAU
MARKAGUNT PLATEAU
Bryce Gorge
ZION NATIONAL PARK
Orderville
Kanab
Lake Powell
Virgin
Meadow Valley
PAHUTE MESA
Alamo
Washington
St. George
Beatty
Mercury
Mesquite
Muddy
SHEEP RANGE
Logandale
Overton
DEATH VALLEY NATIONAL MONUMENT
Las Vegas
North Las Vegas
Henderson
Boulder City
Hoover Dam
Lake Mead
Colorado
SPRING MOUNTAINS
Searchlight
Davis Dam
Fresno
Bakersfield
SIERRA NEVADA
Los Angeles


111°
109°
107°
105°
103°
101°
41°
39°
37°
35°
WYOMING
NEBRASKA
KANSAS
OKLAHOMA
TEXAS
NEW MEXICO
COLORADO
ROCKY MOUNTAINS
PARK RANGE
FRONT RANGE
MOUNT EVANS RANGE
SAWATCH RANGE
SANGRE DE CRISTO MOUNTAINS
WET MOUNTAINS
SAN JUAN MTS.
UNCOMPAHGRE PLATEAU
GRAND MESA
WHITE RIVER PLATEAU
ROAN PLATEAU
COLORADO PLATEAU
UINTA MOUNTAINS
DANFORTH HILLS
HENRY MOUNTAINS
SAN RAFAEL SWELL
GLEN CANYON
SAN LUIS VALLEY
Cheyenne
North Platte
South Platte
Sterling
Holyoke
Fort Collins
Trail Ridge
High Point
Greeley
Loveland
Estes Park
Longmont
Boulder
Fort Morgan
Yuma
Flaming Gorge Dam
DINOSAUR NATIONAL MONUMENT
Kings Peak 4123
Vernal
Craig
Steamboat Springs
Yampa
Rabbit Ears Pass
ROCKY MOUNTAIN NATIONAL PARK
Roosevelt
Duchesne
Starvation Reservoir
White
Meeker
Central City
Georgetown
Golden
Aurora
Denver
Vail
Mount Evans 4348
Eagle
Glenwood Springs
Rifle
Carbondale
Climax
Leadville
Aspen
Mount Elbert 4399
Castle Rock
Limon
Burlington
Price
Huntington
Grand Junction
Pikes Peak 4301
Colorado Springs
Manitou Springs
Cheyenne Wells
Big Sandy
Cripple Creek
Fountain
Canon City
Pueblo
Ordway
Rocky Ford
La Junta
Las Animas
Lamar
Two Butte
Springfield
Walsenburg
Trinidad
Purgatoire
ARCHES NATIONAL PARK
Moab
Delta
Gunnison
Montrose
Salida
Blue Mesa Reservoir
Uncompahgre Peak 4361
CANYONLANDS NATIONAL PARK
Monticello
Blanding
GREAT SAND DUNES NATIONAL MONUMENT
Del Norte
Alamosa
HOVENWEEP NATIONAL MONUMENT
Cortez
MESA VERDE NATIONAL PARK
Durango
Mexican Hat
UTE MOUNTAIN IND. RES.
SOUTHERN UTE INDIAN RESERVATION
UINTAH AND OURAY INDIAN RESERVATION
NAVAJO INDIAN RESERVATION
HOPI INDIAN RESERVATION
Santa Fe
Albuquerque
Amarillo
Rio Grande
Arkansas
Colorado
Gunnison
Dolores
San Juan
Animas
Green
Price
San Rafael
Scale 1:5,307,000
1 inch equals approx. 84 miles
0 25 50 75 100 mi
0 50 100 150 km
Albers Equal-Area Projection
© Encyclopædia Britannica Inc.
8

127°
125°
123°
121°
119°
117°
115°
49°
47°
45°
43°
41°
CANADA
Strait of Georgia
Vancouver
Lynden
Ferndale
Bellingham
NORTH CASCADES NATIONAL PARK
Colville
Victoria
Anacortes
Sedro Woolley
Burlington
Mount Vernon
Omak
Okanogan
COLVILLE INDIAN RESERVATION
Chewelah
Cape Flattery
Strait of Juan de Fuca
Puget Sound
Oak Harbour
Port Angeles
Sequim
Port Townsend
Franklin D. Roosevelt Lake
SPOKANE INDIAN RESERVATION
Grand Coulee
Coulee Dam
Grand Coulee Dam
MONTANA
Forks
Everett
Edmonds
CASCADE RANGE
Lake Chelan
Banks Lake
Columbia
Chelan
Spokane
OLYMPIC MTS.
OLYMPIC NATIONAL PARK
Poulsbo
Lake Washington
Seattle
Kirkland
Bellevue
Renton
Bremerton
Port Orchard
Davenport
Spokane
Medical Lake
Opportunity
Cheney
Snoqualmie
Wenatchee
Ephrata
QUINAULT IND. RES.
Kent
Auburn
Tacoma
Puyallup
Shelton
Olympia
Lacey
Parkland
Yakima
Quincy
Ritzville
Moses Lake
Hoquiam
Montesano
Tumwater
Aberdeen
Grays Harbor
Mount Rainier 4392
MOUNT RAINIER NATIONAL PARK
Ellensburg
Othello
Colfax
Pullman
Willapa Bay
Raymond
Centralia
Chehalis
Selah
Yakima
Union Gap
Wapato
YAKIMA INDIAN RESERVATION
SADDLE MTS.
WASHINGTON
Snake
Pomeroy
Clarkston
Cowlitz
Cape Disappointment
Mount Saint Helens 2550
Toppenish
Sunnyside
Richland
Dayton
Kelso
Longview
Astoria
Seaside
MOUNT ST. HELENS NATIONAL VOLCANIC MONUMENT
Grandview
Prosser
Pasco
Kennewick
College Place
Walla Walla
Milton-Freewater
Columbia
Goldendale
Umatilla
Hermiston
Stanfield
WALLOWA MTS.
St. Helens
Woodland
Bonneville Dam
Hood River
Pendleton
PACIFIC OCEAN
Vancouver
Camas
Portland
Washougal
Gresham
The Dalles
John Day Lock and Dam
UMATILLA INDIAN RESERVATION
Heppner
Enterprise
Tillamook
Hillsboro
Beaverton
Lake Oswego
Milwaukie
Oregon City
Newberg
Mount Hood 3426
La Grande
BLUE MOUNTAINS
HELLS CANYON
IDAHO
McMinnville
Sheridan
Hubbard
Woodburn
Keizer
Lincoln City
Dallas
Salem
Independence
WARM SPRINGS INDIAN RESERVATION
Powder
Baker
Cape Foulweather
Newport
Albany
Mount Jefferson 3199
John Day
Madras
Snake
Toledo
Corvallis
Lebanon
Philomath
Marys Peak 1249
Sweet Home
Prineville
John Day
Harrisburg
Redmond
Ontario
Florence
Veneta
Eugene
Springfield
Bend
Deschutes
Malheur
Cottage Grove
Reedsport
OREGON
Boise
Burns
North Bend
Sutherlin
HARNEY
Malheur Lake
Lake Owyhee
Coos Bay
Coquille
Umpqua
Roseburg
Harney Lake
Bandon
CRATER LAKE NATIONAL PARK
Crater Lake
Jordan Valley
Cape Blanco
Riddle
Summer Lake
BASIN
Owyhee
Lake Abert
Rogue
Grants Pass
Eagle Point
Gold Beach
Medford
Upper Klamath Lake
Jacksonville
Ashland
Klamath Falls
Lakeview
TROUT CREEK MTS.
Brookings
Klamath
CALIFORNIA
NEVADA
Scale 1:5,708,000
1 inch equals approx. 90 miles
0 25 50 75 100 mi
0 50 100 150 km
Albers Equal-Area Projection
©Encyclopædia Britannica Inc.

OREGON
IDAHO
NEVADA
UTAH
CALIFORNIA
ARIZONA
MEXICO
PACIFIC OCEAN
GREAT BASIN
SIERRA NEVADA
COAST RANGES
SACRAMENTO VALLEY
SAN JOAQUIN VALLEY
MOJAVE DESERT
KLAMATH MOUNTAINS
WARNER MOUNTAINS
TEHACHAPI MTS.
DEATH VALLEY
SALTON SEA
IMPERIAL VALLEY
CHOCOLATE MTS.
SAN BERNARDINO MTS.
Crescent City
Eureka
Arcata
Fortuna
Redding
Anderson
Red Bluff
Chico
Oroville
Willows
Fort Bragg
Ukiah
Lakeport
Point Arena
Cape Mendocino
Santa Rosa
Petaluma
Napa
Fairfield
Vallejo
Novato
Richmond
Berkeley
Oakland
San Francisco
San Leandro
Hayward
San Mateo
Redwood City
Palo Alto
Mountain View
Sunnyvale
San Jose
Santa Clara
Saratoga
Concord
Walnut Creek
Pittsburg
Antioch
Stockton
Lodi
Manteca
Tracy
Modesto
Turlock
Merced
Mariposa
Sonora
Woodland
Davis
Sacramento
Roseville
Auburn
Placerville
Jackson
Grass Valley
Nevada City
Marysville
Yuba City
Colusa
Susanville
Reno
Carson City
South Lake Tahoe
Bishop
Mammoth Lakes
Santa Cruz
Monterey
Pacific Grove
Carmel
Point Sur
Salinas
Hollister
Gilroy
San Simeon
Atascadero
San Luis Obispo
Lompoc
Santa Maria
Santa Barbara
Ventura
Oxnard
Los Angeles
Santa Monica
Glendale
Pasadena
Burbank
San Fernando
Simi Valley
Pomona
Downey
Inglewood
Torrance
Long Beach
Huntington Beach
Fullerton
Anaheim
Santa Ana
Costa Mesa
San Clemente
Ontario
Riverside
San Bernardino
Victorville
Barstow
Lancaster
Palmdale
Needles
Las Vegas
Ridgecrest
Bakersfield
Delano
Wasco
Porterville
Tulare
Visalia
Hanford
Kings City
Fresno
Selma
Sanger
Clovis
Madera
Los Banos
Oceanside
Carlsbad
Escondido
San Diego
La Mesa
El Cajon
National City
Chula Vista
Tijuana
Ensenada
Mexicali
Calexico
Brawley
El Centro
Blythe
Indio
Palm Springs
Cathedral City
Coachella
Yosemite National Park
Sequoia National Park
Kings Canyon National Park
Death Valley National Monument
Lassen Volcanic National Park
Lava Beds National Monument
Redwood National Park
Point Reyes National Seashore
Channel Islands National Park
Joshua Tree National Monument
Santa Barbara Channel
San Miguel Island
Santa Cruz I.
Santa Rosa I.
Santa Barbara I.
Santa Catalina Island
San Clemente Island
Mount Shasta 4317
Lassen Peak 3187
Mount Whitney 4418
Mount Ritter 4010
Mount Morgan 4190
Clear Lake
Lake Tahoe
Mono Lake
Owens Lake
Lake Mead
Goose Lake
Eagle Lake
Honey Lake
Lake Almanor
Clear Lake Res.
Trinity
Klamath
Sacramento
Feather
Eel
Merced
Tuolumne
Stanislaus
Kings
Salinas
Pit
Humboldt
White Mts.
Owens
Colorado River Aqueduct
California Aqueduct
Los Angeles Aqueduct
Friant-Kern Canal
Coachella Canal
Colorado
Mexicali
Tijuana
Ensenada
Scale 1:5,708,000
1 inch equals approx. 90 miles
0 25 50 75 100 mi
0 50 100 150 km
Albers Equal-Area Projection
©Encyclopædia Britannica Inc.
125°
123°
121°
119°
117°
115°
113°
43°
41°
39°
37°
35°
33°

Scale 1:6,336,000
1 inch equals approx 100 miles
0 25 50 75 100 mi
0 50 100 150 km
Albers Equal-Area Projection
Scale 1:15,840,000
1 inch equals approx 250 miles
0 50 100 150 200 250 mi.
0 100 200 300 400 km.
Albers Equal-Area Projection
PACIFIC OCEAN
ARCTIC OCEAN
BEAUFORT SEA
BERING SEA
Gulf of Alaska
HAWAII
NIIHAU
KAUAI
OAHU
MOLOKAI
MAUI
LANAI
KAHOOLAWE
Honolulu
HALEAKALA NATIONAL PARK
HAWAII VOLCANOES NATIONAL PARK
ALASKA
CANADA
U.S.S.R.
BROOKS RANGE
ALASKA RANGE
KUSKOKWIM MOUNTAINS
WRANGELL MTS.
CHUGACH RANGE
ST. ELIAS MOUNTAINS
ALEUTIAN RANGE
ALEUTIAN ISLANDS
ANDREANOF ISLANDS
ISLANDS OF THE FOUR MOUNTAINS
SEWARD PENINSULA
ST. LAWRENCE ISLAND
ST. MATTHEW ISLAND
NUNIVAK ISLAND
PRIBILOF ISLANDS
KENAI PENINSULA
ALASKA PENINSULA
KODIAK ISLAND
CHIRIKOF ISLAND
TRINITY ISLANDS
MATANUSKA VALLEY
Anchorage
Fairbanks
Juneau
Skagway
Haines
Sitka
Ketchikan
Barrow
Wainwright
Point Hope
Kotzebue
Nome
Savoonga
Bethel
Dillingham
Kodiak
Unalaska
Dutch Harbor
Attu
Shishaldin Volcano
Mount McKinley 6194
Mount St. Elias 5489
Cordova Peak 2356
Kama Kou 1512
Puu Kukui 1764
Red Hill 3055
Mauna Kea 4205
Mauna Loa 4169
Kilauea Crater 1248
Kawaihae Bay
Kauai Channel
Alenuihaha Channel
Pearl Harbor
Kaiwi Channel
Upolu Point
Kauna Point
Ka Lae
Point Hope
Barrow
Wainwright
Prudhoe Bay
Colville
Porcupine
Arctic Circle
Mackenzie
Yukon
Koyukuk
Tanana
Innoko
Kuskokwim
Norton Sound
Bering Strait
Kotzebue Sound
Hooper Bay
Chevak
Emmonak
Alakanuk
Unalakleet
Galena
Nenana
Anderson
North Pole
Delta Junction
Fort Yukon
Whitehorse
Valdez
Cordova
Seward
Homer
Seldovia
Soldotna
Kenai
Palmer
Wasilla
Houston
Willow
Glennallen
Tok
Togiak
Kwethluk
Mountain Village
St. Marys
Savoonga
Nome
Selawik
Noorvik
Kaktovik
Atka
Adak
Amchitka Island
Agattu I.
Attu I.
Kiska I.
Unimak I.
King Cove
Sand Point
Unalaska I.
Umnak I.
Togiak
Bristol Bay
Kuskokwim Bay
Lake Clark
Iliamna Lake
Becharof Lake
Shelikof Strait
Three Saints Bay
Cook Inlet
Prince William Sound
Chatham Strait
Dixon Entrance
ALEXANDER ARCHIPELAGO
GLACIER BAY NATIONAL PARK AND PRESERVE
ADMIRALTY ISLAND NATIONAL MONUMENT
MISTY FJORDS NATIONAL MONUMENT
KOBUK VALLEY NATIONAL PARK
NOATAK NATIONAL PRESERVE
CAPE KRUSENSTERN NATIONAL MONUMENT
BERING LAND BRIDGE NATIONAL PRESERVE
GATES OF THE ARCTIC NATIONAL PARK AND PRESERVE
VENETIE INDIAN RESERVATION
YUKON-CHARLEY RIVERS NATIONAL PRESERVE
DENALI NATIONAL PARK AND PRESERVE
WRANGELL-ST. ELIAS NATIONAL PARK AND PRESERVE
LAKE CLARK NATIONAL PARK AND PRESERVE
KATMAI NATIONAL PARK AND PRESERVE
KENAI FJORDS NATIONAL PARK
ANIAKCHAK NATIONAL MONUMENT AND PRESERVE
©Encyclopædia Britannica Inc.

character not only for the Upper Midwest, from Ohio to the Dakotas, but also in the Pacific Northwest in general, though to a lesser degree.

The South. By far the largest of the three original Anglo-American culture areas, the South is also the most idiosyncratic with respect to national norms—or slowest to accept them. The South was once so distinct from the non-South in almost every observable or quantifiable feature and so fiercely proud of its peculiarities that for some years the question of whether it could maintain political and social unity with the non-South was in serious doubt. These differences are still observable in almost every realm of human activity, including rural economy, dialect, diet, costume, folklore, politics, architecture, social customs, and recreation. Only during the 20th century can an argument be made that it has achieved a decisive convergence with the rest of the nation, at least in terms of economic behaviour and material culture.

Southern deviation from the mainstream

A persistent deviation from the national mainstream probably began in the first years of settlement. The first settlers of the South were almost purely British, not outwardly different from those who flocked to New England or the Midland, but almost certainly distinct in terms of motives and social values and more conservative in retaining the rurality and the family and social structure of premodern Europe. The vast importation of African slaves was also a major factor, as was a degree of contact with the Indians that was less pronounced farther north. In addition, the unusual pattern of economy (much different from that of northwestern Europe), settlement, and social organization, which were in part an adaptation to a starkly unfamiliar physical habitat, accentuated the South's deviation from other culture areas.

In both origin and spatial structure, the South has been characterized by diffuseness. In the search for a single cultural hearth, the most plausible choice is the Chesapeake Bay area and the northeastern corner of North Carolina, the earliest area of recognizably Southern character. Early components of Southern population and culture also arrived from other sources. A narrow coastal strip from North Carolina to the Georgia–Florida border and including the Sea Islands is decidedly Southern in character, yet it stands apart self-consciously from other parts of the South. Though colonized directly from Great Britain, it had also significant connections with the West Indies, in which relation the African cultural contribution was strongest and purest. Charleston and Savannah, which nurtured their own distinctive civilizations, dominated this subregion. Similarly, French Louisiana received elements of culture and population—to be stirred into the special Creole mixture—not only, putatively, from the Chesapeake Bay hearth area but also indirectly from France, French Nova Scotia, the French West Indies, and Africa. In south central Texas, the Germanic and Hispanic influx was so heavy that a special subregion can be designated.

It would seem, then, that the Southern culture area may be an example of convergent, or parallel, evolution of a variety of elements arriving along several paths but subject to some single general process that could mold one larger regional consciousness and way of life.

Because of its slowness in joining the national technological mainstream, the South can be subdivided into a much greater number of subregions than is possible for any of the other older traditional regions. Those described above are of lesser order than the two principal Souths, variously called Upper and Lower (or Deep) South, Upland and Lowland South, or Yeoman and Plantation South.

The Upland South, which comprises the southern Appalachians, the upper Appalachian Piedmont, the Cumberland and other low interior plateaus, and the Ozarks and Ouachitas, was colonized culturally and demographically from the Chesapeake Bay hearth area and the Midland; it is most emphatically white Anglo-Saxon Protestant (WASP) in character. The latter area, which contains a large black population, includes the greater part of the South Atlantic and Gulf coastal plains and the lower Appalachian Piedmont. Its early major influences came from the Chesapeake Bay area, with only minor elements from the coastal Carolina–Georgia belt, Louisiana, and elsewhere. The division between the two subregions remains distinct from Virginia to Texas, but each region can be further subdivided. Within the Upland South, the Ozark region might legitimately be detached from the Appalachian; and, within the latter, the proud and prosperous Kentucky Bluegrass, with its emphasis on tobacco and Thoroughbreds, certainly merits special recognition.

Toward the margins of the South, the difficulties in delimiting subregions become greater. The outer limits themselves are a topic of special interest. There seems to be more than an accidental relation between these limits and various climatic factors. The fuzzy northern boundary, definitely not associated with the conventional Mason and Dixon Line or the Ohio River, seems most closely associated with length of frost-free season or with temperature during the winter. As the Southern cultural complex was carried to the West, it not only retained its strength but became more intense, in contrast to the influence of New England and the Midland. But the South finally fades away as one approaches the 100th meridian, with its critical decline in annual precipitation. The apparent correlation between the cultural South and a humid subtropical climatic regime is in many ways valid.

Southern culture and climate

The Texas subregion is so large, distinctive, vigorous, and self-assertive that it presents some vexing classificatory questions. Is Texas simply a subregion of the Greater South, or has it acquired so strong and divergent an identity that it can be regarded as a major region in its own right? It is likely that a major region has been born in a frontier zone in which several distinct cultural communities confront one another and in which the mixture has bred the vigorous, extroverted, aggressive Texas personality so widely celebrated in song and story. Similarly, peninsular Florida may be considered either within or juxtaposed to the South but not necessarily part of it. In the case of Florida, an almost empty territory began to receive significant settlement only after about 1890, and if, like Texas, most of it came from the older South, there were also vigorous infusions from elsewhere.

The Midland. The significance of this region has not been less than that of New England or the South, but its characteristics are the least conspicuous to outsiders as well as to its own residents—reflecting, perhaps, its centrality in the course of U.S. development. The Midland (a term not to be confused with Midwest) comprises portions of Middle Atlantic and Upper Southern states: Pennsylvania, New Jersey, Delaware, and Maryland. Serious European settlement of the Midland began a generation or more after that of the other major cultural centres and after several earlier, relatively ineffectual trials by the Dutch, Swedes, Finns, and British. But once begun late in the 17th century by William Penn and his associates, the colonization of the area was a success. Within southeastern Pennsylvania this culture area first assumed its distinctive character: a prosperous, sober, industrious agricultural society that quickly became a mixed economy as mercantile and later industrial functions came to the fore. By the mid-18th century much of the region had acquired a markedly urban character, resembling in many ways the more advanced portions of the North Sea countries. In this respect, at least, the Midland was well ahead of neighbouring areas to the north and south.

It differed also in its polyglot ethnicity. From almost the beginning, the various ethnic and religious groups of the British Isles were joined by immigrants from the European mainland. This diversity has grown and is likely to continue. The mosaic of colonial ethnic groups has persisted in much of Pennsylvania, New York, New Jersey, and Maryland, as has the remarkable variety of nationalities and churches in coalfields, company towns, cities, and many rural areas. Much of the same ethnic heterogeneity can be seen in New England, the Midwest, and a few other areas, but the Midland stands out as perhaps the most polyglot region of the nation. The Germanic element has always been notably strong, if irregularly distributed, in the Midland, accounting for more than 70 percent of the population of many towns. Had the Anglo-American culture not triumphed, the area might well have been designated Pennsylvania German.

Polyglot nature of the Midland culture

Physiography and migration carried the Midland culture area into the Maryland Piedmont. Although its width tapers quickly below the Potomac, it reaches into parts of Virginia and West Virginia, with traces legible far down the Appalachian zone and into the South.

The northern half of the greater Midland region (the New York subregion, or New England Extended) cannot be assigned unequivocally to either New England or this Midland. Essentially it is a hybrid formed mainly from two regional strains of almost equal strength: New England and the post-1660 British element moving up the Hudson valley and beyond. There has also been a persistent, if slight, residue of early Dutch culture and some subtle filtering northward of Pennsylvanian influences. Apparently within the New York subregion occurred the first major fusion of American regional cultures, especially within the early 19th-century "Burned-Over District," around the Finger Lakes and Genesee areas of central and western New York. This locality, the seedbed for a number of important social innovations, was a major staging area for westward migration and possibly a major source for the people and notions that were to build the Midwestern culture area.

Toward the west the Midland retains its integrity for only a short distance—certainly no further than eastern Ohio—as it becomes submerged within the Midwest. Still, its significance in the genesis of the Midwest and the national culture should not be minimized. Its success in projecting its image upon so much of the country may have drawn attention away from the source area. As both name and location suggest, the Midland is intermediate in character in many respects, lying between New England and the South. Its residents are much less concerned with, or conscious of, a strong regional identity (excepting the Pennsylvania Dutch caricatures) than is true for the other regions, and, in addition, the Midland lacks their strong political and literary traditions, though it is unmistakable in its distinctive townscapes and farmsteads.

The newer culture areas. *The Midwest.* There is no such self-effacement in the Midwest, that large triangular region justly regarded as the most nearly representative of the national average. Everyone within or outside of the Midwest knows of its existence, but no one is certain where it begins or ends. The older apex of the eastward-pointing triangle appears to rest around Pittsburgh, while the two western corners melt away somewhere in the Great Plains, possibly in southern Manitoba in the north and southern Kansas in the south. The eastern terminus and the southern and western borders are broad, indistinct transitional zones.

Serious study of the historical geography of the Midwest began only in the 20th century, but it seems likely that this culture region was the combination of all three colonial regions and that this combination first took place in the upper Ohio valley. The early routes of travel—the Ohio and its tributaries, the Great Lakes, and the low, level corridor along the Mohawk and the coastal plains of Lake Ontario and Lake Erie—converge upon Ohio. There, the people and cultural traits from New England, the Midland, and the South were first funneled together. There seems to have been a fanlike widening of the new hybrid area into the West as settlers worked their way frontierward.

Contrasts of Upper and Lower Midwest

Two major subregions are readily discerned, the Upper and Lower Midwest. They are separated by a line, roughly approximating the 41st parallel, that persists as far west as Colorado in terms of speech patterns and indicates differences in regional provenance in ethnic and religious terms as well. Much of the Upper Midwest retains a faint New England character, although Midland influences are probably as important. A rich mixture of German, Scandinavian, Slavic, and other non-WASP elements has greatly diversified a stock in which the British element usually remains dominant and the range of church denominations is great. The Lower Midwest, except for the relative scarcity of blacks, tends to resemble the South in its predominantly Protestant and British makeup. There are some areas with sizable Roman Catholic and non-WASP populations, but on the whole the subregion tends to be more WASP in inclination than most other parts of the nation.

The problem of "the West." The foregoing culture areas account for roughly the eastern half of the coterminous United States. There is a dilemma in classifying the remaining half. The concept of the American West, strong in the popular imagination, is reinforced constantly by romanticized cinematic and television images of the cowboy. It is facile to accept the widespread Western livestock complex as epitomizing the full gamut of Western life, because although the cattle industry may have once accounted for more than one-half of the active Western domain as measured in acres, it employed only a relatively small fraction of the total population. As a single subculture, it cannot represent the total regional culture.

It is not clear whether there is a genuine, single, grand Western culture region. Unlike the East, where virtually all the land is developed and culture areas and subregions abut and overlap in splendid confusion, the eight major and many lesser nodes of population in the western United States resemble oases, separated from one another by wide expanses of nearly unpopulated mountain or arid desert. The only obvious properties these isolated clusters have in common are, first, the intermixture of several strains of culture, primarily from the East but with additions from Europe, Mexico, and East Asia, and, second, except for one subregion, a general modernity, having been settled in a serious way no earlier than the 1840s. Some areas may be viewed as inchoate, or partially formed, cultural entities; the others have acquired definite personalities but are difficult to classify as first-order or lesser order culture areas.

There are several major tracts in the western United States that reveal a genuine cultural identity: the Upper Rio Grande region, the Mormon region, southern California, and, by some accounts, northern California. To this group one might add the anomalous Texan and Oklahoman subregions, which have elements of both the West and the South.

Upper Rio Grande and Mormon regions

The term Upper Rio Grande region was coined to denote the oldest and strongest of the three sectors of Hispanic-American activity in the Southwest, the others being southern California and portions of Texas. Although covering the valley of the upper Rio Grande, the region also embraces segments of Arizona and Colorado as well as other parts of New Mexico. European communities and culture have been present there, with only one interruption, since the late 16th century. The initial sources were Spain and Mexico, but after 1848 at least three distinct strains of Anglo-American culture were increasingly well represented—the Southern, Mormon, and a general undifferentiated Northeastern culture—plus a distinct Texan subcategory. For once this has occurred without obliterating the Indians, whose culture endures in various stages of dilution, from the strongly Americanized or Hispanicized to the almost undisturbed.

The general mosaic is a fabric of Indian, Anglo, and Hispanic elements, and all three major groups, furthermore, are complex in character. The Indian component is made up of Navajo, Pueblo, and several smaller groups, each of which is quite distinct from the others. The Hispanic element is also diverse—modally Mexican mestizo, but ranging from pure Spanish to nearly pure pre-Spanish aboriginal.

The Mormon region is expansive in the religious and demographic realms, though it has ceased to expand territorially as it did in the decades after the first settlement in the Salt Lake valley in 1847. Despite its Great Basin location and an exemplary adaptation to environmental constraints, this cultural complex appears somewhat non-Western in spirit: the Mormons may be in the West, but they are not entirely of it. Their historical derivation from the Midwest and from ultimate sources in New York and New England is still apparent, along with the generous admixture of European converts to their religion.

As in New England, the power of the human will and an intensely cherished abstract design have triumphed over an unfriendly habitat. The Mormon way of life is expressed in the settlement landscape and economic activities within a region more homogeneous internally than any other U.S. culture area.

Northern and southern California

In contrast, northern California has yet to gain its own strong cultural coloration. From the beginning of the great 1849 gold rush the area drew a diverse population from Europe and Asia as well as the older portions of the United States. Whether the greater part of northern California has produced a culture amounting to more than the sum of the contributions brought by immigrants is questionable. San Francisco, the regional metropolis, may have crossed the qualitative threshold. An unusually cosmopolitan outlook that includes an awareness of the Orient stronger than that of any other U.S. city, a fierce self-esteem, and a unique townscape may be symptomatic of a genuinely new, emergent local culture.

Southern California is the most spectacular of the Western regions, not only in terms of economic and population growth but also for the luxuriance, regional particularism, and general avant-garde character of its swiftly evolving cultural pattern. Until the coming of a direct transcontinental rail connection in 1885, the region was remote, rural, and largely inconsequential. Since then, the invasion by persons from virtually every corner of North America and by the world has been massive, but since the 1960s inmigration has slackened perceptibly, and many residents have begun to question the doctrine of unlimited growth. In any event, a loosely articulated series of urban and suburban developments continue to encroach upon what little is left of arable or habitable land in the Coast Ranges and valleys from Santa Barbara to the Mexican border.

Although every major ethnic and racial group and every other U.S. culture area is amply represented in southern California, there is reason to suspect that a process of selection for certain types of people, attitudes, and personality traits may have been at work at both source and destination. The region is distinct from, or perhaps in the vanguard of, the remainder of the nation. One might view southern California as the super-American region or the outpost of a postindustrial future, but its cultural distinctiveness is very evident in landscape and social behaviour. Southern California in no way approaches being a "traditional region," or even the smudged facsimile of such, but rather the largest, boldest experiment in creating a "voluntary region," one built through the self-selection of immigrants and their subsequent interaction.

The remaining identifiable Western regions—the Willamette valley of Oregon, the Puget Sound region, the Inland Empire of eastern Washington and adjacent tracts of Idaho and Oregon, central Arizona, and the Colorado Piedmont—can be treated jointly as potential, or emergent, culture areas, still too close to the national mean to display any cultural distinctiveness. In all of these regions is evident the arrival of a cross section of the national population and the growth of regional life around one or more major metropolises. A New England element is noteworthy in the Willamette valley and Puget Sound regions, while a Hispanic-American component appears in the Colorado Piedmont and central Arizona. Only time and further study will reveal whether any of these regions, so distant from the historic sources of U.S. population and culture, have the capacity to become an independent cultural area. (W.Ze.)

The people

As the 20th century drew to a close, the majority of the U.S. population had achieved a high level of material comfort, prosperity, and security. Americans were not, however, prepared to cope with the unexpected problems of relative affluence or with the persistent difficulties created by residual pockets of poverty. Crime, racial conflict, urban decay, proliferation of nuclear weapons, pollution of the environment, drug abuse, and rising costs of living remained continuing subjects of concern.

Many Americans perceive social tensions as the products of their society's failure to extend the traditional dream of equality of opportunity to all the people. Ideally, social, political, economic, and religious freedom would assure the like treatment of all persons, so that all could achieve goals in accord with their individual talents, if only they worked hard enough. A shared belief in this idea is the strongest bond that has united Americans through the centuries. The fact that some ethnic groups have not achieved full equality troubles many citizens.

ETHNIC DISTRIBUTION

The ethnics. Although current usage confines the term "ethnic" to the descendants of the newest immigrants, its proper, more comprehensive meaning applies to all groups unified by their cultural heritage and by their experience in the New World.

In the 19th century, Yankees formed one such group, marked by common religion and by habits shaped by the original Puritan settlers. From New England, the Yankees spread westward through New York, northern Ohio, Indiana, and Illinois and on to Iowa and Kansas. Tightly knit communities, firm religious values, and a belief in the value of education earned them prominent positions in business, in literature and law, and in cultural and philanthropic institutions. They long identified with the Republican Party.

Southern whites and their descendants, by contrast, generation after generation remained preponderantly rural as migration took them westward across Tennessee and Kentucky to Arkansas, Missouri, Oklahoma, and Texas. These people remained primarily rural until the industrialization of the South in the 20th century, and they preserved affiliations with the Democratic Party until the 1960s.

German immigrant groups

The colonial population also contained other elements that long sustained their identities as groups. The Pennsylvania Germans, held together by religion and language, still pursue their own way of life after three centuries. The great 19th-century German migrations, however, contained a variety of elements that dispersed in the cities as well as in the agricultural areas of the West; to the extent that ethnic ties have survived they are largely sentimental. That is also true of the Scots, Scotch-Irish, Welsh, and Dutch, whose colonial nuclei received some reinforcement after 1800 but who gradually adapted to the ways of the larger groups among whom they lived.

Distinctive language and religion have preserved some coherence among the descendants of the Norwegian, Danish, Swedish, Finnish, and Icelandic newcomers of the 19th century. Where these people clustered in sizeable settlements, as in Minnesota, they transmitted a sense of identity beyond the second generation; and emotional attachments to the lands of origin lingered much longer.

Religion was a powerful force for cohesion among the Roman Catholic Irish and the Jews, both tiny groups before 1840, both reinforced by mass migration thereafter. Both have now become strikingly heterogeneous, displaying a wide variety of economic and social conditions as well as a degree of conformity to the styles of life of other Americans. But the pull of external concerns—in the one case, unification of Ireland; in the other, Israel's security—have helped to preserve group loyalty.

More recently established ethnic groups have preserved greater visibility and greater cohesion. Indeed, by the 1970s and '80s, "ethnic" had come to be used to describe the Americans of Polish, Italian, Lithuanian, Bohemian, Slovakian, and other extraction, most of whom live in the northern and Midwestern cities. They tend to be Roman Catholic and middle-class. Most workers are either part of the blue-collar labour force or holders of low-level white-collar jobs. The neighbourhoods in which many of them live have their roots in the "Little Italys" and "Polish Hills" established by the immigrants. Their strong ethnic ties are apparent in the pattern of their lives: spouses, friends, neighbours, fellow churchgoers, and even coworkers often are also Polish, Italian, or Slovakian.

Group identity of white ethnics

Their ethnic group identity is not, however, merely a holdover from the era of mass immigration. It is based not only upon a common cultural heritage but also on the common interests, needs, and problems they face in the present-day United States. White ethnics are concentrated in the inner cities, though increasingly less so, and are affected by high crime rates, deteriorating municipal services, inferior schools, and urban unrest. They fear losing their jobs and neighbourhoods to other minority groups. Neither rich nor poor, they anxiously watch their purchas-

ing power being diminished by inflation and rising taxes.

As the children and grandchildren of immigrants, they have been taught to believe that the road to success in the United States is achieved through individual effort. They believe in equality of opportunity and self-improvement, and they attribute poverty to the failing of the individual and not to inequities in society. This attitude makes them largely unsympathetic to the demands of blacks, to student protest (which they tend to view as the antics of spoiled children of the rich), and during the Vietnam War, to the peace movement, which they regarded as un-American, especially since their own sons, many of whom did not attend college, made up a large percentage of the fighting force in Vietnam.

As the ethnics have become more vocal, the public has become aware of the problems and concerns of the urban ethnic minorities and stopped dismissing them as merely "racist" or "uneducated." Ethnic groups have begun to be included in the planning and administration of social-welfare programs of government or foundations, and an ethnic identity is no longer looked upon as somehow un-American and vaguely shameful. It has become legitimate to be an "ethnic."

The blacks. The Civil Rights movement that gained momentum in the early 1960s awakened the nation's conscience to the plight of black Americans, who had long been denied first-class citizenship. By the late 20th century, despite government poverty programs and equal-opportunity laws that outlawed discrimination in education, housing, and employment, blacks remained unequal partners in U.S. society. Their median income and education are below those of whites, and their average rate of unemployment is far greater. The black population nonetheless has made important advances. Their share of higher paying jobs, their median income, and their college enrollment have increased dramatically.

Militancy among the blacks

Before the mid-1960s the Civil Rights movement was largely middle-class and interracial; it used nonviolence and passive resistance to change discriminatory laws and practices, primarily in the South. It aimed for an integration of blacks into American life. For the militants who later arose from the urban poor in the North and West, separatism replaced integration as the primary objective. The militants rejected the American cultural mainstream and spoke instead of black pride, of "soul," of African-American history and culture. Instead of attempting to bring change through the moral persuasion of white governmental bodies and institutions, the militants aimed for black autonomy and community control. They turned their attention to developing black political organizations that would give them a position of bargaining strength and political control over their own communities. Most frightening to whites, the militants rejected nonviolence and proclaimed the need for self-defense to protect themselves from police harassment in the urban ghettoes. The militants' concern with the daily problems of ghetto living spoke directly to the three out of five black households living in inner-city areas.

The older civil rights tradition, however, remained vital, sustained by steadily widening access to professional and business opportunities. In Chicago, Los Angeles, Cleveland, Baltimore, Atlanta, and Washington, D.C., black mayors gained election with white support, demonstrating the openness of the political system; and government programs at all levels promised improvement of the condition of the disadvantaged. Above all, the nonviolent message of the martyred Martin Luther King, Jr., retained its attraction long after his death. In 1988 a follower of King, Jesse Jackson, vigorously campaigned for the presidency on a platform that rejected separatism and addressed the needs of many blacks, but also of other underprivileged voters.

The Hispanics. Persons with Spanish surnames make up more than 7 percent of the U.S. population, but they hardly form a coherent group. The majority are of Mexican origin, some are descendants of ancestors who had lived in areas once part of Mexico—Texas, Arizona, New Mexico, and California—others, legal and illegal migrants from across the loosely guarded border. The greater opportunities and higher living standards of the northern republic continue to attract people who despair of rising above their poverty at home.

The Hispanics of Puerto Rican origin are similar to the Mexican-Americans only in language. These Puerto Ricans are citizens of the United States who leave the overcrowded island for the mainland in the hope of raising their conditions of life. Migration for them is like the continuous process of movement by which Americans have always moved to where chances seem best.

The example of black activism and growing self-confidence has induced Mexican-Americans and Puerto Ricans to draw together to further group influence through politics and social organization. They have lacked the cohesion to form any large nationwide associations, but rather have developed spontaneous local groupings that fight for better health, housing, and municipal services, for bilingual school programs, and for improved education for their children.

Cuban immigrants

Quite different, although also Spanish-speaking, are the Cubans and their children who fled Fidel Castro's revolution of 1959. Although among them are representatives of every social group, these Cubans are distinctive in the large number of professional and middle-class people who migrated. Their social and political attitudes differ significantly from those of Mexican-Americans and Puerto Ricans.

After 1960 easy air travel and political and economic instability stimulated a significant migration from Central and South America. Colour defined the newcomers from Haiti, Jamaica, and Barbados as black, although each group of islanders has retained its own identity. By contrast, the arrivals from Colombia, El Salvador, and Nicaragua, many of them political refugees, have tended to identify with Spanish-speaking groups.

Asian-Americans. Asian-Americans as a group have confounded earlier expectations that they would form an indigestible mass in the society. The Chinese, earliest to arrive, and the Japanese were long victims of racial discrimination. In 1924 the law barred further entries; those already in the United States had been ineligible for citizenship since the previous year. In 1942 thousands of Japanese, many born in the United States and therefore American citizens, were interned in relocation camps because their loyalty was suspect. In the decades since World War II attitudes have changed; anti-Asian prejudice has dissolved, and these groups along with others like the Vietnamese have adjusted and advanced, despite occasional local outbreaks of hostility.

The Indians. The American Indians form an ethnic group only in a very general sense. In actuality, these Native Americans, as they are sometimes called, encompass numerous tribes that are widely separated in language, inherited cultures, and experiences in adaptation. In the East, centuries of coexistence with the dominant whites has led to some degree of intermarriage and assimilation and to various patterns of stable adjustment. In the West the hasty expansion of agricultural settlement crowded the Indians into reservations, where federal policy has vacillated between efforts at assimilation and the desire to preserve tribal cultural identity, with unhappy consequences.

The Indian reservations are often enclaves of deep poverty and social distress. The median Indian family income is far below the national average; and, since Indian families tend to be large, most live near destitution in overcrowded housing conditions. A large proportion of Indian children do not finish high school, and many of those who do graduate often lag behind the national educational norms by several years. The infant mortality rate among Indians is higher than the national figure, and the average age at death is much younger than for the rest of the population.

Indians in urban areas

The physical and social isolation of the reservation has caused a cultural hiatus that has left Indians unprepared educationally and culturally to take part in the urbanized, technical United States. The poverty and frustration of life on the reservation has prompted many Indians to migrate to large cities, especially Los Angeles and Chicago. In such environments they often possess neither the occupational skills nor the cultural background necessary to sustain themselves, and social workers report a high percentage

of family disintegration, alcoholism, and suicide among them. A few activist Indian groups have begun to organize to call attention to their condition and press for change through political and legal action.

RELIGIOUS GROUPS

The U.S. government has never supported an established church, and the diversity of the population has discouraged any tendency toward uniformity in worship. Well over 1,200 religious bodies thrive within the country. Some are the products of native development—among them the Disciples of Christ (founded in the early 19th century), Church of Jesus Christ of Latter-day Saints (Mormons; 1830), Seventh-day Adventists (officially established 1863), Jehovah's Witnesses (1872), and Christian Scientists (1879).

Others had their origins in the Old World, but even these have taken distinctive American forms. Roman Catholics (about a third of the U.S. population) look to Rome for guidance; but some national parishes recognize local distinctions, and there are variations in practice from diocese to diocese. Some 6,000,000 Jews are affiliated with three national organizations (Orthodox, Conservative, and Reform) as well as with many smaller sects. Most Protestant denominations also have European roots, the largest being the Baptists, Methodists, and Lutherans. Among other groups are Pentecostals, Presbyterians, Episcopalians, various Eastern churches (including Orthodox), Congregationalists, Reformed, Mennonites and Amish, various Brethren, Unitarians, and the Friends (Quakers). More recently, substantial numbers of Americans, native and immigrant, have been drawn to Islām and to Buddhism.

IMMIGRATION

The era of mass immigration came to an abrupt end soon after World War I, when the Immigration Act of 1924 established an annual quota, fixed in 1929 at 150,000. The act also established the national-origins system, which was to characterize national immigration policy until 1968.

Quota system of immigration

Under it, quotas were established for each country based on the number of persons of that national origin who were living in the United States in 1920. The quotas reduced drastically the flow of immigrants from southeastern Europe—who, since they were still relatively new arrivals in the United States, formed only a small percentage of the population in 1920—and discriminated in favour of the countries of northwestern Europe. Under this system Great Britain, Ireland, and Germany were allotted more than 70 percent of the quota, an allotment that rarely was filled.

The quota system was liberalized in December 1965, and in 1968 it was finally abolished in favour of a first-come, first-served policy. An annual ceiling of 170,000 immigrant visas for nations outside the Western Hemisphere was established, with 20,000 the maximum allowed to any one nation. A ceiling of 120,000 was set for persons from the Western Hemisphere.

The new policy radically changed the pattern of immigration. In 1965, the last year during which the old national-origins system was fully in effect, Mexico, Canada, and Great Britain headed the list of countries sending immigrants to the United States. They have since been displaced by a number of other countries.

The more open attitude toward newcomers has tended to dissolve the older suspicion against refugees. Increasing numbers have entered from repressive regimes in eastern Europe, the Caribbean, Central and South America, and Asia. Sympathy for their plight has reduced racial inhibitions against such people as the Vietnamese, and successful adjustment has demonstrated the wisdom of the welcoming policy.

Illegal immigration is the subject of greater concern. A law in 1986 granted amnesty to many of those already in the United States, regularized their status, and sought to prevent further unauthorized immigration. Since the national-origins quota had lapsed in 1968 and preference went to migration for family reunification, the law favoured the most recent immigrants and particularly Asians and diminished opportunities for the Europeans formerly favoured. (Jo.N./T.K.F./O.H.)

The economy

Economic power

The United States is the world's greatest economic power in terms of gross national product (GNP) and is among the greatest powers in terms of GNP per capita. The nation's wealth is partly a reflection of its rich natural resources. With only 5 percent of the world's population, the United States produces nearly one-fifth of the world's output of coal, copper, and crude petroleum. The agricultural sector produces nearly one-half of the world's corn (maize); nearly one-fifth of its beef, pork, mutton, and lamb; and more than one-tenth of its wheat. The United States owes its economic position more to its highly developed industry, however, than to its natural resources or agricultural output.

Despite its relative self-sufficiency, the United States is the most important single factor in world trade by virtue of the sheer size of its economy. Its exports represent more than 10 percent of the world total. The United States impinges on the economy of the rest of the world not only as a trading power but also as a source of investment capital. Direct investment abroad by U.S. firms is a dominant factor in the economies of Canada and many Latin-American countries and is also important in Europe and in Asia.

Government and private enterprise. The U.S. government plays only a small direct part in economic activity, being largely restricted to such agencies as the U.S. Postal Service, the Nuclear Regulatory Commission, and the Tennessee Valley Authority. Enterprises that are often in public hands in other countries, such as airlines and telephone systems, are run privately in the United States.

A principal effort of the government traditionally has been the fostering of competition through enforcement of antitrust laws. These are designed to combat collusion among companies with respect to prices, output levels, or market shares and, where feasible, to prevent mergers that significantly reduce competition. The vigour with which antitrust laws and regulations are to be enforced is a matter of perennial political debate.

The major area of government regulation of economic activity is through fiscal and monetary policy. The government also exerts considerable leverage on certain sectors of the economy as a purchaser of goods, notably in the aircraft and aerospace industries. Proposals for governmental controls of prices and incomes have been a frequent source of much controversy.

Farming is a field in which the government strongly influences private economic activity. It endeavours to support farm incomes through payments to farmers, controls on output, price supports, and the provision of storage and marketing facilities. One disadvantage of the system is that payments are related to farm output, so that the benefit often goes to the larger commercial farms rather than to the so-called family farms that were originally the main object of governmental concern.

Types of taxation

Taxation. Nearly all of the federal government's revenues come from taxes. By far the most important source of tax revenue is the personal income tax. Gross receipts from corporate income taxes yield a far smaller percentage of total federal receipts. Excise duties yield another small percentage of total federal revenue, but this is offset by the fact that individual states levy their own excise and sales taxes. Federal excises rest heavily on alcohol, gasoline, and tobacco. Another major source of revenue is social-insurance taxes and contributions; estate and gift taxes yield a tiny percentage of the total.

Trade unions. The labour force in the United States is not highly organized. About three-fourths of those belonging to unions are affiliated with the American Federation of Labor-Congress of Industrial Organizations (AFL-CIO), the nationwide federation of unions. Most unions in manufacturing bargain on a plant- or companywide scale, although the older unions, such as those of the carpenters and the electricians, bargain by crafts. Although the freedom to strike is qualified with legislative provisions for cooling-off periods and in some cases compulsory arbitration, major unions are able and sometimes willing to embark on long strikes.

RESOURCES

Mineral resources. The United States is the world's second largest petroleum-producing nation. The major producing fields are in Alaska, California, the Gulf of Mexico, Louisiana, and Oklahoma. Important reserves of natural gas are found in most of these same areas.

Recoverable coal deposits are concentrated largely in the Appalachians and in Wyoming. Nearly half of the bituminous coal is mined in West Virginia and Kentucky. Pennsylvania produces the country's only anthracite. Illinois, Indiana, and Ohio also produce coal. Iron ore is mined predominantly in Minnesota and Michigan; there has been a long-term shift away from the mining of hematite toward the mining of magnetite.

The United States also has important reserves of copper, magnesium, lead, and zinc. Copper production is concentrated in the mountainous states of the West—in Arizona, Utah, Montana, Nevada, and New Mexico. Zinc is more scattered, being mined in Tennessee, Missouri, Idaho, and New York. Lead mining is concentrated in Missouri. Other metals mined in the United States are gold, silver, molybdenum, manganese, tungsten, bauxite, uranium, vanadium, and nickel. Important nonmetallic minerals produced are phosphates, potash, sulfur, stone, and clays.

Biological resources. Of the total land area, somewhat less than half is devoted to farming (including pasture and range). Tobacco is produced in the Southeast and in Kentucky and cotton in the South and Southwest; California is noted for its vineyards, citrus groves, and truck gardens; the Midwest is the centre of corn and wheat farming, while dairy herds are concentrated in the Northern states. The Southwestern and Rocky Mountain states support large herds of livestock.

About two-thirds of the vast forested areas in the United States are in commercial forestland. The area with the most forestland is the West, including Alaska, but there are large areas also in the South and the North. Almost half of the hardwood is located in the North. Of total commercial forestland, more than 70 percent is in private ownership. About 20 percent is owned or controlled by the federal government, the remainder being under the control of state and local administrations.

Hydroelectric resources. Hydroelectric resources are heavily concentrated in the Pacific and Mountain regions, which together account for almost three-fifths of the nation's hydroelectric installed capacity. Alaska is estimated to have nearly one-fifth of the nation's unrealized potential capacity.

SOURCES OF NATIONAL INCOME

Agriculture, forestry, and fishing. Despite the enormous output of U.S. agriculture, the sector of agriculture, forestry, and fishing altogether produces less than 3 percent of the GNP. Farm productivity has grown at a rapid rate, enabling a smaller labour force to produce more than ever before. Farm manpower has fallen, while mechanization and concentration of farm holdings has increased. Among the most important crops are corn (maize), wheat, barley, grain sorghums, cotton, rice, soybeans, and tobacco. The general improvement in yields over the years has been accompanied by a large increase in the use of commercial fertilizers, pesticides, and herbicides.

Changing structure of agriculture

The United States is the world's major producer of timber. More than 80 percent of the production is made up of softwoods and the remainder of hardwoods; the principal softwood trees are Douglas fir and southern yellow pine; the major hardwood, oak.

The United States is one of the most important countries in the world in fishing. Fish for human consumption accounts for more than half of the tonnage landed. Shellfish account for less than a fifth of the annual catch but for nearly half of the total value.

Mining and quarrying. Less than 2 percent of the GNP comes from mining and quarrying, despite the fact that the United States is a major world producer of a number of metals and of coal and petroleum.

Industry. Manufacturing output has grown at approximately the same rate as the economy as a whole. The increase in productivity over the years has been in part a function of the increased level of capital investment. Manufacturing accounts for about a fifth of the GNP.

One of the most important sectors in terms of value added is the manufacture of transportation equipment, including motor vehicles, aircraft, and space equipment. Other important sectors include nonelectrical machinery, electrical machinery, food products, and chemicals. Steel mill products go largely to the automotive industry and to the construction industry.

In the second half of the 20th century, services have grown faster than any other sector of the U.S. economy. Services are now second only to manufacturing in contribution to the GNP. The most important components are health and business services.

Finance. Under the Federal Reserve System, central banking functions are exercised by 12 Federal Reserve banks, each serving one area of the country, supervised by the Board of Governors in Washington, D.C. The governors are appointed by the president, subject to confirmation by the Senate, but are by no means invariably in accord with the administration's views on economic policy. The Federal Reserve System regulates bank credit and the money supply by changes in the discount rate charged on loans to member banks, by changes in the reserve requirements imposed on commercial banks, and by open-market operations in government securities. The Treasury is not, however, without influence on the working of the monetary system; it influences market interest rates through its management of the national debt, while, by changing its own deposits with the Federal Reserve banks, it can affect the volume of credit.

The capital market

Laws hindering the formation of branch banks have led to a proliferation of individual domestic commercial banks. A liberalizing trend in state banking laws in the 1970s and '80s has encouraged both intra- and interstate expansion of existing bank facilities and bank holding companies. About two-fifths of all commercial banks belong to the Federal Reserve System, but these banks hold almost three-fourths of all commercial bank deposits. Banks incorporated under national charter must be members of the system, while banks incorporated under state charters may become members. Member banks must maintain minimum legal reserves with a Federal Reserve bank and must deposit a percentage (which can vary within established limits) of their time (savings) deposits and checking accounts with a Federal Reserve bank.

In addition to commercial banks, there are thousands of nonbank credit agencies (mostly personal credit institutions and savings and loan associations). Other financial intermediaries include insurance companies and security and commodity brokers. The federal government sponsors other credit agencies in the fields of housing and farming.

New York City has two organized stock (share) exchanges, the New York Stock Exchange and the American Stock Exchange, which account for the bulk of all stock sales in the United States. The country's major grain market is the Chicago Board of Trade, and the leading market for currencies and financial instruments is the Chicago Mercantile Exchange. There are other, smaller exchanges in a number of American cities.

Foreign trade. International trade plays an important part in the U.S. economy, as in other major industrial countries. Indeed, in both exports and imports the United States is the most important single trading country. Since 1976 the United States has had a negative trade balance, and in 1985, for the first time since 1914, U.S. debts owed to foreign creditors exceeded foreign debts owed to U.S. creditors. (E.I.U./Ed.)

TRANSPORTATION

The economic and social complexion of life in the United States mirrors the nation's extraordinary mobility. A pervasive transportation network has helped bring together in the vast geographic expanse of the country a surprisingly homogeneous and close-knit social and economic environment. This freedom to move explains in large measure the dynamism of the U.S. economy. Mobility has made possible vast metropolises, spreading suburbs, a

The mobility of the population

lengthening radius of commuter travel, dispersal of business and industry, and the growing millions of nonfarm rural residents who constitute a new kind of urbanization without a strong centre. Mobility has also had destructive effects. It has accelerated the decay of older urban areas, multiplied traffic congestion, intensified pollution of the environment, and helped to undermine public transportation systems.

The most spectacular part of the U.S. transportation network is perhaps the 44,000-mile Interstate Highway System, a national network of multiple-lane expressways found in all 50 states and connecting 90 percent of all cities of at least 50,000 population. This system, begun in the 1950s, now carries one-fifth of the nation's motor traffic and has made it possible to drive from coast to coast without stopping for a traffic light. The network of roads is densest on the Eastern Seaboard, along the Mississippi and Ohio valleys, and on the West Coast.

Nearly 90 percent of all households own at least one automobile or truck, and many own two or more. While most trips in metropolitan areas are made by automobile, public transit and rail commuter lines play an important role in the most populous cities. The majority of home-to-work travel in the rush hours is by public carrier in such large centres as New York City, Chicago, Philadelphia, and Boston. At the same time, most city transit systems have undergone a sharp decline: bus and subway rides have generally decreased despite a large increase in the urban population of some areas.

Railroads as bulk carriers

Although railroads once dominated both freight and passenger traffic in the United States, government regulation and increased competition from trucking substantially reduced their role in transportation. Railroads now move about one-third of the nation's intercity freight traffic, the most important items carried being coal, grain, chemicals, and motor vehicles. Many rail companies had given up passenger service by 1970, in which year Congress created the National Railroad Passenger Corporation (known as Amtrak) to take over passenger service. Amtrak operates a 24,000-mile system serving nearly 500 stations across the country.

Navigable waterways are extensive and centre upon the Mississippi River system in the country's interior, the Great Lakes–St. Lawrence Seaway system in the north, and the Gulf Coast waterways along the Gulf of Mexico. Barges carry more than two-thirds of domestic waterborne traffic, the major products moved being petroleum products, coal and coke, and grain. The country's largest ports in tonnage handled are New York City; New Orleans, La.; Valdez, Alaska; and Houston, Texas.

Airplane traffic has experienced spectacular growth in the United States since the mid-20th century. From 1970 to 1985, for example, passenger traffic on certified air carriers increased 126 percent. There are nearly 500 public airports, the busiest being Chicago and Atlanta, Ga.

(W.O./Ed.)

Administration and social conditions

GOVERNMENT

The national government. The U.S. Constitution defines a federal system of government in which certain powers are delegated to the national government; other powers fall to the states. The national government consists of executive, legislative, and judicial branches that are designed to check and balance one another; all are interrelated and overlapping yet each is quite distinct.

Since the Constitution was ratified in 1788, there have been 26 amendments to it. The first 10, known as the Bill of Rights, established a number of individual liberties. Notable among the other amendments are the 13th, 14th, and 15th, which abolished slavery and declared former slaves citizens with the right to vote; the 17th, which provided for the direct election of U.S. senators; and the 19th, which effected women's suffrage. Amending the Constitution requires a proposal by a two-thirds vote of both houses of Congress or by a national convention, followed by ratification by three-fourths of the state legislatures or state conventions.

The executive branch. The executive branch of the government is headed by the president, who must be a natural-born citizen of the United States, at least 35 years old, and a resident of the country for at least 14 years. The formal responsibilities of the president include those of chief executive, treaty maker, commander in chief of the army, and head of state. In practice, they have grown to include drafting legislation, formulating foreign policy, personal diplomacy, and leadership of his political party. The members of the president's Cabinet—the secretaries of State, Treasury, Defense, Interior, Agriculture, Commerce, Labor, Health and Human Services, Housing and Urban Development, Transportation, Education, Energy, and Veterans Affairs and the attorney general—are appointed by the president with the approval of the Senate; they are described in the Twenty-fifth Amendment as "the principal officers of the executive departments," but much power has come to be exercised by presidential aides who are not in the Cabinet. Thus, the president's Executive Office includes the Office of Management and Budget, the Council of Economic Advisers, and the National Security Council.

Role of the president

The legislative branch. The legislative branch of the government is the Congress, which has two houses: the Senate and the House of Representatives. Powers granted Congress under the Constitution include the power to levy taxes, borrow money, regulate interstate commerce, declare war, seat members, discipline its own membership, and determine its rules of procedure.

Powers of Congress

With the exception of revenue bills, which must originate in the House of Representatives, legislative bills may be introduced in and amended by either house; a bill—with its amendments—must pass both houses and be signed by the president before it becomes law. The president may veto a bill, but a veto can be overridden by a two-thirds vote of both houses.

The House of Representatives is chosen by the direct vote of the electorate in single-member districts in each state, the number of representatives allotted to each state being based on population and the overall total never exceeding 435. Members must be 25 years old, residents of the states from which they are elected, and previously citizens of the United States for at least seven years. It has become practically imperative, though not constitutionally required, that they be inhabitants of the districts that elect them. They serve for a two-year period. The speaker of the House, who is chosen by the majority party, presides over debate, appoints members of select and conference committees, and performs other important duties. The parliamentary leaders of the two parties are the majority floor leader and the minority floor leader; they are helped by party whips who maintain contact between the leadership and the members of the House. Bills introduced by members in the House of Representatives are received by the standing committees, which meet in private executive session and can amend, expedite, delay, or kill the bills. The committee chairmen traditionally have attained their positions on the basis of seniority, but this practice has been challenged. Among the most important committees are those on Appropriations, Ways and Means, and Rules. The Rules Committee, traditionally conservative, has great power to determine which bills will be brought to the floor of the House for consideration.

Each state elects two senators at large. Senators must be at least 30 years old, residents of the state from which they are elected, and previously citizens of the United States for at least nine years. Each term of service is for six years, and terms are so arranged that one-third of the members are elected every two years.

The Senate has 16 standing committees, among which the most prominent are those on Foreign Relations, Finance, Appropriations, and Governmental Affairs. Debate is almost unlimited and may be used to delay the vote on a bill indefinitely. Such a delay is known as a filibuster and in most instances can be brought to an end if three-fifths of the Senate agree. Treaties made by the president with other governments must be ratified by a two-thirds vote of the Senate.

The judicial branch. The judicial branch of the federal

The Supreme Court

government is headed by the U.S. Supreme Court, which interprets the meaning of the Constitution and of federal laws. It consists of nine justices (including the chief justice) appointed for life by the president with the consent of the Senate. It has appellate jurisdiction for the lower federal courts and from state courts of last resort if a federal question is involved. The court has original jurisdiction over cases involving foreign ambassadors, ministers, consuls, and cases to which a state is a party.

Three types of cases commonly reach the Supreme Court: cases involving litigants of different states, cases involving the interpretation of federal law, and cases involving the interpretation of the Constitution. The court can take official action with as few as six judges joining in deliberation, and a majority vote of the entire court is decisive; a tie vote sustains a lower-court decision. Often the minority judges write a dissenting report.

The Supreme Court has often been criticized for its decisions. In the 1930s, for example, a conservative court overturned much of President Franklin D. Roosevelt's New Deal legislation. In the area of civil rights it has received criticism from various groups at different times. After a 1954 ruling against school segregation, Southern political leaders attacked it harshly. Later, they were joined by Northern conservatives. A number of decisions involving the pretrial rights of prisoners also came under attack on the ground that the court had made it difficult to convict criminals.

Below the Supreme Court are the U.S. courts of appeals. Special courts handle property and contract damage suits against the United States (U.S. Claims Court), review customs rulings (U.S. Court of International Trade), and apply the Uniform Code of Military Justice (U.S. Court of Military Appeals). Each state has at least one federal district court and at least one federal judge. District judges are appointed for life by the president with Senate consent. Appeals from district-court decisions are carried to the courts of appeals.

Revenue. The great majority of the federal government's revenues come from taxes. The most important source is the personal income tax, administered by the Internal Revenue Service. The receipts from corporate income taxes yield a much smaller percentage of total federal receipts. Another small percentage comes from federal excise taxes, but this is offset by the fact that most of the individual states levy their own excise and sales taxes. Federal excise revenues come primarily from taxes on alcohol, gasoline, and tobacco. Social-insurance taxes and contributions constitute another important source of revenue; estate and gift taxes account for only a tiny fraction of the total.

State and municipal governments. The governments of the 50 states have structures closely paralleling those of the federal government. Each state has a governor, a legislature, and a judiciary. Each state has its own constitution.

All state legislatures but one have two houses, Nebraska's being unicameral. Traditionally, state legislatures have been dominated by rural representatives who may not always be sympathetic to the needs of growing urban areas. Most state judicial systems are based upon elected justices of the peace (although in many states this term is not used), above whom come major trial courts, often called district courts, and appellate courts. In addition, there are probate courts concerned with wills, estates, and guardianships. Most state judges are elected.

State governments have a wide array of functions, encompassing agriculture and conservation, highway and motor-vehicle supervision, public safety and corrections, professional licensing, regulation of intrastate business and industry, and certain aspects of education, public health, and welfare. These activities require a large administrative organization, headed by the governor. In most states there is also a lieutenant governor, not always of the same party as the governor, who serves as the presiding officer of the Senate. Other elected officials commonly include a secretary of state, state treasurer, state auditor, attorney general, and superintendent of public instruction.

Municipal governments are more diverse in structure than state governments. There are three basic types: mayor–council governments, commission governments, and council–manager governments. In the first type, the mayor and the council are elected; although the council is nominally responsible for formulating city ordinances, which the mayor enforces, the mayor often controls the actions of the council. Boston, New York City, Philadelphia, Chicago, and Seattle, Wash., are among those cities having the mayor–council type of government. In the commission type, voters elect a number of commissioners, each of whom serves as head of a city department; the presiding commissioner is generally the mayor. Tulsa, Okla., and Salt Lake City, Utah, are included among the cities with commission governments. In the council–manager type, an elected council hires a city manager to administer the city departments. The mayor, elected by the council, simply chairs it and officiates at important functions. Des Moines, Iowa, and Cincinnati, Ohio, have council–manager governments.

The three types of municipal government

Political parties. There are two major political parties, the Democratic Party and the Republican Party. Other parties have occasionally challenged these two but without permanent success. One reason for their failure is that in order to win a national election, a party must appeal to a broad base of voters and a wide spectrum of interests. The two major parties thus tend to be moderate in their programs, and there may often be little difference between them on some issues. Each has a conservative wing, and each has a wing that is considered liberal. The conservative Democrats tend to be more conservative on racial issues, for example, than their Republican counterparts; the liberal Democrats are more radical on economic issues than the liberal Republicans. The national parties contest presidential elections every four years, but, between their quadrennial national conventions, they are often little more than loose alliances of state and local party organizations.

In elections for president and vice president, voters actually choose among electors committed to the support of a particular candidate, a system called the electoral college. Each state is allotted one electoral vote for each senator and representative in Congress.

State and local politics

At the state level, political parties reflect the diversity of the population. Large urban centres are more likely to support a Democratic ticket, whereas rural areas, small cities, and suburban areas tend more often to vote Republican. In many states rural areas and smaller towns control the state legislatures, even though the more populous city areas provide the greater proportion of tax revenue. A Supreme Court ruling in 1964 sought to remedy this situation by ordering states to reapportion their legislatures more closely by population. Some states have traditionally given majorities to one particular party. Until the mid-20th century, for example, the 11 Southern states of the old Confederacy voted almost solidly for Democratic candidates; in other states, such as Maine and South Dakota, Republicans are more likely to win.

Municipal political parties have a pyramidal structure based, at the lowest level, on districts or precincts. The leaders of these units are responsible to ward leaders, who form the governing body of the municipal party. All of these party functionaries are responsible for getting their voters to the polls at election time, often on the basis of a return for services rendered. It is to them, after all, that voters have gone with requests for better municipal services, jobs, and assistance in minor difficulties. One route to political office for the ordinary citizen has traditionally been through the organization: belonging to a neighbourhood party club, helping to raise funds, getting out the vote, watching the polls, and gradually rising through the system to committeeman, city councilman, representative to the state legislature, or—depending on chance, talent, political expediency, and a host of other factors—to higher positions.

As society has become increasingly urban, politics and government have become more complex. Many problems of the cities, including the problems of transportation, housing, education, health, and welfare, can no longer be handled entirely on the local level. Since even the states do not have the necessary resources, cities have often turned to the federal government for assistance.

Armed forces and security. The military forces consist of the U.S. Army, Navy (including the Marine Corps), and Air Force, under the umbrella of the Department of Defense with its headquarters in the Pentagon building in Arlington county, Va. (A related force, the Coast Guard, is governed by the Department of Transportation.) Conscription was ended in 1973, and since that time the United States has had volunteer military forces. The armed services also maintain reserve forces that may be called upon in time of war. Each state has a National Guard consisting of reserve groups subject to call at any time by the governor of the state.

Since a large part of the military budget is spent on matériel and research and development, military programs have considerable economic and political impact. The influence of the military also extends to other countries through a variety of multilateral and bilateral treaties for mutual defense and military assistance. The United States has military bases in Africa, Asia, Europe, and Latin America.

The National Security Act of 1947 created a coordinated command for security and intelligence-gathering activities. The act established the National Security Council (NSC) and the Central Intelligence Agency, the latter being under the NSC and responsible for foreign intelligence. The National Security Agency, an agency of the Department of Defense, is responsible for cryptographic and communications intelligence.

SOCIAL SERVICES

Education. The interplay of local, state, and national programs and policies is particularly evident in the field of education. Historically, education has been considered the province of the state and local governments. Of the more than 3,000 colleges and universities, the academies of the armed services are among the few federal institutions. (The federal government also administers, among others, the College of the Virgin Islands.) For years, however, the federal government has been involved in education at all levels, beginning in 1862 with the grant of public lands to the states for the purpose of establishing colleges of agricultural and mechanical arts, called land-grant colleges. Additionally, the federal government supports school lunch programs, administers Indian education, makes research grants to universities, underwrites loans to college students, and finances education for veterans. Whether the government should also give assistance to private and parochial (religious) schools has been widely debated. The Supreme Court has ruled that direct assistance to parochial schools is barred by the First Amendment to the Constitution, which states that "Congress shall make no law respecting an establishment of religion," although this has not been extended to the use of textbooks and so-called supplementary educational centres.

Colleges and universities

Although responsibility for elementary education still rests primarily with local government, it is increasingly affected by state and national policies. The 1964 Civil Rights Act, for example, required federal agencies to discontinue financial aid to school districts that are not racially integrated. For many school systems where blacks live in largely segregated enclaves, children must be transported long distances to achieve nonsegregated schools.

Trends in education have been toward being more responsive to the needs of a complex society: preschool programs; classes in the community; summer and night schools; increased facilities for exceptional children; and educational programs geared to the culturally deprived and the disaffected student. Such programs, however, have been only partially successful.

Health and welfare. For all the country's wealth, poverty remains a reality for many people in the United States—more than one-tenth of the population. About half of the poor are in homes in which the householder is a full- or part-time wage earner. Of the others, many are too old to work or are disabled, while a large percentage of the householders are mothers of young children. The states provide assistance to the poor in varying amounts. The U.S. Department of Agriculture subsidizes the distribution of low-cost food and food stamps to the poor through the state and local governments.

The poor

Increasing public concern with poverty and welfare led to new federal legislation beginning in the 1960s. Work, training, and rehabilitation programs were established in 1964 for welfare recipients. Between 1964 and 1969 the Office of Economic Opportunity began a number of programs, including the Head Start program for preschool children, the Neighborhood Youth Corps, and the Teacher Corps.

Persons who have been employed are eligible for retirement pensions under the federal Social Security program established in the 1930s. Many employers also provide additional retirement benefits, as well as health insurance and life insurance, usually based on contributions by both the employer and the worker. In addition, more than 35,000,000 persons are covered by private pension plans.

The provision of medical and health care is one of the largest industries in the United States. There are, nevertheless, many inadequacies in medical services, particularly in rural and poor areas. About one-sixth of the population, including members of the armed forces and their families, receive medical care paid for or subsidized by the federal government. Many people are not covered by any form of health insurance.

The federal Department of Health and Human Services through its National Institutes of Health supports much of the biomedical research in the United States. Grants are made also to researchers in clinics and medical schools.

Housing. About two-thirds of the housing units in the United States are single-family houses. Most houses are of wood construction, though often covered with shingles or brick veneer.

Housing, like health, has long been considered a private rather than public concern. The growth of urban slums, however, led many municipal governments to enact stricter building codes and sanitary regulations. In 1934 the Federal Housing Administration was established to make loans to institutions that would build low-rent dwellings. Efforts to reduce the slums in large cities by developing low-cost housing in other areas frequently have been resisted by property owners. For many years the restrictive covenant, by which property owners pledged not to sell to certain racial or religious groups, served to bar those groups from many communities. A 1948 Supreme Court decision declared covenants unenforceable; a 1962 executive order by President John F. Kennedy prohibited discrimination in housing built with federal aid; and many states and cities have adopted fair-housing laws and set up fair-housing commissions. These efforts, however, have had relatively little effect in large cities, where minority groups and the poor often live in ghettos.

Quality of housing

Law enforcement. The enforcement of law in the United States has traditionally been centred in the hands of local police officials. The bulk of the work is done by the police and detectives in the cities and by sheriffs and constables in rural areas. Many of the state governments also have law-enforcement agencies, and all of them have highway-patrol systems for traffic-law enforcement.

Crimes that come under federal jurisdiction (for example, those committed in more than one state) are the responsibility of the Federal Bureau of Investigation (FBI), which also provides assistance through fingerprint identification and technical laboratory services to state and local law-enforcement agencies. In addition, certain federal agencies—such as the Drug Enforcement Agency Administration of the Department of Justice and the Bureau of Alcohol, Tobacco and Firearms of the Department of the Treasury—are empowered to enforce specific federal laws. (J.T.H./Ed.)

Cultural life

The culture that endures in the United States, as in any country, is made not by vast impersonal forces but by uniquely talented men and women; and many of the most gifted artists in the United States, as elsewhere, have chosen to make their art far from the shared realities of daily life. The work of some of the greatest American artists and writers has been done in deliberate seclusion and has taken as its material the interior life of the mind and heart that shapes and precedes national experience.

Even if it is true that these habits of retreat are, in turn, themselves in part traditions and are culturally shaped, it is also true that the least illuminating way to approach the poems of Emily Dickinson or the paintings of Winslow Homer, to take only two imposing examples, is as the consequence of large-scale mass sociological phenomena. Still, many, perhaps even most, American artists have chosen to situate themselves in the common life of their country. Their involvement with the problems they share with their neighbours has given their art a common shape and often a common substance.

For most of the 20th century the common quarrel that has absorbed many American artists and thinkers has been one between the values of a mass, democratic popular culture and those of a refined elite culture accessible only to the few—the quarrel between "low" and "high." In part, this was a problem that science left on the doorstep of the arts: beginning at the turn of the century, the growth of the technology of mass communications—motion pictures, the phonograph, radio and, eventually, television—created a potential audience for stories and music and theatre larger than anyone could previously have imagined. At the beginning of the 20th century, new machines began to appear that made it possible for music and drama and pictures to reach more people than had ever before been possible in human history. People in San Francisco could look at the latest pictures or hear the latest music from New York City months, or even moments, after they were made; a great performance now demanded a pilgrimage no longer than the trip to a corner movie theatre.

Effects of technology

Some of these machines were invented in Europe, and their story is as much a story of the Old World as of the New. But in the United States the growth and dissemination of the new means of mass communication had a special excitement, for the new machines came not simply as a new or threatening force but also as the fulfillment of an American dream. Mass culture seemed to promise a democratic culture, a cultural life directed not to an aristocracy but to all men and women. It was not that the new machines produced new ideals, but that they made the old dreams seem suddenly a practical possibility. Much American art at the turn of the century and through the 1920s, from the paintings of Charles Sheeler to the poetry of Hart Crane, hymned the power of the new technology and the dream of a common culture.

By the mid-20th century, however, many people recoiled in dismay at what had happened to the American arts, high and low. The new technology of mass communications for the most part seemed to have achieved not a generous democratization but a bland homogenization of culture. Many people thought that the control of culture had passed wholly into the hands of advertisers, people who used the means of a common culture just to make money. It was not only that most of the new music and drama that had been made for motion pictures and radio, and later for television, seemed shallow; it was also that the high, or serious, culture that had become available through the means of mass reproduction seemed to have been reduced to a string of popularized hits, which concealed the real complexity of art. Culture, made democratic, had become too easy.

As a consequence, many intellectuals and artists at the end of World War II began to try to construct new kinds of elite culture, art that would be deliberately difficult—and to many people it seemed that this new work was merely difficult; much of the new art and dance seemed puzzling and deliberately obscure. Difficult art happened, above all, in New York City. During World War II, the city had seen an influx of avant-garde artists escaping Adolf Hitler's Europe, including the painters Max Ernst and Piet Mondrian and the composer Igor Stravinsky. They imported many of the ideals of the European avant-garde, particularly the belief that art should always be difficult and "ahead of its time." It is a paradox that the avant-garde movement in Europe had begun in the late 19th century in rebellion against what artists thought were the oppressive and stifling standards of high official culture in Europe and that Europeans had often looked to American mass culture for inspiration. In the United States, however, the practice of avant-garde art became a way for artists and intellectuals to isolate themselves from what they thought was the cheapening of standards.

Yet this counterculture had, by the 1960s, become in large American cities an official culture of its own. For many intellectuals and thinkers at that time, this gloomy situation seemed to be permanent. One could choose between an undemanding low culture and an austere but isolated high culture. For much of the 20th century, scholars saw these two worlds—the public world of popular culture and the private world of modern art—as irreconcilable antagonists and thought that American culture was too often defined by the abyss between them.

The two American cultures

But more and more scholars have begun to see in the most enduring inventions of American culture patterns of cyclic renewal between high and low. It has become apparent, as scholars study particular cases instead of abstract ideas, that the contrast has often been overdrawn. Instead of a simple opposition between popular culture and elite culture, it is possible to recognize in the prolix and varied forms of popular culture innovations and inspirations that have enlivened the most original high American culture and to see how the inventions of high culture circulate back into the street in a spiraling, creative flow. In the achievements of American jazz musicians, who took the songs of Tin Pan Alley and the Broadway musical and inflected them with their own improvisational genius; in the works of great choreographers like Paul Taylor and George Balanchine, who found in tap dances and marches and ballroom bebop new kinds of movement that they then incorporated into the language of high dance; in the "shadow boxes" of the artist Joseph Cornell, who took for his material the mundane goods of the five-and-ten and the department store and used them as private symbols in surreal dioramas—in the work of all of these artists and many more can be seen the same kind of inspiring dialogue between the austere discipline of avant-garde art and the enlivening touch of the vernacular.

Perhaps this circular traffic between high and low also helps account for another remarkable feature of American cultural life since World War II. It is the phenomenon that the cultural historian Thomas Bender has called the triumph of "the culture of the eye and ear." Since World War II, American achievements in dance, music, and painting have been among the most remarkable in the world. This seems paradoxical: precisely those parts of American cultural life that might have seemed most threatened by mass means of cultural communication have been the most triumphantly successful. Yet perhaps it is the provocative distance between the "high" European traditions of dance and art and the "low" desire to take part in American life, the distance between the studio and the street, that has challenged and inspired so many choreographers and painters.

LITERATURE

Because the most articulate artists are, by definition, writers, most of the arguments about what culture is and ought to do have been about literature—which can skew the perception of American culture, since the most memorable American art has not always appeared in books, novels, stories, or plays. In part, perhaps, this is because writing was the first art form to undergo a revolution of mass technology; books were being printed in thousands of copies while to hear a symphony or see a painting required a trip for most Americans. The basic dispute between mass experience and individual experience has therefore perhaps been less keenly felt as an everyday fact in writing in the 20th century than it has been in other art forms. Still, writers have seen and recorded this quarrel as a feature of the world around them, and the evolution of American writing since World War II has shown some of the same basic patterns that are found in painting and dance and the theatre.

In the United States after World War II, many writers, in opposition to what they perceived as the bland flattening-out of cultural life, made their subject the things that set Americans apart from one another. Although, for many Americans, ethnic and even religious differences

had become increasingly less important—holiday rather than everyday material—as the century developed, many postwar writers seized on these differences to achieve a detached point of view on American life. Beginning in the 1940s and '50s, three groups in particular seemed to be "outsider-insiders," who could bring a special vision to fiction: Southerners, Jews, and blacks.

Southern, Jewish, and black writers

Each group had a sense of uncertainty, mixed emotions, and stifled aspirations that lent a questioning counterpoint to the general chorus of affirmation in American life. The Southerners—William Faulkner, Eudora Welty, and Flannery O'Connor in particular—thought that a noble tradition of defeat and failure had been part of the fabric of Southern life since the Civil War; at a time when "official" American culture often insisted that the American story was one of endless triumphs and optimism, they told stories of tragic fate. Jewish writers—most prominently the Chicago novelist Saul Bellow, who won the Nobel Prize in 1976, Bernard Malamud, and Philip Roth—found in the "golden exile" of Jews in the United States a juxtaposition of surface affluence with deeper unease and perplexity that seemed to many of their countrymen to offer a common predicament in a heightened form.

For black Americans the promise of American life had in many respects never been fulfilled. "What happens to a dream deferred?" the poet Langston Hughes asked, and many black writers attempted to answer that question through stories that mingled pride, perplexity, and rage. Black literature achieved one of the unquestioned masterpieces of 20th-century American fiction in Ralph Ellison's *Invisible Man* (1952). More recently, the rise of feminism as a political movement has given many women a sense that their experience, too, is rich and important; since the 1970s there has been an explosion of women's fiction, including the much-admired work of Toni Morrison (*Beloved;* 1987), Anne Tyler (*The Accidental Tourist;* 1985), and Louise Erdrich (*Love Medicine;* 1984).

Perhaps precisely because so many novelists sought to make their fiction from experiences that were deliberately imagined as "marginal," set aside from the general condition of American life, many other writers had the sense that fiction, and particularly the novel, might no longer be the best way to record American life. For many writers, the novel seemed to have become above all a form of private, interior expression that could no longer keep up with the extravagant oddities of the United States. Many writers took up journalism with some of the passion for perfection of style that had once been reserved for fiction. The exemplars of this form of poetic journalism included the masters of *The New Yorker* magazine, most notably A.J. Liebling, whose books included *The Earl of Louisiana* (1961), a study of an election in that state, as well as Joseph Mitchell, who in his books *The Bottom of the Harbor* (1959) and *Joe Gould's Secret* (1965) offered dark and perplexing accounts of the American metropolis. The dream of combining facts and lyrical fire also achieved a masterpiece in James Agee's *Let Us Now Praise Famous Men* (1941), an account of sharecropper life in the South that is a landmark in the struggle for imparting to nonfiction the beauty and permanence of poetry.

The nonfiction novel

As the century developed, this genre of imaginative nonfiction (sometimes called the "nonfiction novel" or "documentary novel") took on different forms. Truman Capote's *In Cold Blood* (1965), for example, recreated a multiple murder in Kansas. By contrast, Tom Wolfe, whose influential books included *The Right Stuff* (1979), an account of the early days of the U.S. space program, and Norman Mailer, whose books included *The Armies of the Night* (1968), a ruminative piece about the political conventions in 1968, took on huge public events and made them subject to the insights (and, many people thought, the idiosyncratic whims) of a personal sensibility.

As the nonfiction novel often pursued extremes of grandiosity and hyperbole, the short story assumed a previously unexpected importance in the life of American writing; the short story became the voice of private vision and private lives. The short story, with its natural insistence on the unique moment and the glimpse of something private and fragile, came to have a new prominence. The rise of the American short story is bracketed by two remarkable books: J.D. Salinger's *Nine Stories* of 1953 and Raymond Carver's collection *What We Talk About When We Talk About Love* (1981), almost exactly a quarter century later. Salinger inspired a generation by imagining that a serious search for a spiritual life could be reconciled with an art of gaiety and charm; Carver confirmed in the next generation their sense of a loss of spirituality through an art of taciturn reserve and cloaked emotions.

THE VISUAL ARTS

Perhaps the greatest, and certainly the loudest, event of American cultural life after World War II was what the critic Irving Sandler has called "the triumph of American painting"—the emergence of a new form of art that allowed American painting to become dominant in the world. This dominance lasted for at least 40 years, from the birth of the so-called New York school, or Abstract Expressionism, around 1945 until at least the mid-1980s, and took in many different kinds of art and artists. In its first flowering, in the epic-scaled abstractions of Jackson Pollock, Mark Rothko, Willem de Kooning, and the other members of the New York school, this new painting seemed abstract, rarefied, and constructed from a series of negations, from saying "no" to everything except the purest elements of painting. Abstract Expressionism seemed to stand at the farthest possible remove from the common life of American culture, and particularly from the life of American popular culture. Even this painting, however, in later years came under a new and perhaps less austere scrutiny; and the art historian Robert Rosenblum has argued that many of the elements of Abstract Expressionism, for all of their apparent hermetic distance from common experience, are inspired by the scale and light of the American landscape and of 19th-century landscape painting—by elements that run deep and centrally in Americans' sense of themselves and their country.

Abstract Expressionism

It is certainly true that the next generation of American painters, who throughout the 1950s continued the unparalleled dominance of American influence in the visual arts, made their art aggressively and unmistakably of the dialogue between the studio and the street. The painter Jasper Johns, for instance, took as his subject the most common and even banal of American symbols—maps of the 48 continental states, the flag itself—and depicted the quickly read and immediately identifiable common icons with a slow, meditative painterly scrutiny. His contemporary and occasional partner, Robert Rauschenberg, took up the same dialogue in a different form; his art consisted of dreamlike collages of images silk-screened from the mass media, combined with personal artifacts and symbols, all combined in a mélange of jokes and deliberately perverse associations. In a remarkably similar spirit, the eccentric Surrealist Joseph Cornell took images from popular culture and made them into a language of nostalgia and poetic reverie. Although Cornell, like William Blake, whom he in many ways resembled, worked largely in isolation, his sense of the poetry that lurks unseen in everyday objects had a profound effect on other artists.

By the early 1960s, with the explosion of the new kind of art called Pop art, the engagement of painting and drawing with popular culture seemed so explicit as to be almost overwhelming and at times risked losing any sense of private life and personal inflection at all—risked becoming all street and no studio. Artists such as Andy Warhol, Roy Lichtenstein, and Claes Oldenburg took the styles and objects of popular culture—everything from comic books to lipstick tubes—and treated them with the absorption and grave seriousness previously reserved for religious icons. But this art, too, had its secrets and strong individual voices and visions. In his series of drawings called "*Proposals for Monuments and Buildings, 1965–69*" Oldenburg drew ordinary things—fire hydrants, ice cream bars, bananas—as though they were as large as skyscrapers; his pictures combined a virtuoso's gift for drawing with a vision, at once celebratory and satirical, of the P.T. Barnum spirit of American life. Warhol silk-screened images of popular movie stars and Campbell's soup cans; in replicating them, he suggested that their reiteration by mass

Pop art

production had emptied them of their humanity but also had given them a hieratic immortality. Lichtenstein used the techniques of comic-book illustration to paraphrase some of the monuments of modern painting, making a coolly witty art in which the forms of Henri Matisse were embodied in the costume of Captain Marvel.

But those artists who self-consciously chose to make their art out of popular materials and images were not the only ones who had something to say about the traffic between mass and elite culture. The so-called Minimalists, who made abstract art out of simple, and usually hard-edged, geometric forms, from one point of view carried on the tradition of austere abstraction. But it was also the Minimalists, as art historians have pointed out, who carried over the vocabulary of the new International style of unornamented architecture into the world of the fine arts; Minimalism imagined the dialogue between street and studio in terms of hard edges and simple forms rather than in terms of imagery, but it took part in the same dialogue. In some cases, the play between high and low has been carried out as a dialogue between Pop and Minimal styles themselves. Frank Stella, thought by many to be the preeminent American painter since the 1970s, began as a Minimalist, making extremely simple paintings of black chevrons from which everything was banished except the barest minimum of painterly cues. Yet in his subsequent work he became almost extravagantly "maximalist" and, as he began to make bas-reliefs, added to the stark elegance of his early paintings wild, Pop-art elements of outthrusting spirals and Day-Glo colours—even sequins and glitter—that deliberately suggested the invigorating vulgarity of the Las Vegas "Strip." Stella's flamboyant reliefs combined the spare elegance of abstraction with the greedy vitality of the American street.

Some artists made their art public by borrowing from images and icons of the street; other artists wanted art to take on a new responsibility by making an art for the street. Many artists in the 1970s and '80s, among them Mary Miss, Alice Aycock, James Turrell, Robert Irwin, and Elyn Zimmerman, tried to bridge the gulf between American art and life through the simple means of making sculpture for public spaces. This movement, called "site-sculpture" or "public art," rejected the idea of public sculpture as forbidding monuments set in the middle of arid plazas and instead tried to construct environments through which viewers could pass, so that their experience of the work took place over time. The most powerful and moving monument of these years—Maya Lin's Vietnam Veterans Memorial in Washington, D.C., where every day thousands of visitors pass through a sunken black polished V inscribed with the names of the Vietnam War dead—used highly abstract and even minimal means to orchestrate a powerfully emotional drama and should be seen as an expression of the site-sculpture movement.

American photography

Formerly, most surveys of American culture might not have thought photography of much importance, but photography has begun to lay a new claim to attention as a serious art form. For most of the first part of the 20th century, the most remarkable American photographers had, on the whole, tried to make photography into a fine art by divorcing it from its ubiquitous presence as a recorder of moments and by splicing it onto older, painterly traditions. After World War II, however, a few gifted photographers were able to transcend the distinction between media image and aesthetic object—between art and photojournalism—and to make from a single, pregnant moment a complete and enduring image. Walker Evans and Robert Frank (the latter, like so many artists of the postwar period, an immigrant), rather than trying to make photography as calculated and considered as the traditional fine arts, found in the instantaneous vision of the camera something at once personal and permanent. Frank's book *The Americans* (1959), the record of a tour of the United States that combined an apparent casualness with a sense of the ominous worthy of the Italian painter Giorgio de Chirico, was the masterpiece of this vision, and no work of the postwar era was more influential in all fields of visual expression. Robert Mapplethorpe, Diane Arbus, and above all Richard Avedon and Irving Penn, who together dominated both fashion and portrait photography for almost half a century and who straddled the lines between museum and magazine, high portraiture and low commercials—all of these photographers came to seem, in their oscillations between glamour and gloom and commercial and personal work, exemplary of the predicaments and potential triumphs facing the American artist.

THE THEATRE

Perhaps more than any other art form, the American theatre has suffered from the invention of the new technologies of mass reproduction. Whereas painting and writing can choose their distance from (or intimacy with) mass culture, much of the age-old material of the theatre has been subsumed by motion pictures and television; what the theatre can do that cannot be done elsewhere is not always clear. As a consequence, the Broadway theatre, which in the 1920s seemed a vital area of American culture and, in the high period of the playwright Eugene O'Neill, a place of cultural renaissance, has become nearly defunct. A brief, and largely false, spring took place in the period just after World War II. Tennessee Williams and Arthur Miller, in particular, both wrote movingly and even courageously about the lives of "left-out" Americans, demanding attention for the outcasts of a relentlessly commercial society. Half a century later, however, they both seemed more traditional and less profoundly innovative than their contemporaries in the other arts—more profoundly tied to the conventions of European naturalistic theatre and less inclined or able to renew and rejuvenate the language of their form. The one completely original American contribution to the stage, the musical theatre, blossomed in the works of Frank Loesser (*Guys and Dolls* in particular, which the critic Kenneth Tynan regarded as one of the greatest of American plays), but it exists today largely as a revival art and in the brave "holdout" works of the composer and lyricist Stephen Sondheim (*Company, Sweeney Todd, Into the Woods*). When new kinds of popular music (particularly rock) took the place of the Broadway theatre song as the soundtrack of American romance, the stage lost its revivifying contact with the street and became an increasingly distant form.

MOTION PICTURES

Reevaluation of the American motion picture

In some respects the motion picture is the American art form par excellence, and no area of art has undergone a more dramatic revision in critical appraisal since the 1970s. Throughout most of the 1940s and '50s, serious critics, with a few honourable exceptions (Agee in particular), even those who considered the motion picture seriously as a potential artistic medium, took it for granted that the Hollywood movie was, judged as art, hopelessly compromised by commerce. In the 1950s in France, however, a generation of critics associated with the magazine *Cahiers du Cinéma* (most of whom later would become well-known filmmakers themselves, including François Truffaut and Jean-Luc Godard) argued that the American film, precisely because its need to please a mass audience had helped it break out of the limiting gentility of the European cinema, had a vitality and a set of masters without equal in the world. New studies and appreciation of such Hollywood filmmakers as John Ford, Howard Hawks, and William Wyler resulted, and, eventually, this new evaluation worked its way back into the United States itself: another demonstration that one country's low art can become another country's high. Imported back to the United States, this reevaluation of Hollywood motion pictures changed and amended preconceptions that had hardened into prejudices. The new appreciation of the individual vision of the Hollywood film was to inspire a generation of young American filmmakers, including Francis Ford Coppola, Martin Scorsese, and George Lucas, to use the commercial film as at once a form of personal expression and a means of empire building, with results that were, predictably, mixed.

DANCE

Serious dance hardly existed in the United States until after the mid-20th century. One remarkable American,

Isadora Duncan, had played as large a role as anyone at the turn of the century and after in the transformation of classical dance into a form of intense and improvisational personal expression. But most of Duncan's life and career were spent in Europe, and she bequeathed to the American imagination a shining, influential image rather than a set of steps. Ruth St. Denis and Ted Shawn, throughout the 1920s, kept dance in the United States alive; but it was in the work of the choreographer Martha Graham that the tradition of modern American dance which Duncan had invented found its first and most influential master. Graham's work, like that of her contemporaries among the Abstract Expressionist painters, sought a basic, timeless vocabulary of primal expression, but in the most direct sense she founded a tradition. A Graham dancer, Paul Taylor, became the most influential modern-dance master of the next generation, and a Taylor dancer, Twyla Tharp, in turn became the most influential choreographer of the generation after that. Where Graham had deliberately turned her back on popular culture, however, both Taylor and Tharp viewed it quizzically, admiringly, and hungrily. Whether the low inspiration came from music, as in Tharp's "Nine Sinatra Songs," choreographed to recordings by the pop singer Frank Sinatra and employing and transforming the language of ballroom dance, or directly off the street, as in a famous section of Taylor's dance "Cloven Kingdom," in which the dancer's movements are inspired by the way Americans walk and strut and fight, both Taylor and Tharp continued to feed upon popular culture without being consumed by it. Perhaps for this reason their art has continued to grow in stature around the world; they are intensely local and yet prized elsewhere.

The influence of Martha Graham

George Balanchine, the choreographer who dominated the greatest of American ballet troupes, the New York City Ballet, from its founding in 1948 until his death in 1983, might be considered outside the bounds of purely American culture. Yet this only serves as a reminder of how limited and provisional such national groupings must always be, for though Balanchine was born and educated in Russia and took his inspiration from a language of movement codified in Europe in the 19th century, no one imagined the gestures of American life with more verve, love, or originality. His art was open to everything from the austere and demanding music of Charles Ives to the marches of John Philip Sousa. Balanchine created new standards of beauty for both men and women dancers (and not incidentally, helped spread a new standard of athletic beauty into the culture at large) and invented an audience for dance in the United States where none had existed before. By the end of his life, Balanchine was perhaps the greatest and certainly among the most American of all artists.

AUDIENCES

Art is made by artists but is possible only with audiences, and perhaps the most worrying trait of American culture since the mid-20th century has been the threatened disappearance of a broad middle audience for the arts. Many weekly magazines that had helped sustain a sense of community among educated readers—*Collier's, The Saturday Evening Post, Look, Life*—have stopped publishing since the 1970s (although *Life* is still published as a monthly). Others, including *Harpers' Magazine* and *Atlantic,* continue more as philanthropies.

Television has become king and, like many despots, insists that its tyranny is simply an expression of the popular will. The fierce competition among the national networks for ratings may seem absurd to future generations, since they involve relatively small differences among huge numbers; a show that consistently attracts two million devoted viewers—for example, more people watching a Mozart opera in a single night than might have seen it in all of the 19th century—is characterized, on these despotic grounds, as a failure and as a "minority taste."

Arts in the American university

As the elephantine growth and devouring appetite of television have reduced the middle audience, there has also been a concurrent growth in the support of the arts in the university. The public support of higher education in the United States, although its ostensible purposes are often merely pragmatic, has had the perhaps unintended effect of making the universities into cathedrals of culture. The positive side of this development should never be overlooked; things that have begun as scholarly pursuits—for example, the enthusiasm for authentic performances of early music—have after their incubation in the academy, given pleasure to increasingly larger audiences. The growth of the universities has also, for good or ill, helped decentralize culture; the Guthrie Theater in Minneapolis, Minn., for example, or the regional opera companies of St. Louis, Mo., and Santa Fe, N.M., are difficult to imagine without the support and involvement of local universities. But many people believe that the "academicization" of the arts has also had the negative effect of encouraging art made by college professors for other professors. Some people believe, for example, that this has led to the development of a literature which is valued less for its engagement with the world than for its engagement with other kinds of writing.

Yet a broad, middle-class audience for the arts, if it is endangered, continues to flourish. The establishment of the Lincoln Center for the Performing Arts in New York City in the early 1960s provided a model for subsequent centres across the country, including the John F. Kennedy Center for the Performing Arts in Washington, D.C., which opened in 1971. It is sometimes said that the audiences who attend concerts and recitals at these centres are mere "consumers" of culture, rather than people engaged passionately in the ongoing life of the arts. But it seems probable that the motives that lead Americans to the concert hall or opera house are no more or less mixed today than have been those of any other people in any other period: a desire for prestige, a sense of duty, and real love of the form all commingle.

The deeper problem that has led to one financial crisis after another for theatre companies and dance troupes and museums (the Twyla Tharp dance company, despite its worldwide reputation, for instance, and a popular orientation that included several successful seasons on Broadway, was compelled in the 1980s to be absorbed into American Ballet Theatre) rests on hard and fixed facts about the economics of the arts and about the economics of the performing arts in particular. Ballet, opera, symphony, and drama are labour-intensive "industries" in an era of labour-saving devices. Other industries have remained competitive by substituting automated labour for human labour; but for all that new stage devices can help cut costs, the basic demands of the old art forms are difficult to alter. A corps de ballet cannot be mechanized or stored on software; voices belong to singers, and singers cannot be replicated. Many Americans, accustomed to the simple connection between popularity and financial success, have had a difficult time grasping this fact; perhaps it is one of the reasons for the uniquely impoverished condition of government funding for the arts in the United States. (Canadian government sources, for example, spend, on a per capita basis, more than 10 times as much money on the arts as does the U.S. government.)

Government funding

Yet the anxiety about the future of the arts in the United States reflects, in part, the extraordinary demands Americans have come to make on them. No country has ever before, for good or ill, invested so much in the ideal of a common culture; the arts for most Americans are imagined as therapy, as education, as a common inheritance, as, in some sense, the definition of life itself and the sum of all good things. Americans have increasingly asked art to play the role that religion played in older cultures. Perhaps this is because it is, in the United States, only in art that all the disparate parts of the culture, can, at least for the length of a story or play or ballet, at last come together. Culture in other countries has always expressed ideals of hierarchy and fixed place: they suggest a ladder ascending from low folk culture to high aristocratic refinement. At its best, American culture has suggested a ring, not a ladder. High and low join hands. (A.Go.)

For statistical data on the land and people of the United States of America, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

The territory represented by the continental United States had, of course, been discovered, perhaps several times, before the voyages of Columbus. When Columbus came, he found the New World inhabited by peoples who in all likelihood had originally come from the continent of Asia. Probably these first inhabitants had arrived 20,000 to 35,000 years before in a series of migrations from Asia to North America by way of the Bering Strait. By the time the first Europeans appeared, the aborigines (commonly referred to as Indians) had spread and occupied all portions of the New World.

The foods and other resources available in each physiographic region largely determined the type of culture prevailing there. Fish and sea mammals, for example, contributed the bulk of the food supply of coastal tribes, although the acorn was a staple for California Indians; plant life and wild game (especially the American bison, or buffalo) were sources for the Plains Indians; small-game hunting and fishing (depending again on local resources) provided for Midwestern and Eastern tribes. These foods were supplemented by corn, which was a staple food for the Indians of the Southwest. The procurement of these foods called for the employment of fishing, hunting, plant and berry gathering, and farming techniques, the application of which depended, in turn, upon the food resources utilized in given areas.

Foods and other raw materials likewise conditioned the material culture of the respective regional groups. All Indians transported goods by human carrier; the use of dogs to pull sleds or travois was widespread; and rafts, boats, and canoes were used where water facilities were available. The horse, imported by the Spanish in the early 16th century, was quickly adopted by the Indians once it had made its appearance. The horse came to be used widely by the buffalo-hunting Indians of the Great Plains.

Indian culture groups are distinguished among other ways by house types. The dome-shaped ice houses were developed by the Eskimos; rectangular plank houses were produced by the Northwestern Indians; earth and skin lodges and tepees by plains and prairie tribes; flat-roofed and often multistoried houses by some of the Pueblo Indians of the Southwest; barrel houses by the natives in the Northeast. Clothing, or the lack of it, likewise varied with native groups, as did crafts, weapons, and tribal economic, social, and religious customs.

At the time of Columbus' arrival there were probably roughly 1,500,000 Indians in what is now the continental United States, although estimates vary greatly. In order to assess the role and the impact of the American Indian upon the subsequent history of the United States in any meaningful way, one must understand the differentiating factors, such as those mentioned above. Generally speaking it may be said, however, that the American Indians as a whole exercised an important influence upon the white civilization transplanted from Europe to the New World. Indian foods and herbs, articles of manufacture, methods of raising some crops, war techniques, words, a rich folklore, and racial infusions are among the more obvious general contributions of the Indians to their European conquerors. The protracted and brutal westward-moving conflict caused by white expansionism and Indian resistance constitutes one of the most tragic chapters in the history of the United States. (O.O.W./Ed.)

Colonial America to 1763

THE EUROPEAN BACKGROUND

The English colonization of North America was but one chapter in the larger story of European expansion throughout the globe. The Portuguese, beginning with a voyage to Porto Santo off the coast of West Africa in 1418, were the first Europeans to promote overseas exploration and colonization. By 1487 the Portuguese had traveled all the way to the southern tip of Africa, establishing trading stations at Arguin, Sierra Leone, and El Mina. In 1497 Vasco da Gama rounded the Cape of Good Hope and sailed up the eastern coast of Africa, laying the groundwork for Portugal's later commercial control of India. By 1500, when Pedro Álvares Cabral stumbled across the coast of Brazil en route to India, Portuguese influence had expanded to the New World as well.

Though initially lagging behind the Portuguese in the arts of navigation and exploration, the Spanish quickly closed that gap in the decades following Columbus' voyages to America (see COLUMBUS). First in the Caribbean and then in spectacular conquests of New Spain and Peru, they captured the imagination, and the envy, of the European world.

French settlement

France, occupied with wars in Europe to preserve its own territorial integrity, was not able to devote as much time or effort to overseas expansion as did Spain and Portugal. Beginning in the early 16th century, however, French fishermen established an outpost in Newfoundland, and in 1534 Jacques Cartier began exploring the Gulf of St. Lawrence. By 1543 the French had ceased their efforts to colonize the northeast portion of the New World. In the last half of the 16th century, France attempted to found colonies in Florida and Brazil; but each of these efforts failed, and by the end of the century Spain and Portugal remained the only two European nations to have established successful colonies in America.

The English, although anxious to duplicate the Spanish and Portuguese successes, nevertheless lagged far behind in their colonization efforts. The English possessed a theoretical claim to the North American mainland by dint of the 1497 voyage of John Cabot off the coast of Nova Scotia, but in fact they had neither the means nor the desire to back up that claim during the 16th century. Thus it was that England relied instead on private trading companies, which were interested principally in commercial rather than territorial expansion, to defend its interests in the expanding European world. The first of these commercial ventures began with the formation of the Muscovy Company in 1554. In 1576–78 the English mariner Martin Frobisher undertook three voyages in search of a Northwest Passage to the Far East. In 1577 Sir Francis Drake made his famous voyage around the world, plundering the western coast of South America en route. A year later Sir Humphrey Gilbert, one of the most dedicated of Elizabethan imperialists, began a series of ventures aimed at establishing permanent colonies in North America. All of his efforts met with what was, at best, limited success. Finally, in September 1583, Gilbert, with five vessels and 260 men, disappeared in the North Atlantic. With the failure of Gilbert's voyage, the English turned to a new man, Sir Walter Raleigh, and a new strategy—a southern rather than a northern route to North America—to advance England's fortunes in the New World. Raleigh's efforts to found a permanent colony off the coast of Virginia, although they did finally fail with the mysterious destruction of the Roanoke Island colony in 1587, awakened popular interest in a permanent colonizing venture.

Motives for English colonization

During the years separating the failure of the Roanoke colony and the establishment in 1607 of the English settlement in Jamestown, English propagandists worked hard to convince the public that a colony in America would yield instant and easily exploitable wealth. Even men like the English geographer Richard Hakluyt were not certain that the Spanish colonization experience could or should be imitated but hoped nevertheless that the English colonies in the New World would prove to be a source of immediate commercial gain. There were, of course, other motives for colonization. Some hoped to discover the much-sought-after route to the Orient in North America. English imperialists thought it necessary to settle in the New World in order to limit Spanish expansion. Once it was proven that America was a suitable place for settlement, some Englishmen would travel to those particular colonies that promised to free them from religious persecution. There were also Englishmen, primarily of lower- and middle-

class origin, who hoped the New World would provide them with increased economic opportunity in the form of free or inexpensive land. These last two motives, while they have been given considerable attention by historians, appear not to have been so much original motives for English colonization as they were shifts of attitude once colonization had begun.

SETTLEMENT

Virginia. The leaders of the Virginia Company of London, a joint-stock company in charge of the Jamestown enterprise, were for the most part wealthy and wellborn commercial and military adventurers eager to find new outlets for investment. During the first two years of its existence, the Virginia colony, under the charter of 1607, proved an extraordinarily bad investment. This was principally due to the unwillingness of the early colonizers to do the necessary work of providing for themselves and to the chronic shortage of capital for supply of the venture.

A new charter in 1609 significantly broadened membership in the Virginia Company, thereby increasing temporarily the supply of capital at the disposal of its directors; but most of the settlers continued to act as though they expected the Indians to provide for their existence, a notion that the Indians fiercely rejected. As a result, the enterprise still failed to yield any profits, and the number of investors again declined.

The crown issued a third charter in 1612, authorizing the company to institute a lottery to raise more capital for the floundering enterprise. In that same year John Rolfe harvested the first crop of a high-grade and therefore potentially profitable strain of tobacco. At about the same time, with the arrival of Sir Thomas Dale in the colony as governor in 1611, the settlers gradually began to practice the discipline necessary for their survival, though at an enormous personal cost.

The administration of Sir Thomas Dale

Dale carried with him the "Laws Divine, Morall and Martial," which were intended to supervise nearly every aspect of the settlers' lives. Each person in Virginia, including women and children, was given a military rank, with duties spelled out in minute detail. Penalties imposed for violating these rules were severe: those who failed to obey the work regulations were to be forced to lie with neck and heels together all night for the first offense, whipped for the second, and sent to a year's service in English galleys (convict ships) for the third. The settlers could hardly protest the harshness of the code, for that might be deemed slander against the company—an offense punishable by service in the galleys or by death.

Dale's code brought order to the Virginia experiment, but it hardly served to attract new settlers. To increase incentive the company, beginning in 1618, offered 50 acres of land to those settlers who could pay their transportation to Virginia and a promise of 50 acres after seven years of service to those who could not pay their passage. Concurrently, the new governor of Virginia, Sir George Yeardley, issued a call for the election of representatives to a House of Burgesses, which was to convene in Jamestown in July 1619. In its original form, the House of Burgesses was little more than an agency of the governing board of the Virginia Company, but it would later expand its powers and prerogatives and become an important force for colonial self-government.

Virginia made a crown colony

Despite the introduction of these reforms, the years from 1619 to 1624 proved fatal to the future of the Virginia Company. Epidemics, constant warfare with the Indians, and internal disputes took a heavy toll on the colony. In 1624 the crown finally revoked the charter of the company and placed the colony under royal control. The introduction of royal government into Virginia, while it was to have important long-range consequences, did not produce an immediate change in the character of the colony. The economic and political life of the colony continued as it had in the past. The House of Burgesses, though its future under the royal commission of 1624 was uncertain, continued to meet on an informal basis; by 1629 it was officially reestablished. The crown also grudgingly acquiesced to the decision of the Virginia settlers to continue to direct most of their energies to the growth and exportation of tobacco. By 1630 the Virginia colony, while not prosperous, at least showed signs that it was capable of surviving without royal subsidy.

Maryland. Maryland, Virginia's neighbour to the north, was the first English colony to be controlled by a single proprietor rather than by a joint-stock company. George Calvert (Lord Baltimore) had been an investor in a number of colonizing schemes before being given a grant of land from the crown in 1632. Baltimore was given a sizable grant of power to go along with his grant of land; he had control over the trade and political system of the colony so long as he did nothing to deviate from the laws of England. Baltimore's son Cecilius Calvert took over the project at his father's death and promoted a settlement at St. Mary's on the Potomac. Supplied in part by Virginia, the Maryland colonists managed to sustain their settlement in modest fashion from the beginning. As in Virginia, however, the early 17th-century settlement in Maryland was often unstable and unrefined; composed overwhelmingly of young, single males—many of them indentured servants—it lacked the stabilizing force of a strong family structure to temper the rigours of life in the wilderness.

The colony was intended to serve at least two purposes. Baltimore, a Roman Catholic, was anxious to found a colony where Catholics could live in peace, but he was also eager to see his colony yield him as large a profit as possible. From the outset Protestants outnumbered Catholics, although a few prominent Catholics tended to own an inordinate share of the land in the colony. Despite this favouritism in the area of land policy, Baltimore was for the most part a good and fair administrator.

Following the accession of William and Mary to the English throne, however, control of the colony was taken away from the Calvert family and entrusted to the royal government. Shortly thereafter the crown decreed that Anglicanism would be the established religion of the colony. In 1715, after the Calvert family had renounced Catholicism and embraced Anglicanism, the colony reverted back to a proprietary form of government.

The New England colonies. Although lacking a charter, the founders of Plymouth in Massachusetts were, like their counterparts in Virginia, dependent upon private investments from profit-minded backers to finance their colony. The nucleus of that settlement was drawn from an enclave of English émigrés in Leyden, Holland. These religious Separatists believed that the true church was a voluntary company of the faithful under the "guidance" of a pastor and tended to be exceedingly individualistic in matters of church doctrine. Unlike the settlers of Massachusetts Bay, the Pilgrims chose to "separate" from the Church of England rather than to reform it from within.

In 1620, the first year of settlement, nearly half the settlers died of disease. From that time forward, however, and despite decreasing support from English investors, the health and the economic position of the colonists improved. The Pilgrims soon secured peace treaties with most of the Indians around them, enabling them to devote their time to building a strong, stable economic base rather than diverting their efforts toward costly and time-consuming problems of defending the colony from attack. Although none of their principal economic pursuits—farming, fishing, and trading—promised them lavish wealth, the Pilgrims in America were, after only five years, self-sufficient.

Although the Pilgrims were always a minority in Plymouth, they nevertheless controlled the entire governmental structure of their colony during the first four decades of settlement. Before disembarking from the *Mayflower* in 1620, the Pilgrim founders, led by William Bradford, demanded that all the adult males aboard who were able to do so sign a compact promising obedience to the laws and ordinances drafted by the leaders of the enterprise. Although the Mayflower Compact has been interpreted as an important step in the evolution of democratic government in America, it is a fact that the compact represented a one-sided arrangement, with the settlers promising obedience and the Pilgrim founders promising very little. Although nearly all the male inhabitants were permitted to vote for deputies to a provincial assembly and for a governor,

Mayflower Compact

the colony, for at least the first 40 years of its existence, remained in the tight control of a few men. After 1660 the people of Plymouth gradually gained a greater voice in both their church and civic affairs, and by 1691, when Plymouth colony was annexed to Massachusetts Bay, the Plymouth settlers had distinguished themselves by their quiet, orderly ways.

The Puritans of Massachusetts Bay, like the Pilgrims, sailed to America principally to free themselves from religious restraints. Unlike the Pilgrims, the Puritans did not desire to "separate" themselves from the Church of England but, rather, hoped by their example to reform it. Nonetheless, one of the recurring problems facing the leaders of the Massachusetts Bay colony was to be the tendency of some, in their desire to free themselves from the alleged corruption of the Church of England, to espouse Separatist doctrine. When these tendencies or any other hinting of deviation from orthodox Puritan doctrine developed, those holding them were either quickly corrected or expelled from the colony. The leaders of the Massachusetts Bay enterprise never intended their colony to be an outpost of toleration in the New World; rather, they intended it to be a "Zion in the wilderness," a model of purity and orthodoxy, with all backsliders subject to immediate correction.

John Winthrop and Massachusetts Bay

The civil government of the colony was guided by a similar authoritarian spirit. Men like John Winthrop, the first governor of Massachusetts Bay, believed that it was not the duty of the governors of society to act as the direct representatives of their constituents but rather to decide, independently, what measures were in the best interests of the total society. The original charter of 1629 gave all power in the colony to a General Court composed of only a small number of shareholders in the company. On arriving in Massachusetts, many disfranchised settlers immediately protested against this provision and caused the franchise to be widened to include all church members. These "freemen" were given the right to vote in the General Court once each year for a governor and a Council of Assistants. Although the charter of 1629 technically gave the General Court the power to decide on all matters affecting the colony, the members of the ruling elite initially refused to allow the freemen in the General Court to take part in the lawmaking process on the grounds that their numbers would render the court inefficient.

In 1634 the General Court adopted a new plan of representation whereby the freemen of each town would be permitted to select two or three delegates and assistants, elected separately but sitting together in the General Court, who would be responsible for all legislation. There was always tension existing between the smaller, more prestigious group of assistants and the larger group of deputies. In 1644, as a result of this continuing tension, the two groups were officially lodged in separate houses of the General Court, with each house reserving a veto power over the other.

Despite the authoritarian tendencies of the Massachusetts Bay colony, a spirit of community developed there as perhaps in no other colony. The same spirit that caused the residents of Massachusetts to report on their neighbours for deviation from the true principles of Puritan morality also prompted them to be extraordinarily solicitous about their neighbours' needs. Although life in Massachusetts was made difficult for those who dissented from the prevailing orthodoxy, it was marked by a feeling of attachment and community for those who lived within the enforced consensus of the society.

Connecticut and Rhode Island

Many New Englanders, however, refused to live within the orthodoxy imposed by the ruling elite of Massachusetts, and both Connecticut and Rhode Island were founded as a by-product of their discontent. The Reverend Thomas Hooker, who had arrived in Massachusetts Bay in 1633, soon found himself in opposition to the colony's restrictive policy regarding the admission of church members and to the oligarchic power of the leaders of the colony. Motivated both by a distaste for the religious and political structure of Massachusetts and by a desire to open up new land, Hooker and his followers began moving into the Connecticut valley in 1635. By 1636 they had succeeded in founding three towns—Hartford, Windsor, and Wethersford. In 1638 the separate colony of New Haven was founded, and in 1662 Connecticut and Rhode Island merged under one charter.

Roger Williams, the man closely associated with the founding of Rhode Island, was banished from Massachusetts because of his unwillingness to conform to the orthodoxy established in that colony. Williams' views conflicted with those of the ruling hierarchy of Massachusetts in several important ways. His own strict criteria for determining who was regenerate, and therefore eligible for church membership, finally led him to deny any practical way to admit anyone into the church. Once he recognized that no church could ensure the purity of its congregation, he ceased using purity as a criterion and instead opened church membership to nearly everyone in the community. Moreover, Williams showed distinctly Separatist leanings, preaching that the Puritan church could not possibly achieve purity as long as it remained within the Church of England. Finally, and perhaps most serious, he openly disputed the right of the Massachusetts leaders to occupy Indian land without first purchasing it from the natives.

The unpopularity of Williams' views forced him to flee Massachusetts Bay for Providence in 1636. In 1639 William Coddington, another dissenter in Massachusetts, settled his congregation in Newport. Four years later Samuel Gorton, another minister banished from Massachusetts Bay because of his differences with the ruling oligarchy, settled in Shawomet (later renamed Warwick). In 1644 these three communities joined with a fourth in Portsmouth under one charter to become one colony called Providence Plantation in Narragansett Bay.

The early settlers of New Hampshire and Maine were also ruled by the government of Massachusetts Bay. New Hampshire was permanently separated from Massachusetts in 1692, although it was not until 1741 that it was given its own royal governor. Maine remained under the jurisdiction of Massachusetts until 1820.

The middle colonies. New Netherland, founded in 1624 at Fort Orange (now Albany) by the Dutch West India Company, was but one element in a wider program of Dutch expansion in the first half of the 17th century. The English captured the colony of New Netherland in 1664; it was renamed New York, after James, Duke of York, brother of Charles II, and was placed under the proprietary control of the duke. In return for an annual gift to the king of 40 beaver skins, the Duke of York and his resident Board of Governors were given extraordinary discretion in the ruling of the colony. Although the grant to the Duke of York made mention of a representative assembly, the duke was not legally obliged to summon it and in fact did not summon it until 1683. The duke's interest in the colony was chiefly economic, not political, but most of his efforts to derive economic gain from New York proved futile. Indians, foreign interlopers (the Dutch actually recaptured New York in 1673 and held it for more than a year), and the success of the colonists in evading taxes made the proprietor's job a frustrating one.

New York

In February 1685 the Duke of York found himself not only proprietor of New York but also king of England, a fact that changed the status of New York from that of a proprietary to a royal colony. The process of royal consolidation was accelerated when in 1688 the colony, along with the New England and New Jersey colonies, was made part of the ill-fated Dominion of New England. In 1691 Jacob Leisler, a German merchant living on Long Island, led a successful revolt against the rule of the deputy governor, Francis Nicholson. The revolt, which was a product of dissatisfaction with a small aristocratic ruling elite and a more general dislike of the consolidated scheme of government of the Dominion of New England, served to hasten the demise of the dominion.

Pennsylvania, in part because of the liberal policies of its founder, William Penn, was destined to become the most diverse, dynamic, and prosperous of all the North American colonies. Penn himself was a liberal, but by no means radical, English Whig. His Quaker faith was marked not by the religious extremism of some Quaker leaders of the day but rather by an adherence to certain dominant

Pennsylvania

tenets of the faith—liberty of conscience and pacifism—and by an attachment to some of the basic tenets of Whig doctrine. William Penn sought to implement these ideals in his "holy experiment" in the New World.

Penn received his grant of land along the Delaware River in 1681 from Charles II. The first "frame of government" proposed by Penn in 1682 provided for a council and an assembly, each to be elected by the freeholders of the colony. The council was to have the sole power of initiating legislation; the lower house could only approve or veto bills submitted by the council. After numerous objections about the "oligarchic" nature of this form of government, Penn issued a second frame of government in 1682 and then a third in 1696, but even these did not wholly satisfy the residents of the colony. Finally, in 1701, a Charter of Privileges, giving the lower house all legislative power and transforming the council into an appointive body with advisory functions only, was approved by the citizens. The Charter of Privileges, like the other three frames of government, continued to guarantee the principle of religious toleration to all Protestants.

Pennsylvania prospered from the outset. Although there was some jealousy between the original settlers (who had received the best land and important commercial privileges) and the later arrivals, economic opportunity in Pennsylvania was on the whole greater than in any other colony. Beginning in 1683 with the immigration of Germans into the Delaware valley and continuing with an enormous influx of Irish and Scotch-Irish in the 1720s and '30s, the population of Pennsylvania increased and diversified. The fertile soil of the countryside, in conjunction with a generous government land policy, kept immigration at high levels throughout the 18th century. Ultimately, however, the continuing influx of European settlers hungry for land spelled doom for the pacific Indian policy initially envisioned by William Penn. "Economic opportunity" for European settlers often depended on the dislocation, and frequent extermination, of the Indian residents who had initially occupied the land in Penn's colony.

New Jersey

New Jersey remained in the shadow of both New York and Pennsylvania throughout most of the colonial period. Part of the territory ceded to the Duke of York by the English crown in 1664 lay in what would later become the colony of New Jersey. The Duke of York in turn granted that portion of his lands to John Berkeley and George Carteret, two close friends and allies of the king. In 1665 Berkeley and Carteret established a proprietary government under their own direction. Constant clashes, however, developed between the New Jersey and the New York proprietors over the precise nature of the New Jersey grant. The legal status of New Jersey became even more tangled when Berkeley sold his half interest in the colony to two Quakers, who in turn placed the management of the colony in the hands of three trustees, one of whom was William Penn. The area was then divided into East Jersey, controlled by Carteret, and West Jersey, controlled by Penn and the other Quaker trustees. In 1682 the Quakers bought East Jersey. A multiplicity of owners and an uncertainty of administration caused both colonists and colonizers to feel dissatisfied with the proprietary arrangement, and in 1702 the crown united the two Jerseys into a single royal province.

When the Quakers purchased East Jersey, they also acquired the tract of land that was to become Delaware, in order to protect their water route to Pennsylvania. That territory remained part of the Pennsylvania colony until 1704, when it was given an assembly of its own. It remained under the Pennsylvania governor, however, until the Revolution.

The Carolinas and Georgia. The English crown had issued grants to the Carolina territory as early as 1629, but it was not until 1663 that a group of eight proprietors—most of them men of extraordinary wealth and power even by English standards—actually began colonizing the area. The proprietors hoped to grow silk in the warm climate of the Carolinas, but all efforts to produce that valuable commodity failed. Moreover, it proved difficult to attract settlers to the Carolinas; it was not until 1718, after a series of violent Indian wars had subsided, that the population began to increase substantially. The pattern of settlement, once begun, followed two paths. North Carolina, which was largely cut off from the European and Caribbean trade by its unpromising coastline, developed into a colony of small to medium farms. South Carolina, with close ties to both the Caribbean and Europe, produced rice and, after 1742, indigo for a world market. The early settlers in both areas came primarily from the West Indian colonies. This pattern of migration was not, however, as distinctive in North Carolina, where many of the residents were part of the spillover from the natural expansion of Virginians southward.

The Fundamental Constitutions

The original frame of government for the Carolinas, the Fundamental Constitutions, drafted in 1669 by Anthony Ashley Cooper (Lord Shaftesbury) with the help of the philosopher John Locke, was largely ineffective because of its restrictive and feudal nature. The Fundamental Constitutions was abandoned in 1693 and replaced by a frame of government diminishing the powers of the proprietors and increasing the prerogatives of the provincial assembly. In 1729, primarily because of the proprietors' inability to meet the pressing problems of defense, the Carolinas were converted into the two separate royal colonies of North and South Carolina.

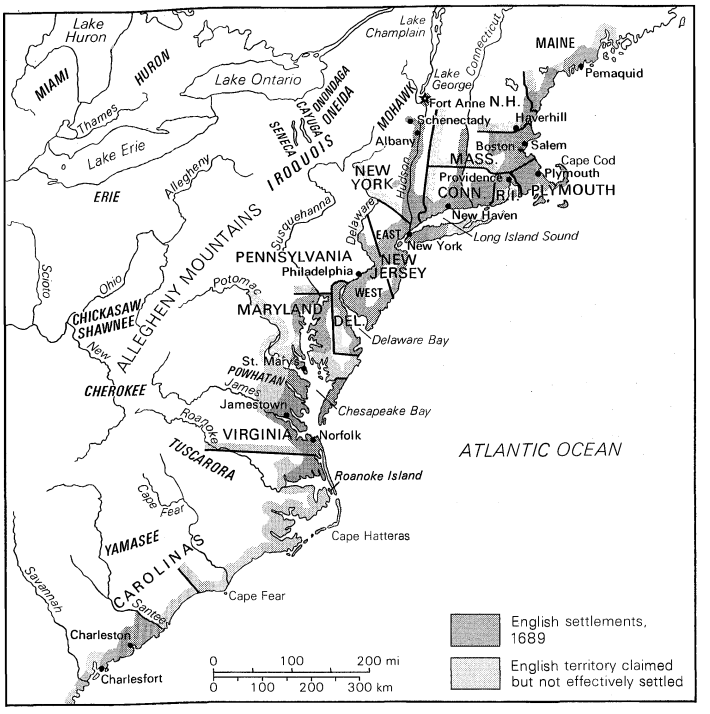
The proprietors of Georgia, led by James Oglethorpe, were wealthy philanthropic English gentlemen. It was Oglethorpe's plan to transport imprisoned debtors to Georgia where they could rehabilitate themselves by profitable labour and make money for the proprietors in the process. Those who actually settled in Georgia—and by no means all of them were impoverished debtors—encountered a highly restrictive economic and social system. Oglethorpe and his partners limited the size of individual landholdings to 500 acres, prohibited slavery, forbade the drinking of rum, and instituted a system of inheritance that further restricted the accumulation of large estates. The regulations, though noble in intention, created considerable tension between some of the more enterprising settlers and the proprietors. Moreover, the economy did not live up to the expectations of the colony's promoters. The silk industry in Georgia, like that in the Carolinas, failed to produce even one profitable crop.

The settlers were also dissatisfied with the political structure of the colony; the proprietors, concerned primarily with keeping close control over their utopian experiment, failed to provide for local institutions of self-government. As protests against the proprietors' policies mounted, the crown in 1752 assumed control over the colony; subsequently, many of the restrictions that the settlers had complained about, notably those discouraging the institution of slavery, were lifted.

IMPERIAL ORGANIZATION

British policy toward the American colonies was inevitably affected by the domestic politics of England; since the politics of England in the 17th and 18th centuries were never wholly stable, it is not surprising that British colonial policy during those years never developed along clear and consistent lines. During the first half century of colonization, it was even more difficult for England to establish an intelligent colonial policy because of the very disorganization of the colonies themselves. It was nearly impossible for England to predict what role Virginia, Maryland, Massachusetts, Connecticut, and Rhode Island would play in the overall scheme of empire because of the diversity of the aims and governmental structures of those colonies. By 1660, however, England had taken the first steps in reorganizing her empire in a more profitable manner. The Navigation Act of 1660, a modification and amplification of a temporary series of acts passed in 1651, provided that goods bound to England or to English colonies, regardless of origin, must be shipped only in English vessels; that three-fourths of the personnel of those ships be Englishmen; and that certain "enumerated articles," such as sugar, cotton, and tobacco, be shipped only to England, with trade in those items with other nations prohibited. This last provision hit Virginia and Maryland particularly hard; although those two colonies were awarded a monopoly over the English tobacco market at the same time that they

Navigation Acts



English colonies in 17th-century North America.

Adapted from R. Treharne and H. Fullard (eds.), *Muir's Historical Atlas: Medieval and Modern*, 9th ed. (1962); George Philip and Son, Ltd., London

were prohibited from marketing their tobacco elsewhere, there was no way that England alone could absorb their tobacco production.

The 1660 act proved inadequate to safeguard the entire British commercial empire, and in subsequent years other navigation acts were passed, strengthening the system. In 1663 Parliament passed an act requiring all vessels with European goods bound for the colonies to pass first through English ports to pay customs duties. In 1673 Parliament, in order to prevent merchants from shipping the enumerated articles from colony to colony in the coastal trade and then taking them to a foreign country, required that merchants post bond guaranteeing that those goods be taken only to England. Finally, in 1696 Parliament established a Board of Trade to oversee Britain's commercial empire, instituted mechanisms to ensure that the colonial governors aided in the enforcement of trade regulations, and set up vice admiralty courts in America for the prosecution of those who violated the Navigation Acts. On the whole, this attempt at imperial consolidation—what some historians have called the process of Anglicization—was successful in bringing the economic activities of the colonies under closer crown control. While a significant amount of colonial trade continued to evade British regulation, it is nevertheless clear that the British were at least partially successful in imposing greater commercial and political order on the American colonies during the period from the late-17th to the mid-18th century.

In addition to the agencies of royal control in England, there were a number of royal officials in America responsible not only for aiding in the regulation of England's commercial empire but also for overseeing the internal affairs of the colonies. The weaknesses of royal authority in the politics of provincial America were striking. In some areas, particularly in the corporate colonies of New England during the 17th century and in the proprietary colonies throughout their entire existence, direct royal authority in the person of a governor responsible to the crown was nonexistent. The absence of a royal governor in those colonies had a particularly deleterious effect on the enforcement of trade regulations. In fact, the lack of royal control over the political and commercial activities of New England prompted the Board of Trade to overturn the Massachusetts Bay charter in 1684 and to consolidate Massachusetts, along with the other New England colonies and New York, into the Dominion of New England. After the colonists, aided by the turmoil of the Revolution of 1688 in England, succeeded in overthrowing the dominion scheme, the crown installed a royal governor in Massachusetts to protect its interests.

Powers of the royal governors

In those colonies with royal governors—the number of those colonies grew from one in 1650 to eight in 1760—the crown possessed a mechanism by which to ensure that royal policy was enforced. The Privy Council issued each royal governor in America a set of instructions carefully defining the limits of provincial authority. The royal governors were to have the power to decide when to call the provincial assemblies together, to prorogue, or dissolve, the assemblies, and to veto any legislation passed by those assemblies. The governor's power over other aspects of the political structure of the colony was just as great. In most royal colonies he was the one official primarily responsible for the composition of the upper houses of the colonial legislatures and for the appointment of important provincial officials, such as the treasurer, attorney general, and all colonial judges. Moreover, the governor had enormous patronage powers over the local agencies of government. The officials of the county court, who were the principal agents of local government, were appointed by the governor in most of the royal colonies. Thus, the governor had direct or indirect control over every agency of government in America.

THE GROWTH OF PROVINCIAL POWER

Political growth. The distance separating England and America, the powerful pressures exerted on royal officials by Americans, and the inevitable inefficiency of any large bureaucracy all served to weaken royal power and to strengthen the hold of provincial leaders on the affairs of their respective colonies. During the 18th century the colonial legislatures gained control over their own parliamentary prerogatives, achieved primary responsibility for legislation affecting taxation and defense, and ultimately took control over the salaries paid to royal officials. Provincial leaders also made significant inroads into the governor's patronage powers. Although theoretically the governor continued to control the appointments of local officials, in reality he most often automatically followed the recommendations of the provincial leaders in the localities in question. Similarly, the governor's councils, theoretically agents of royal authority, came to be dominated by prominent provincial leaders who tended to reflect the interests of the leadership of the lower house of assembly rather than those of the royal government in London.

Thus, by the mid-18th century most political power in America was concentrated in the hands of provincial rather than royal officials. These provincial leaders undoubtedly represented the interests of their constituents more faithfully than any royal official could, but it is clear that the politics of provincial America were hardly democratic by modern standards. In general, both social prestige and political power tended to be determined by economic standing; and the economic resources of colonial America, though not as unevenly distributed as in Europe, were nevertheless controlled by relatively few men.

Rise of the planter class

In the Chesapeake societies of Virginia and Maryland, and particularly in the regions east of the Blue Ridge mountains, a planter class came to dominate nearly every aspect of those colonies' economic life. These same planters, joined by a few prominent merchants and lawyers, dominated the two most important agencies of local government—the county courts and the provincial assemblies. This extraordinary concentration of power in the hands of a wealthy few occurred in spite of the fact that a large percentage of the free adult male population (some have estimated as high as 80 to 90 percent) was able to participate in the political process. The ordinary citizens of the Chesapeake society, and those of most colonies, nevertheless continued to defer to those whom they considered to be their "betters." Although the societal ethic that enabled power to be concentrated in the hands of a few was hardly a democratic one, there is little evidence, at least for Virginia and Maryland, that the people of those societies were dissatisfied with their rulers. In general they believed that their local officials ruled responsively.

In the Carolinas a small group of rice and indigo planters monopolized much of the wealth. As in Virginia and Maryland, the planter class came to constitute a social elite. As a rule the planter class of the Carolinas did not have the same long tradition of responsible government as did the ruling oligarchies of Virginia and Maryland, and, as a consequence, they tended to be absentee landlords and governors, often passing much of their time in Charleston, away from their plantations and their political responsibilities.

The western regions of both the Chesapeake and Carolina societies displayed distinctive characteristics of their own. Ruling traditions were fewer, accumulations of land and wealth less striking, and the social hierarchy less rigid in the west. In fact, in some western areas antagonism toward the restrictiveness of the east and toward eastern control of the political structure led to actual conflict. In both North and South Carolina armed risings of varying intensity erupted against the unresponsive nature of the eastern ruling elite. As the 18th century progressed, however, and as more men accumulated wealth and social prestige, the societies of the west came more closely to resemble those of the east.

New England society was more diverse and the political system less oligarchic than that of the South. In New England the mechanisms of town government served to broaden popular participation in government beyond the narrow base of the county courts.

New England town meetings

The town meetings, which elected the members of the provincial assemblies, were open to nearly all free adult males. Despite this, a relatively small group of men dominated the provincial governments of New England. As in the South, men of high occupational status and social prestige were closely concentrated in leadership positions in their respective colonies; in New England, merchants, lawyers, and to a lesser extent clergymen made up the bulk of the social and political elite.

The social and political structure of the middle colonies was more diverse than that of any other region in America. New York, with its extensive system of manors and manor lords, often displayed genuinely feudal characteristics. The tenants on large manors often found it impossible to escape the influence of their manor lords. The administration of justice, the election of representatives, and the collection of taxes often took place on the manor itself. As a consequence, the large landowning families exercised an inordinate amount of economic and political power. The Great Rebellion of 1766, a short-lived outburst directed against the manor lords, was a symptom of the widespread discontent among the lower and middle classes. By contrast, Pennsylvania's governmental system was more open and responsive than that of any other colony in America. A unicameral legislature, free from the restraints imposed by a powerful governor's council, allowed Pennsylvania to be relatively independent of the influence of both the crown and the proprietor. This fact, in combination with the tolerant and relatively egalitarian bent of the early Quaker settlers and the subsequent immigration of large numbers of Europeans, made the social and political structure of Pennsylvania more democratic but more faction ridden than that of any other colony.

Population growth. The increasing political autonomy of the American colonies was a natural reflection of their increased stature in the overall scheme of the British Empire. In 1650 the population of the colonies had been about 52,000; in 1700 it was perhaps 250,000, and by 1760 it was approaching 1,700,000. Virginia had increased from about 54,000 in 1700 to approximately 340,000 in 1760. Pennsylvania had begun with about 500 settlers in 1681 and had attracted at least 250,000 people by 1760. And America's cities were beginning to grow as well. By 1765 Boston had reached 15,000; New York City, 16,000–17,000; and Philadelphia, the largest city in the colonies, 20,000.

Part of that population growth was the result of the involuntary immigration of African slaves. During the 17th century slaves remained a tiny minority of the population. By the mid-18th century, after Southern colonists discovered that the profits generated by their plantations could support the relatively large initial investments needed for slave labour, the volume of the slave trade increased markedly. In Virginia the slave population leaped from about 2,000 in 1670 to perhaps 23,000 in 1715 and reached 150,000 on the eve of the American Revolution. In South Carolina it was even more dramatic. In 1700 there were probably no more than 2,500 blacks in the population; by 1765 there were 80,000–90,000, with blacks outnumbering whites by about 2 to 1.

African slaves

One of the principal attractions to the immigrants who moved to America voluntarily was the availability of inexpensive arable land. The westward migration to America's frontier—in the early 17th century all of America was a frontier, and by the 18th century the frontier ranged anywhere from 10 to 200 miles from the coastline—was to become one of the distinctive elements in American history. English Puritans, beginning in 1629 and continuing through 1640, were the first to immigrate in large numbers to America. Throughout the 17th century most of the immigrants were English; but, beginning in the second decade of the 18th century, a wave of Germans, principally from the Rhineland Palatinate, arrived in America: by 1770 between 225,000 and 250,000 Germans had emigrated to America, more than 70 percent of them settling in the middle colonies, where generous land policies and religious toleration made life more comfortable for them. The Scotch-Irish and Irish immigration, which began on a large scale after 1713 and continued past the American Revolution, was more evenly distributed. By 1750 both Scotch-Irish and Irish could be found in the western portions of nearly every colony. In almost all the regions in which Europeans sought greater economic opportunity, however, that same quest for independence and self-sufficiency led to tragic conflict with Indians over the control of land. And in nearly every instance the outcome was similar: the Europeans, failing to respect Indian claims either to land or to cultural autonomy, pushed the Indians of North America further and further into the periphery.

Economic growth. Provincial America came to be less dependent upon subsistence agriculture and more on the cultivation and manufacture of products for the world market. Land, which initially served only individual needs, came to be the fundamental source of economic enterprise. The independent yeoman farmer continued to exist, particularly in New England and the middle colonies, but most settled land in North America by 1750 was devoted to the cultivation of a cash crop. New England turned its land over to the raising of meat products for export. The middle colonies were the principal producers of grains. By 1700 Philadelphia exported more than 350,000 bushels of wheat and more than 18,000 tons of flour annually. The Southern colonies were, of course, even more closely tied to the cash crop system. South Carolina, aided by British incentives, turned to the production of rice and indigo. North Carolina, although less oriented toward the market economy than South Carolina, was nevertheless one of the principal suppliers of naval stores. Virginia and Maryland steadily increased their economic dependence on tobacco and on the London merchants who purchased that tobacco; and for the most part they ignored those who recommended that they diversify their economies by turning part of their land over to the cultivation of wheat. Their near-total dependence upon the world tobacco price would ultimately prove disastrous, but for most of the 18th century Virginia and Maryland soil remained productive enough to make a single-crop system reasonably profitable.

Predominance of cash crops

As America evolved from subsistence to commercial agriculture, an influential commercial class increased its power in nearly every colony. Boston was the centre of the merchant elite of New England, who not only dominated economic life but also wielded social and political power as well. Merchants like James De Lancey and Philip Livingston in New York and Joseph Galloway, Robert Morris, and Thomas Wharton in Philadelphia exerted an influence far beyond the confines of their occupations. In Charleston the Pinckney, Rutledge, and Lowndes families controlled much of the trade that passed through that port. Even in Virginia, where a strong merchant class

was nonexistent, those people with the most economic and political power were those commercial farmers who best combined the occupations of merchant and farmer. And it is clear that the commercial importance of the colonies was increasing. During the years 1700–10, approximately £265,000 sterling was exported annually to Great Britain from the colonies, with roughly the same amount being imported by the Americans from Great Britain. By the decade 1760–70, that figure had risen to more than £1,000,000 sterling of goods exported annually to Great Britain and £1,760,000 annually imported from Great Britain.

CULTURAL AND RELIGIOUS DEVELOPMENT

Colonial culture. America's intellectual attainments during the 17th and 18th centuries, while not inferior to those of the nations of Europe, were nevertheless of a decidedly different character. It was the techniques of applied science that most excited the minds of Americans, who, faced with the problem of subduing an often wild and unruly land, saw in science the best way to explain, and eventually to harness, those forces around them. Ultimately this scientific mode of thought might be applied to the problems of civil society as well, but for the most part the emphasis in colonial America remained on science and technology, not politics or metaphysics. Typical of America's peculiar scientific genius was John Bartram of Pennsylvania, who collected and classified important botanical data from the New World. The American Philosophical Society, founded in 1744, is justly remembered as the focus of intellectual life in America. Men like David Rittenhouse, an astronomer who built the first planetarium in America; Cadwallader Colden, the lieutenant governor of New York, whose accomplishments as a botanist and as an anthropologist probably outmatched his achievements as a politician; and Benjamin Rush, a pioneer in numerous areas of social reform as well as one of colonial America's foremost physicians, were among the many active members of the society. At the centre of the American Philosophical Society was one of its founders, Benjamin Franklin, who (in his experiments concerning the flow of electricity) proved to be one of the few American scientists to achieve a major theoretical breakthrough but who was more adept at the kinds of applied research that resulted in the manufacture of more efficient fireplaces and the development of the lightning rod.

American Philosophical Society

American cultural achievements in nonscientific fields were less impressive. American literature, at least in the traditional European forms, was nearly nonexistent. The most important American contribution to literature was neither in fiction nor in metaphysics but rather in such histories as Robert Beverley's *History and Present State of Virginia* (1705) or William Byrd's *History of the Dividing Line* (1728–29, but not published until 1841). The most important cultural medium in America was not the book but the newspaper. The high cost of printing tended to eliminate all but the most vital news, and thus local gossip or extended speculative efforts were sacrificed so that more important material such as classified advertisements and reports of crop prices could be included. Next to newspapers, almanacs were the most popular literary form in America, Franklin's *Poor Richard's* being only the most famous among scores of similar projects. Not until 1741 and the first installment of Franklin's *General Magazine* did literary magazines begin to make their first appearance in America. Most of the 18th-century magazines, however, failed to attract subscribers, and nearly all of them collapsed after only a few years of operation.

Art and drama

The visual and performing arts, though flourishing somewhat more than literature, were nevertheless slow to achieve real distinction in America. America did produce one good historical painter in Benjamin West and two excellent portrait painters in John Copley and Gilbert Stuart; but it is not without significance that all three men passed much of their lives in London, where they received more attention and higher fees.

The Southern colonies, particularly Charleston, seemed to be more interested in providing good theatre for their residents than did other regions, but in no colony did the theatre approach the excellence of that of Europe. In New England, Puritan influence was an obstacle to the performance of plays, and even in cosmopolitan Philadelphia the Quakers for a long time discouraged the development of the dramatic arts.

If Americans in the colonial period did not excel in achieving a high level of traditional cultural attainment, they did manage at least to disseminate what culture they had in a manner slightly more equitable than that of most nations of the world. Newspapers and almanacs, though hardly on the same intellectual level as the *Encyclopédie* produced by the European philosophes, probably had a wider audience than any European cultural medium. The New England colonies, although they did not always manage to keep pace with population growth, pioneered in the field of public education. Outside of New England, education remained the preserve of those who could afford to send their children to private schools, although the existence of privately supported but tuition-free charity schools and of relatively inexpensive "academies" made it possible for the children of the American middle class to receive at least some education. The principal institutions of higher learning—Harvard (1636), William and Mary (1693), Yale (1701), Princeton (1747), Pennsylvania (a college since 1755), King's College (1754, now Columbia), Rhode Island College (1764, now Brown), Queen's College (1766, now Rutgers), and Dartmouth (1769)—served the upper class almost exclusively; and most of them had a close relationship with a particular religious point of view (*e.g.*, Harvard was a training ground for Congregational ministers, and Princeton was closely associated with Presbyterianism).

Public education

The Great Awakening. A series of religious revivals known collectively as the Great Awakening swept over the colonies in the 1730s and '40s. Its impact was first felt in the middle colonies, where Theodore J. Frelinghuysen, a minister of the Dutch Reformed church, began preaching in the 1720s. In New England, in the early 1730s men such as Jonathan Edwards, perhaps the most learned theologian of the 18th century, were responsible for a reawakening of religious fervour. By the late 1740s the movement had extended into the Southern colonies, where itinerant preachers such as Samuel Davies and George Whitefield exerted considerable influence, particularly in the backcountry.

The Great Awakening represented a reaction against the increasing secularization of society and against the corporate and materialistic nature of the principal churches of American society. By making conversion the initial step on the road to salvation and by opening up the conversion experience to all who recognized their own sinfulness, the ministers of the Great Awakening, some intentionally and others unwittingly, democratized Calvinist theology. The technique of many of the preachers of the Great Awakening was to inspire in their listeners a fear of the consequences of their sinful lives and a respect for the omnipotence of God. This sense of the ferocity of God was often tempered by the implied promise that a rejection of worldliness and a return to faith would result in a return to grace and an avoidance of the horrible punishments of an angry God. There was a certain contradictory quality about these two strains of Great Awakening theology, however. Predestination, one of the principal tenets of the Calvinist theology of most of the ministers of the Great Awakening, was ultimately incompatible with the promise that man could, by a voluntary act of faith, achieve salvation by his own efforts. Furthermore, the call for a return to complete faith and the emphasis on the omnipotence of God was the very antithesis of Enlightenment thought, which called for a greater questioning of faith and a diminishing role for God in the daily affairs of man. On the other hand, Jonathan Edwards, one of the principal figures of the Great Awakening in America, explicitly drew on the thought of men like John Locke and Isaac Newton in an attempt to make religion rational. Perhaps most important, the evangelical styles of religious worship promoted by the Great Awakening helped make the religious doctrines of many of the insurgent church denominations—particularly those of the Baptists and the Methodists—more accessible to a wider cross section of

the American population. This expansion in church membership extended to blacks as well as whites, and the ritual forms of Evangelical Protestantism possessed features that facilitated the syncretism of African and American forms of religious worship.

AMERICA, ENGLAND, AND THE WIDER WORLD

The American colonies, though in many ways isolated from the nations of Europe, were nevertheless continually subject to diplomatic and military pressures from abroad. In particular, Spain and France were always nearby, waiting to exploit any signs of English weakness in America in order to increase their commercial and territorial designs on the North American mainland. The Great War for the Empire, or the French and Indian War as the Americans called it (see EUROPE), was but another round in a century of warfare between the major European powers. First in King William's War (1689–97), then in Queen Anne's War (1702–13), and in King George's War (1744–48; the American phase of the War of the Austrian Succession), Englishmen and Frenchmen had vied for control over the Indians, for possession of the territory lying to the north of the North American colonies, for access to the trade in the Northwest, and for commercial superiority in the West Indies. In most of these encounters France had been aided by her ally, Spain. Because of its own holdings immediately south and west of the British colonies and in the Caribbean, Spain realized that it was in its own interest to join with the French in limiting British expansion. The culmination of these struggles came in 1754 with the Great War for the Empire. Whereas previous contests between Great Britain and France in America had been mostly provincial affairs, with American colonists doing most of the fighting for the British, the Great War for the Empire saw sizable commitments of English troops to America. The strategy of the English under William Pitt was to allow their ally, Prussia, to carry the brunt of the fighting in Europe, thus freeing the English to concentrate their troops in America.

The French and Indian War

The French, despite the fact that they were outnumbered 15 to 1 by the English colonial population in America, were nevertheless well equipped to hold their own against the British. They had a larger military organization in America than did the English, their troops were better trained, and they were more successful than the British in forming military alliances with the Indians. The early engagements of the war went to the French; the surrender of George Washington to a superior French force at Fort Necessity, the annihilation of General Edward Braddock at the Monongahela River, and French victories at Oswego and Fort William Henry all made it seem as if the war would be a short and unsuccessful one for the British. Even as these defeats took place, however, the English were able to increase their supplies of both men and material in America. By 1758, with its strength finally up to a satisfactory level, England began to implement its larger strategy, which involved sending a combined land and sea force to gain control of the St. Lawrence and a large land force aimed at Fort Ticonderoga to eliminate French control of Lake Champlain. The first expedition against the French at Ticonderoga was a disaster, as General James Abercrombie led about 15,000 British and colonial troops in an attack against the French before his forces were adequately prepared. The English assault on Louisburg, the key to the St. Lawrence, was more successful. In July 1758 Lord Jeffrey Amherst led a naval attack in which his troops landed on the shores from small boats, established beachheads, and then captured the fort at Louisburg.

The capture of Quebec

In 1759, after several months of sporadic fighting, the forces of James Wolfe captured Quebec from the French army led by the Marquis de Montcalm. This was probably the turning point of the war. By the fall of 1760 the British had taken Montreal, and England possessed practical control of all of the North American continent. It took another two years for England to defeat her rivals in other parts of the world, but the contest for control of North America had been settled.

In the Treaty of Paris of 1763, England took possession of all of Canada, East and West Florida, all territory east of the Mississippi in North America, and St. Vincent, Tobago, and Dominica in the Caribbean. At the time, the British victory seemed one of the greatest in its history. The British Empire in North America had not only been secured but also greatly expanded. But, in winning the war, Britain had dissolved the empire's most potent material adhesives. Conflicts arose as the needs and interests of the British Empire began to differ from those of the American colonies; and the colonies, now economically powerful, culturally distinct, and steadily becoming more independent politically, would ultimately rebel before submitting to the English plan of empire. (R.R.B.)

The American Revolution and the early federal republic

PRELUDE TO REVOLUTION

Britain's victory over France in the Great War for the Empire had been won at very great cost. British government expenditures, which had amounted to nearly £6,500,000 annually before the war, rose to about £14,500,000 annually during the war. As a result, the burden of taxation in England was probably the highest in the country's history, much of it borne by the politically influential landed classes. Furthermore, with the acquisition of the vast domain of Canada and the prospect of holding British territories both against the various nations of Indians and against the Spaniards to the south and west, the costs of colonial defense could be expected to continue indefinitely. Parliament, moreover, had voted Massachusetts a generous sum in compensation for its war expenses. It therefore seemed reasonable to British opinion that some of the future burden of payment should be shifted to the colonists themselves—who until then had been lightly taxed and indeed lightly governed.

The prolonged wars had also revealed the need to tighten the administration of the loosely run and widely scattered elements of the British Empire. If the course of the war had confirmed the necessity, the end of the war presented the opportunity. The acquisition of Canada required London officials to take responsibility for the unsettled western territories, now freed from the threat of French occupation. The British soon moved to take charge of the whole field of Indian relations. By royal proclamation (1763) a line was drawn down the Appalachians marking the limit of settlement from the British colonies, beyond which Indian trade was to be conducted strictly through British-appointed commissioners. These steps were not in time to prevent a serious uprising under the Ottawa chief Pontiac, however; and the proclamation, which sprang in part from a respect for Indian rights, caused consternation among British colonists for two reasons. It meant that limits were being set to the prospects of settlement and speculation in western lands, and it took control of the west out of colonial hands. The most ambitious men in the colonies thus saw the proclamation as a loss of power to control their own fortunes.

Proclamation of 1763

The tax controversy. George Grenville, who was named prime minister in 1763, was soon looking to meet the costs of defense by raising revenue in the colonies. The first measure was the Plantation Act of 1764, usually called the Sugar, or Revenue, Act, which reduced to a mere threepence the duty on imported foreign molasses but linked with this a high duty on refined sugar and a prohibition on foreign rum (the needs of the British treasury were carefully balanced with those of West Indies planters and New England distillers). The last measure of this kind (1733) had not been enforced, but this time the government set up a system of customs houses, staffed by British officers, and even established a vice-admiralty court. Sitting at Halifax, N.S., the court heard very few cases, but in principle it appeared to threaten the cherished British privilege of trials by local juries. Boston further objected to the tax's revenue-raising aspect on constitutional grounds, but, despite some expressions of anxiety, the colonies in general acquiesced.

Parliament next affected colonial economic prospects by passing a Currency Act (1764) to withdraw paper currencies, many of them surviving from the war period,

from circulation. This was not done to restrict economic growth so much as to take out currency that was thought to be unsound, but it did severely reduce the circulating medium during the difficult postwar period and further indicated that such matters were subject to British control.

Grenville's next move was a stamp duty, to be raised on a wide variety of transactions, including legal writs, newspaper advertisements, and ships' bills of lading. The colonies were duly consulted and offered no alternative suggestions. The feeling in London, shared by Benjamin Franklin, was that, after making formal objections, the colonies would accept the new taxes as they had the earlier ones. But the Stamp Act (1765) hit harder and deeper than any previous parliamentary measure. As some agents had already pointed out, because of postwar economic difficulties the colonies were short of ready funds. (In Virginia this shortage was so serious that the province's treasurer, John Robinson, who was also speaker of the assembly, manipulated and redistributed paper money that had been officially withdrawn from circulation by the Currency Act; a large proportion of the landed gentry benefited from this largesse.) The Stamp Act struck at vital points of colonial economic operations, affecting transactions in trade. It also affected many of the most articulate and influential people in the colonies (lawyers, journalists, bankers). It was, moreover, the first "internal" tax levied directly on the colonies by Parliament. Previous colonial taxes had been levied by local authorities or had been "external" import duties whose primary aim could be viewed as regulating trade for the benefit of the empire as a whole rather than raising revenue. Yet no one, either in Britain or the colonies, fully anticipated the uproar that followed the imposition of these duties. Mobs in Boston and other towns rioted and forced appointed stamp distributors to renounce their posts; legal business was largely halted. Several colonies sent delegations to a Congress in New York in the summer of 1765, where the Stamp Act was denounced as a violation of the Englishman's right to be taxed only through elected representatives, and plans were adopted to impose a nonimportation embargo on British goods.

The Stamp Act

A change of ministry facilitated a change of British policy on taxation. Parliamentary opinion was angered by what it perceived as colonial lawlessness, but British merchants were worried about the embargo on British imports. The Marquis of Rockingham, succeeding Grenville, was persuaded to repeal the Stamp Act—for domestic reasons rather than out of any sympathy with colonial protests. The repeal was passed, however, on the same day as the Declaratory Act, which declared that Parliament had the power to bind or legislate the colonies "in all cases whatsoever." Parliament would not have voted the repeal without this assertion of its authority.

The colonists, jubilant at the repeal of the Stamp Act, drank innumerable toasts, sounded peals of cannon, and were prepared to ignore the Declaratory Act as face-saving window dressing. John Adams, however, warned in his *Dissertation on the Canon and Feudal Law* that Parliament, armed with this view of its powers, would try to tax the colonies again; and this happened in 1767 when Charles Townshend became Chancellor of the Exchequer in a ministry formed by Pitt, now Earl of Chatham. The problem was that Britain's financial burden had not been lifted. Townshend, claiming to take literally the colonial distinction between external and internal taxes, imposed external duties on a wide range of necessities, including lead, glass, paint, paper, and tea, the principal domestic beverage. One ominous result was that colonists now began to believe that the British were developing a long-term plan to reduce the colonies to a subservient position, which they were soon calling "slavery." This view was ill-informed, however. Grenville's measures had been designed as a carefully considered package; apart from some tidying-up legislation, Grenville had had no further plans for the colonies after the Stamp Act. His successors developed further measures, not as extensions of an original plan but because the Stamp Act had been repealed.

The Townshend duties

Nevertheless, the colonists were outraged. In Pennsylvania the lawyer and legislator John Dickinson wrote a series of essays that, appearing in 1767 and 1768 as *Letters from a Farmer in Pennsylvania*, were widely reprinted and exerted great influence in forming a united colonial opposition. Dickinson agreed that Parliament had supreme power where the whole empire was concerned, but he denied that it had power over internal colonial affairs; he quietly implied that the basis of colonial loyalty lay in its utility among equals rather than in obedience owed to a superior.

It proved easier to unite on opinion than on action. Gradually, after much maneuvering and negotiation, a wide-ranging nonimportation policy against British goods was brought into operation. Agreement had not been easy to reach, and the tensions sometimes broke out in acrimonious charges of noncooperation. In addition, the policy had to be enforced by newly created local committees, a process that put a new disciplinary power in the hands of local men who had not had much previous experience in public affairs. There were, as a result, many signs of discontent with the ordering of domestic affairs in some of the colonies—a development that had obvious implications for the future of colonial politics if more action were needed later.

Constitutional differences with Britain. Very few colonists wanted or even envisaged independence at this stage. (Dickinson had hinted at such a possibility with expressions of pain that were obviously sincere.) The colonial struggle for power, although charged with intense feeling, was not an attempt to change government structure but an argument over legal interpretation. The core of the colonial case was that, as British subjects, they were entitled to the same privileges as their fellow subjects in Britain. They could not constitutionally be taxed without their own consent; and, because they were unrepresented in the Parliament that voted the taxes, they had not given this consent. James Otis, in two long pamphlets, ceded all sovereign power to Parliament with this proviso. Others, however, began to question whether Parliament did have lawful power to legislate over the colonies. These doubts were expressed by the late 1760s, when James Wilson, a Scottish immigrant lawyer living in Philadelphia, wrote an essay on the subject. Because of the withdrawal of the Townshend round of duties in 1770, Wilson kept this essay private until new troubles arose in 1774, when he published it as *Considerations on the Nature and Extent of the Legislative Authority of the British Parliament*. In this he fully articulated a view that had been gathering force in the colonies (it was also the opinion of Benjamin Franklin) that Parliament's lawful sovereignty stopped at the shores of Britain.

The official British reply to the colonial case on representation was that the colonies were "virtually" represented in Parliament in the same sense that the large voteless majority of the British public was represented by those who did vote. To this James Otis snorted that, if the majority of the British people did not have the vote, they ought to have it. The idea of colonial members of Parliament, several times suggested, was never a likely solution because of problems of time and distance and because, from the colonists' point of view, colonial members would not have adequate influence.

Virtual representation

The standpoints of the two sides to the controversy could be traced in the language used. The principle of parliamentary sovereignty was expressed in the language of paternalistic authority; the British referred to themselves as parents and to the colonists as children. Colonial tories, who accepted Parliament's case in the interests of social stability, also used this terminology. From this point of view, colonial insubordination was "unnatural," just as the revolt of children against parents was unnatural. The colonists replied to all this in the language of rights. They held that Parliament could do nothing in the colonies that it could not do in Britain because the Americans were protected by all the common-law rights of the British. (When the First Continental Congress met in September 1774, one of its first acts was to affirm that the colonies were entitled to the common law of England.)

Rights, as Richard Bland of Virginia insisted in *The Colonel Dismounted* (as early as 1764), implied equality.

And here he touched on the underlying source of colonial grievance. Americans were being treated as unequals, which they not only resented but also feared would lead to a loss of control of their own affairs. Colonists perceived legal inequality when writs of assistance—essentially, general search warrants—were authorized in Boston in 1761 while closely related "general warrants" were outlawed in two celebrated cases in Britain. Townshend specifically legalized writs of assistance in the colonies in 1767. Dickinson devoted one of his *Letters from a Farmer* to this issue.

When Lord North became prime minister early in 1770, George III had at last found a minister who could work both with himself and with Parliament. British government began to acquire some stability. In 1770, in the face of the American policy of nonimportation, the Townshend tariffs were withdrawn—all except the tax on tea, which was kept for symbolic reasons. Relative calm returned, though it was ruffled on the New England coastline by frequent incidents of defiance of customs officers, who could get no support from local juries. These outbreaks did not win much sympathy from other colonies, but they were serious enough to call for an increase in the number of British regular forces stationed in Boston. One of the most violent clashes occurred in Boston just before the repeal of the Townshend duties. Threatened by mob harassment, a small British detachment opened fire and killed five people, an incident soon known as the Boston Massacre. The soldiers were charged with murder and were given a civilian trial, in which John Adams conducted a successful defense.

The Boston Massacre

The other serious quarrel with British authority occurred in New York, where the assembly refused to accept all the British demands for quartering troops. Before a compromise was reached, Parliament had threatened to suspend the assembly. The episode was ominous because it indicated that Parliament was taking the Declaratory Act at its word; on no previous occasion had the British legislature intervened in the operation of the constitution in an American colony. (Such interventions, which were rare, had come from the crown.)

British intervention in colonial economic affairs occurred again when in 1773 Lord North's administration tried to rescue the East India Company from difficulties that had nothing to do with America. The Tea Act gave the company, which produced tea in India, a monopoly of distribution in the colonies. The company planned to sell its tea through its own agents, eliminating the system of sale by auction to independent merchants. By thus cutting the costs of middlemen, it hoped to undersell the widely purchased, inferior, smuggled tea. This plan naturally affected colonial merchants, and many colonists denounced the act as a plot to induce Americans to buy—and therefore pay the tax on—legally imported tea. Boston was not the only port to threaten to reject the casks of taxed tea, but its reply was the most dramatic—and provocative.

The Boston Tea Party

On Dec. 16, 1773, a party of Bostonians, thinly disguised as Mohawk Indians, boarded the ships at anchor and dumped some £10,000 worth of tea into the harbour. British opinion was outraged, and America's friends in Parliament were immobilized. (American merchants in other cities were also disturbed. Property was property.) In the spring of 1774, with hardly any opposition, Parliament passed a series of measures designed to reduce Massachusetts to order and imperial discipline. The port of Boston was closed; and, in the Massachusetts Government Act, Parliament for the first time actually altered a colonial charter, substituting an appointive council for the elective one established in 1691 and conferring extensive powers on the governor and council. The famous town meeting, a forum for radical thinkers, was outlawed as a political body. To make matters worse, Parliament also passed the Quebec Act for the government of Canada. To the horror of pious New England Calvinists, the Roman Catholic religion was recognized for the French inhabitants. In addition, Upper Canada (*i.e.*, the southern section) was joined to the Mississippi valley for purposes of administration, permanently blocking the prospect of American control of western settlement.

The Continental Congress. There was widespread agreement that this intervention in colonial government could threaten other provinces and could be countered only by collective action. After much intercolonial correspondence, a Continental Congress came into existence, meeting in Philadelphia in September 1774. Every colonial assembly except that of Georgia appointed and sent a delegation. The Virginia delegation's instructions were drafted by Thomas Jefferson and were later published as *A Summary View of the Rights of British America* (1774). Jefferson insisted on the autonomy of colonial legislative power and set forth a highly individualistic view of the basis of American rights. This belief that the American colonies and other members of the British Empire were distinct states united under the king and thus subject only to the king and not to Parliament was shared by several other delegates, notably James Wilson and John Adams, and strongly influenced the Congress.

The Congress' first important decision was one on procedure: whether to vote by colony, each having one vote, or by wealth calculated on a ratio with population. The decision to vote by colony was made on practical grounds—neither wealth nor population could be satisfactorily ascertained—but it had important consequences. Individual colonies, no matter what their size, retained a degree of autonomy that translated immediately into the language and prerogatives of sovereignty. Under Massachusetts' influence, the Congress next adopted the Suffolk Resolves, recently voted in Suffolk county, Mass., which for the first time put natural rights into the official colonial argument (hitherto all remonstrances had been based on common law and constitutional rights). Apart from this, however, the prevailing mood was cautious.

The Congress' aim was to put such pressure on the British government that it would redress all colonial grievances and restore the harmony that had once prevailed. The Congress thus adopted an Association that committed the colonies to a carefully phased plan of economic pressure, beginning with nonimportation, moving to nonconsumption, and finishing the following September (after the rice harvest had been exported) with nonexportation. A few New England and Virginia delegates were looking toward independence; but the majority went home hoping that these steps, together with new appeals to the king and to the British people, would avert the need for any further such meetings. If these measures failed, however, a second Congress would convene the following spring.

Behind the unity achieved by the Congress lay deep divisions in colonial society. In the mid-1760s upriver New York was disrupted by land riots, which also broke out in parts of New Jersey; much worse disorder ravaged the backcountry of both North and South Carolina, where frontier people were left unprotected by legislatures that taxed them but in which they felt themselves unrepresented. A pitched battle at Alamance Creek in North Carolina in 1771 ended that rising and was followed by executions for treason. Although without such serious disorder, the cities also revealed acute social tensions and resentments of inequalities of economic opportunity and visible status. New York provincial politics were riven by intense rivalry between two great family-based parties, the DeLanceys, who benefited from royal government connections, and their rivals, the Livingstons. (The politics of the quarrel with Britain affected the domestic standing of the parties and eventually eclipsed the DeLanceys.) Another phenomenon was the rapid rise of dissenting religious sects, notably the Baptists; although they carried no political program, their style of preaching suggested a strong undercurrent of social as well as religious dissent. There was no inherent unity to these disturbances, but many leaders of colonial society were reluctant to ally themselves with these disruptive elements even in protest against Britain. They were concerned about the domestic consequences of letting the protests take a revolutionary turn; power shared with these elements might never be recovered.

Local disunity

When the Second Continental Congress met in Philadelphia in May 1775, most of the leaders still hoped for reconciliation with Britain; but news of clashes between British troops and Massachusetts militia at Lexington and

Concord (April 19) stirred the delegates to action. Steps were taken to put the continent on a war footing. While a further appeal was addressed to the British people (mainly on John Dickinson's insistence), the Congress raised an army, adopted a *Declaration of the Causes and Necessity of Taking Up Arms,* and appointed committees to deal with domestic supply and foreign affairs. In August 1775 the king declared a state of rebellion; by the end of the year, all colonial trade was banned. Even yet, General George Washington, commander of the Continental Army, still referred to the British troops as "ministerial" forces, indicating a civil war, not a war looking to separate national identity.

Then in January 1776 the publication of Thomas Paine's irreverent pamphlet *Common Sense* abruptly shattered this hopeful complacency and put independence on the agenda. Paine's eloquent, direct language spoke people's unspoken thoughts; no pamphlet had ever made such an impact on colonial opinion. While the Congress negotiated urgently, but secretly, for a French alliance, power struggles erupted in provinces where conservatives still hoped for relief. The only form relief could take, however, was British concessions; as public opinion hardened in Britain, where a general election in November 1774 had returned a strong majority for Lord North, the hope for reconciliation faded. In the face of British intransigence, men committed to their definition of colonial rights were left with no alternative; and the substantial portion of colonists—about one-fifth—who preferred loyalty to the crown, with all its disadvantages, were localized and outflanked. Where the British armies massed, they found plenty of loyalist support; but, when they moved on, they left the loyalists feeble and exposed.

The most dramatic internal revolution occurred in Pennsylvania, where a strong radical party, based mainly in Philadelphia but with allies in the country, seized power in the course of the controversy over independence itself. Opinion for independence swept the colonies in the spring of 1776. The Congress recommended that colonies form their own governments and assigned a committee to draft a declaration of independence.

The Declaration of Independence

This document, written by Thomas Jefferson but revised in committee, consisted of two parts. The preamble set the claims of the United States on a basis of natural rights, with a dedication to the principle of equality; the second was a long list of grievances against the crown—not Parliament now, since the argument was that Parliament had no lawful power in the colonies. On July 2 the Congress itself voted for independence; on the 4th it adopted the Declaration. (J.R.Po.)

THE WAR OF INDEPENDENCE

Until early in 1778 the War of Independence, also known as the American Revolution, was a civil war within the British Empire; later, as France in 1778, Spain in 1779, and the Netherlands in 1780 joined the colonies against Britain, the war became international. From the beginning sea power was vital in determining the course of the war, lending to British strategy a flexibility that helped compensate for the comparatively small numbers of troops sent to America and ultimately enabling the French to help bring about the final British surrender at Yorktown, Va.

Land campaigns. Americans fought the war on land essentially with two types of organization, the Continental (national) Army and the state militias. The total number of the former provided by quotas from the states throughout the conflict was 231,771 men; the militias totaled 164,087. At any given time, however, the American forces seldom numbered over 20,000; in 1781 there were only about 29,000 insurgents under arms throughout the country. The war was therefore one of small field armies. Militias, poorly disciplined and with elected officers, were summoned for periods usually not exceeding three months. The terms of Continental Army service were only gradually increased from one to three years, and not even bounties and the offer of land kept the army up to strength. Reasons for the difficulty in maintaining an adequate Continental force included the colonists' traditional antipathy to regular armies, the objections of farmers to being away from their fields, the competition of the states with the Continental Congress to keep men in the militia, and the wretched and uncertain pay in a period of inflation.

By contrast, the British army was a reliable, steady force of professionals. Since it numbered only about 42,000, heavy recruiting programs were introduced. Many of the enlisted men were farm boys, like most of the Americans. Others were unemployed persons from the urban slums. Still others joined the army to escape fines or imprisonment. The great majority became efficient soldiers owing to sound training and ferocious discipline. The officers were drawn largely from the gentry and the aristocracy and obtained their commissions and promotions by purchase. Though they received no formal training, they were not so dependent on a book knowledge of military tactics as were many of the Americans. British generals, however, tended toward a lack of imagination and initiative, while those who demonstrated such qualities often were rash.

Because troops were few and conscription unknown, the British government, following a traditional policy, purchased about 30,000 troops from various German princes. The Landgrave of Hesse furnished approximately three-fifths of this total. Few acts by the crown roused so much antagonism in America as this use of foreign mercenaries.

Opening engagements

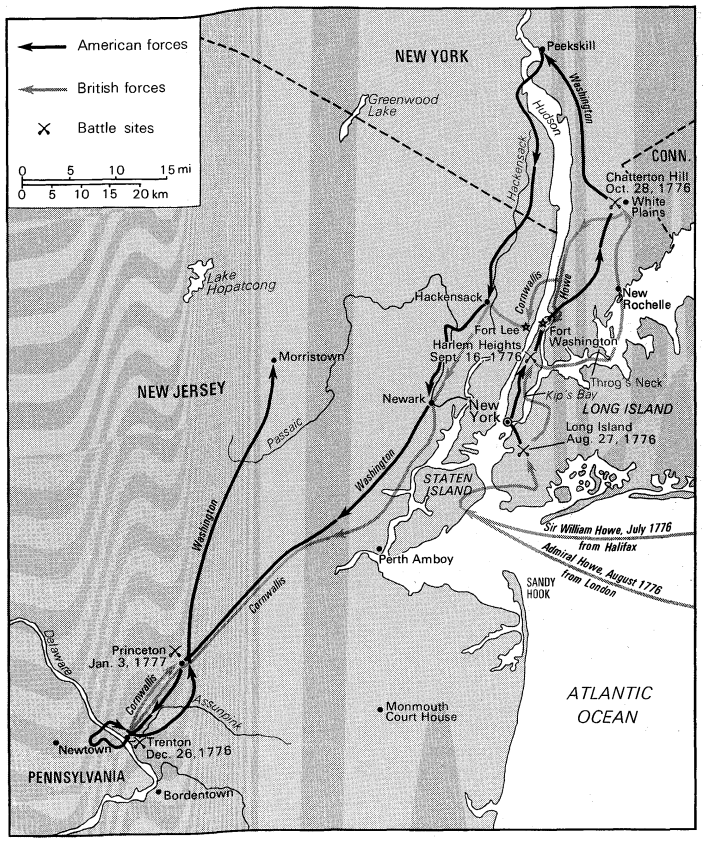
The war began in Massachusetts when General Thomas Gage sent a force from Boston to destroy rebel military stores at Concord. Fighting occurred at Lexington and Concord on April 19, 1775, and only the arrival of reinforcements saved the British original column. Rebel militia then converged on Boston from all over New England. Their entrenching on Breed's Hill led to a British frontal assault on June 17 under General William Howe, who won the hill but at the cost of more than 40 percent of the assault force.

General George Washington was appointed commander in chief of the American forces by the Continental Congress. Not only did he have to contain the British in Boston but he had also to recruit a Continental army. During the winter of 1775–76 recruitment lagged so badly that fresh drafts of militia were called up to help maintain the siege. The balance shifted in late winter, when General Henry Knox arrived with artillery from Fort Ticonderoga in New York, which had been captured from the British in May 1775. Mounted on Dorchester Heights, above Boston, the guns forced Howe, who had replaced Gage in command, to evacuate the city on March 17, 1776. Howe then repaired to Halifax to prepare for an invasion of New York, and Washington moved units southward for its defense.

Meanwhile, action flared in the north. In the fall of 1775 the Americans invaded Canada. One force under General Richard Montgomery captured Montreal on November 13. Another under Benedict Arnold made a remarkable march through the Maine wilderness to Quebec. Unable to take the city, Arnold was presently joined by Montgomery, many of whose troops had gone home because their enlistments had expired. An attack on the city on the last day of the year failed, Montgomery was killed, and many troops were captured. The Americans maintained a siege of the city but withdrew with the arrival of British reinforcements in the spring. Pursued by the British and decimated by smallpox, the Americans fell back to Ticonderoga. General Guy Carleton's hopes of moving quickly down Lake Champlain, however, were frustrated by Arnold's construction of a fighting fleet. Forced to build one of his own, Carleton destroyed most of the American fleet in October 1776 but considered the season too advanced to bring Ticonderoga under siege.

As the Americans suffered defeat in Canada, so did the British in the South. North Carolina patriots trounced a body of loyalists at Moore's Creek Bridge on Feb. 27, 1776. Charleston, S.C., was successfully defended against a British assault by sea in June.

Having made up its mind to crush the rebellion, the British government sent General Howe and his brother, Richard, Admiral Lord Howe, with a large fleet and 34,000 British and German troops to New York. It also gave the Howes a commission to treat with the Americans. The Continental Congress, which had proclaimed the in-



New York–New Jersey Campaign of 1776–77.

Adapted from *The American Heritage Pictorial Atlas of United States History*; copyright © 1966 by American Heritage Publishing Co., Inc.

dependence of the colonies, at first thought the Howes empowered to negotiate peace terms but discovered that they were authorized only to accept submission and assure pardons.

British seizure of New York

Their peace efforts getting nowhere, the Howes turned to force. Under his brother's guns, General Howe landed troops on Long Island and on August 27 scored a smashing victory. Washington evacuated his army from Brooklyn to Manhattan that night under cover of a fog. On September 15 Howe followed up his victory by invading Manhattan. Though checked at Harlem Heights the next day, he drew Washington off the island in October by a move to Throg's Neck and then to New Rochelle, northeast of the city. Leaving garrisons at Fort Washington on Manhattan and at Fort Lee on the opposite shore of the Hudson River, Washington hastened to block Howe. The latter, however, defeated him on October 28 at Chatterton Hill near White Plains. Howe slipped between the American army and Fort Washington and stormed the fort on November 16, seizing nearly 3,000 prisoners, guns, and supplies. Lord Cornwallis then took Fort Lee and on November 24 started to drive the American army across New Jersey. Though Washington escaped to the west bank of the Delaware, his army nearly disappeared. Howe then put his army into winter quarters, with outposts at such towns as Bordentown and Trenton.

On Christmas night Washington struck back with a brilliant riposte. Crossing the ice-strewn Delaware with 4,000 men, he fell upon the Hessian garrison at Trenton at dawn, taking nearly 1,000 prisoners. Though almost trapped by Cornwallis, who recovered Trenton on Jan. 2, 1777, Washington made a skillful escape during the night, won a battle against British reinforcements at Princeton the next day, and went into winter quarters in the defensible area around Morristown. The Trenton–Princeton campaign, which roused the country, saved the struggle for independence from collapse.

Britain's strategy in 1777 aimed at driving a wedge between New England and the other colonies. An army under General John Burgoyne was to march south from Canada and join forces with Howe on the Hudson. But Howe seems to have concluded that Burgoyne was strong enough to operate on his own and left New York in the summer, taking his army by sea to the head of Chesapeake Bay. Once ashore, he defeated Washington badly but not decisively at Brandywine Creek on September 11. Then, feinting westward, he entered Philadelphia, the American capital, on September 25. The Continental Congress fled to York. Washington struck back at Germantown on October 4 but, compelled to withdraw, went into winter quarters at Valley Forge.

In the north the story was different. Burgoyne was to move south to Albany with a force of about 9,000 British, Germans, Indians, and American loyalists; a smaller force under Lieutenant Colonel Barry St. Leger was to converge on Albany through the Mohawk valley. Burgoyne took Ticonderoga handily on July 5 and then, instead of using Lake George, chose a southward route by land. Slowed by the terrain and trees cut down by American axmen under General Philip Schuyler and needing horses, he sent a force of Germans to collect them at Bennington, Vt. The Germans were nearly wiped out on August 16 by New Englanders under General John Stark and Colonel Seth Warner. Meanwhile, St. Leger besieged Fort Schuyler (present Rome, N.Y.), ambushed a relief column of American militia at Oriskany on August 6, but retreated as his Indians gave up the siege and an American force approached under Arnold. Burgoyne himself reached the Hudson, but the Americans, now under General Horatio Gates, checked him at Freeman's Farm on September 19 and, thanks to Arnold's battlefield leadership, decisively defeated him at Bemis Heights on October 7. Ten days later, unable to get help from New York, Burgoyne surrendered at Saratoga.

The Battles of Saratoga

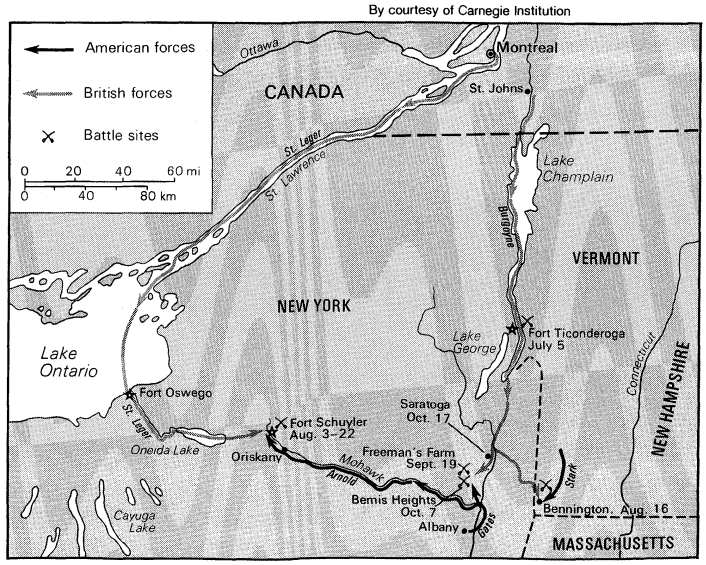
The most significant result of Burgoyne's capitulation was the entrance of France into the war. The French had secretly furnished financial and material aid since 1776. Now they prepared fleets and armies, although they did not formally declare war until June 1778.

Meanwhile, the Americans survived a hungry winter at Valley Forge, made worse by quartermaster and commissary mismanagement, grafting contractors, and the unwillingness of farmers to sell produce for paper money. The situation was improved by the arrival of Baron Friedrich Wilhelm von Steuben, a Prussian officer in the service of France. Von Steuben instituted a training program in which he emphasized drilling by officers, marching in column, and using firearms more effectively.

The program paid off at Monmouth Court House, N.J., on June 28, 1778, when Washington attacked the British withdrawing from Philadelphia to New York. Though Sir Henry Clinton, who had replaced Howe, struck back hard, the Americans stood their ground.

French aid now materialized with the appearance of a

By courtesy of Carnegie Institution



Northern Campaign of 1777.

strong fleet under the Count d'Estaing. Unable to enter New York harbour, d'Estaing tried to assist Major General John Sullivan in dislodging the British from Newport, R.I. Storms and British reinforcements thwarted the joint effort.

Action in the north was largely a stalemate for the rest of the war. The British raided New Bedford, Mass., and New Haven and New London, Conn., while loyalists and Indians attacked settlements in New York and Pennsylvania. On the other hand, the Americans under Anthony Wayne stormed Stony Point, N.Y., on July 16, 1779, and "Light-Horse" Harry Lee took Paulus Hook, N.J., on August 19. More lasting in effect was Sullivan's expedition of August 1779 against Britain's Indian allies in New York, particularly the destruction of their villages and fields of corn. Farther west, Colonel George Rogers Clark seized Vincennes and other posts north of the Ohio in 1778.

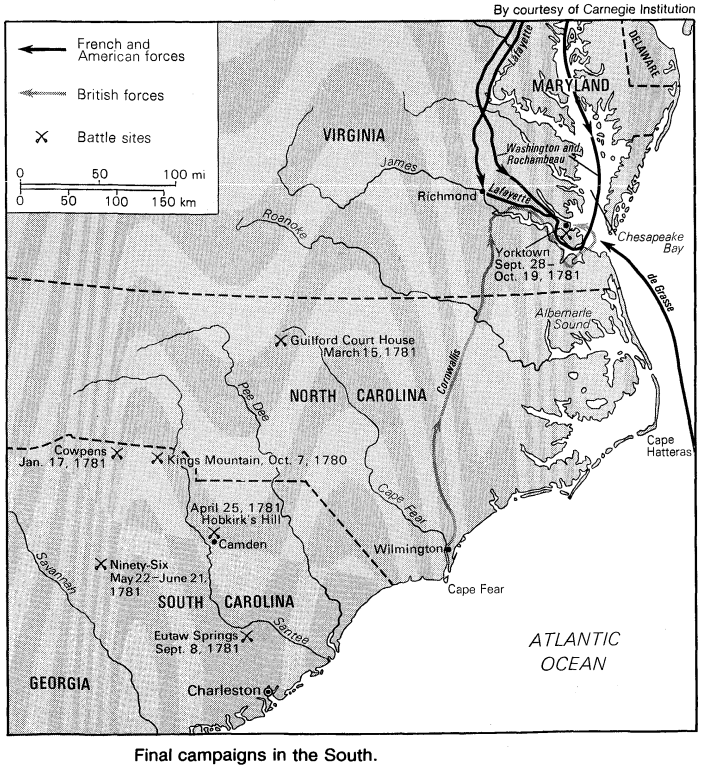
Potentially serious blows to the American cause were Benedict Arnold's defection in 1780 and army mutinies in 1780 and 1781. Arnold's attempt to betray West Point to the British miscarried. Mutinies were sparked by misunderstandings over terms of enlistment, poor food and clothing, gross arrears of pay, and the decline in the purchasing power of the dollar. Suppressed by force or negotiation, the mutinies shook the morale of the army.

The Americans also suffered setbacks in the South. British strategy called for offensives there from 1778 on that were designed to take advantage of the flexibility of sea power and the loyalist sentiment of many of the people. British forces from New York and St. Augustine occupied Georgia by the end of January 1779 and successfully defended Savannah in the fall against d'Estaing and a Franco-American army. Clinton, having withdrawn his Newport garrison, captured Charleston and an American army of 5,000 under General Benjamin Lincoln in May 1780. Learning that Newport was threatened by a French expeditionary force under the Count de Rochambeau, he returned to New York, leaving Cornwallis at Charleston.

Cornwallis, however, took the offensive. On August 16 he shattered General Gates's army at Camden, S.C. The destruction of a force of loyalists at Kings Mountain on October 7 led him to move against the new American commander, Nathanael Greene. When Greene put part of his force under General Daniel Morgan, Cornwallis sent his cavalry leader, Banastre Tarleton, after Morgan. At Cowpens on Jan. 17, 1781, Morgan destroyed practically all of Tarleton's column. Subsequently, on March 15, Greene and Cornwallis fought at Guilford Court House, N.C. Cornwallis won but suffered heavy casualties. After withdrawing to Wilmington, he marched into Virginia to join British forces sent there by Clinton.

Greene then moved to South Carolina, where he was defeated by Lord Rawdon at Hobkirk's Hill on April 25 and at Ninety-Six in June and by Lieutenant Colonel Alexander Stewart at Eutaw Springs on September 8. In spite of this, the British, harassed by partisan leaders like Francis Marion, Thomas Sumter, and Andrew Pickens, soon retired to the coast and remained locked up in Charleston and Savannah.

By courtesy of Carnegie Institution



Final campaigns in the South.

The final campaign

Meanwhile Cornwallis entered Virginia. Sending Tarleton on raids across the state, he started to build a base at Yorktown, at the same time fending off American forces under Wayne, von Steuben, and the Marquis de Lafayette.

Learning that the Count de Grasse had arrived in the Chesapeake with a large fleet and 3,000 French troops, Washington and Rochambeau moved south to Virginia. By mid-September the Franco-American forces had placed Yorktown under siege, and British rescue efforts proved fruitless. Cornwallis surrendered his army of more than 7,000 men on October 19. Thus, for the second time during the war the British had lost an entire army.

Thereafter, land action in America died out, though the war persisted in other theatres and on the high seas. Eventually Clinton was replaced by Sir Guy Carleton. While the peace treaties were under consideration and afterward, Carleton evacuated thousands of loyalists from America, including many from Savannah on July 11, 1782, and others from Charleston on December 14. The last British forces finally left New York on Nov. 25, 1783. Washington then reentered the city in triumph.

Historians' judgments

In explaining the outcome of the war, scholars have pointed out that the British never contrived an overall general strategy for winning it. Also, even if the war could have been terminated by British power in the early stages, the generals during that period, notably Howe, declined to make a prompt, vigorous, intelligent application of that power. They acted, to be sure, within the conventions of their age, but in choosing to take minimal risks (for example, Carleton at Ticonderoga and Howe at Brooklyn Heights and later in New Jersey and Pennsylvania) they lost the opportunity to deal potentially mortal blows to the rebellion. There was also a grave lack of understanding and cooperation at crucial moments (as with Burgoyne and Howe in 1777). Finally, the British counted too strongly on loyalist support they did not receive.

But British mistakes alone could not account for the success of the United States. Feeble as their war effort occasionally became, the Americans were able generally to take advantage of their enemies' mistakes. The Continental Army, moreover, was by no means an inept force even before von Steuben's reforms. The militias, while usually unreliable, could perform admirably under the leadership of men who understood them, like Arnold, Greene, and Morgan, and often reinforced the Continentals in crises. Furthermore, Washington, a rock in adversity, learned slowly but reasonably well the art of generalship. The supplies and funds furnished by France from 1776 to 1778 were invaluable, while French military and naval support after 1778 was essential. The outcome, therefore, resulted from a combination of British blunders, American efforts, and French assistance.

The war at sea. Although the colonists ventured to challenge Britain's naval power from the outbreak of the conflict, the war at sea in its later stages was mainly fought between Britain and America's European allies, the American effort being reduced to privateering.

The Continental Navy

The importance of the sea was recognized early. In October 1775 the Continental Congress authorized the creation of the Continental Navy and established the Marine Corps in November. Taking its direction from the naval and marine committees of the Congress, the navy was only occasionally effective. In 1776 it had 27 ships as against Britain's 270; by the end of the war, the British total had risen close to 500, and the American had dwindled to 20. Many of the best seamen available went off privateering,

and both Continental Navy commanders and crews suffered from a lack of training and discipline.

The first significant blow by the navy was struck by Commodore Esek Hopkins, who captured New Providence (Nassau) in the Bahamas in 1776.

Other captains, such as Lambert Wickes, Gustavus Conyngham, and John Barry, also enjoyed successes, but the Scottish-born John Paul Jones was especially notable. As captain of the *Ranger,* Jones scourged the British coasts in 1778, capturing the man-of-war *Drake.* As captain of the *Bonhomme Richard* in 1779, he intercepted a timber convoy and captured the British frigate *Serapis.*

More injurious to the British were the raids by American privateers on their shipping. American ships, furnished with letters of marque by the Congress or the states, swarmed about the British Isles. By the end of 1777 they had taken 560 British vessels, and by the end of the war they had probably seized 1,500. More than 12,000 British sailors also were captured. One result was that, by 1781, British merchants were clamouring for an end to hostilities.

Most of the naval action occurred at sea. The significant exception was Benedict Arnold's battles against General Carleton's fleet on Lake Champlain at Valcour Island on Oct. 11 and off Split Rock on Oct. 13, 1776. Arnold lost both battles, but his construction of his fleet of tiny vessels, mostly gondolas (gundalows) and galleys, had forced the British to build a larger fleet and hence delayed their attack on Fort Ticonderoga until the following spring. This delay contributed significantly to Burgoyne's capitulation in October 1777.

The entrance of France into the war, followed by Spain in 1779 and the Netherlands in 1780, was an important factor in the naval aspect of the war. The Spanish and Dutch were not particularly active, but their role in keeping British naval forces tied down in Europe was significant. The British navy could not maintain an effective blockade of both the American coast and the enemies' ports. Owing to years of economy and neglect, Britain's ships of the line were neither modern nor sufficiently numerous. An immediate result was that France's Toulon fleet under d'Estaing got safely away to America, where it appeared off New York and later assisted General Sullivan in the unsuccessful siege of Newport. A fierce battle off Ushant, Fr., in July 1778 between the Channel fleet under Admiral Augustus Keppel and the Brest fleet under the Count d'Orvilliers proved inconclusive. Had Keppel won decisively, French aid to the Americans would have diminished and Rochambeau might never have been able to lead his expedition to America.

In the following year England was in real danger. Not only did it have to face the privateers of the United States, France, and Spain off its coasts, as well as the raids of John Paul Jones, but it also lived in fear of invasion. The combined fleets of France and Spain had acquired command of the Channel, and a French army of 50,000 waited for the propitious moment to board their transports. Luckily for the British, storm, sickness among the Allied crews, and changes of plans terminated the threat.

The planned invasion of England

Despite Allied supremacy in the Channel in 1779, the threat of invasion, and the loss of islands in the West Indies, the British maintained control of the North American seaboard for most of 1779 and 1780, which made possible their Southern land campaigns. They also reinforced Gibraltar, which the Spaniards had brought under siege in the fall of 1779, and sent a fleet under Admiral Sir George Rodney to the West Indies in early 1780. After fruitless maneuvering against the Count de Guichen, who had replaced d'Estaing, Rodney sailed for New York.

While Rodney had been in the West Indies, a French squadron slipped out of Brest and sailed to Newport with Rochambeau's army. Rodney, instead of trying to block the approach to Newport, returned to the West Indies, where, upon receiving instructions to attack Dutch possessions, he seized Sint Eustatius, the Dutch island that served as the principal depot for war materials shipped from Europe and transshipped into American vessels. He became so involved in the disposal of the enormous booty that he dallied at the island for six months.

In the meantime, a powerful British fleet relieved Gibraltar in 1781, but the price was the departure of the French fleet at Brest, part of it to India, the larger part under Admiral de Grasse to the West Indies. After maneuvering indecisively against Rodney, de Grasse received a request from Washington and Rochambeau to come to New York or the Chesapeake.

Earlier, in March, a French squadron had tried to bring troops from Newport to the Chesapeake but was forced to return by Admiral Marriot Arbuthnot, who had succeeded Lord Howe. Soon afterward Arbuthnot was replaced by Thomas Graves, a conventional-minded admiral.

Informed that a French squadron would shortly leave the West Indies, Rodney sent Samuel Hood north with a powerful force while he sailed for England, taking with him several formidable ships that might better have been left with Hood.

Soon after Hood dropped anchor in New York, de Grasse appeared in the Chesapeake, where he landed troops to help Lafayette contain Cornwallis until Washington and Rochambeau could arrive. Fearful that the Count de Barras, who was carrying Rochambeau's artillery train from Newport, might join de Grasse, and hoping to intercept him, Graves sailed with Hood to the Chesapeake. Graves had 19 ships of the line against de Grasse's 24. Though the battle that began on September 5 off the Virginia capes was not a skillfully managed affair, Graves had the worst of it and retired to New York. He ventured out again on October 17 with a strong contingent of troops and 25 ships of the line, while de Grasse, reinforced by Barras, now had 36 ships of the line. No battle occurred, however, since off the capes Graves learned that Cornwallis had surrendered.

The Battle of the Virginia Capes

Though Britain subsequently recouped some of its fortunes, Rodney defeating and capturing de Grasse in the Battle of the Saints off Dominica in 1782 and British land and sea forces inflicting defeats in India, the turn of events did not significantly alter the situation in America as it existed after Yorktown. A new government under Lord Shelburne (1st Marquess of Lansdowne) tried to get the American commissioners to agree to a separate peace, but, ultimately, the treaty negotiated with the Americans was not to go into effect until the formal conclusion of a peace with their European allies. (W.M.Wa.)

Treaty of Paris. The military verdict in North America was reflected in the preliminary Anglo-American peace treaty of 1782, which was included in the Treaty of Paris of 1783. Benjamin Franklin, John Adams, John Jay, and Henry Laurens served as the American commissioners. By its terms Britain recognized the independence of the United States with generous boundaries, including the Mississippi River on the west. Britain retained Canada but ceded East and West Florida to Spain. Provisions were inserted calling for the payment of American private debts to British citizens, for American access to the Newfoundland fisheries, and for a recommendation by the Continental Congress to the states in favour of fair treatment of the loyalists.

Most of the loyalists remained in the new nation. Perhaps as many as 80,000 Tories migrated to Canada, England, and the British West Indies. Many of these had served as British soldiers, and many had been banished by the American states. The less ardent and more cautious Tories, staying in the United States, accepted the separation from Britain as final and could not be distinguished from the patriots after the passage of a generation. The loyalists were harshly treated as dangerous enemies by the American states during the war and immediately afterward. They were commonly deprived of civil rights, often fined, and frequently deprived of their property. The more conspicuous were usually banished upon pain of death. The British government compensated more than 4,000 of the exiles for property losses, paying out almost £3,300,000. It also gave them land grants, pensions, and appointments to enable them to reestablish themselves. (Ed.)

FOUNDATIONS OF THE AMERICAN REPUBLIC

It had been far from certain that the Americans could fight a successful war against the might of Britain. The scattered colonies had little inherent unity; their experience of col-

lective action was limited; an army had to be created and maintained; they had no common institutions other than the Continental Congress; and they had almost no experience of continental public finance. The Americans could not have hoped to win the war without French help, and the French monarchy—whose interests were anti-British but not pro-American—had waited watchfully to see what the Americans could do in the field. Although the French began supplying arms, clothing, and loans surreptitiously soon after the Americans declared independence, it was not until 1778 that they entered into a formal alliance.

Most of these problems lasted beyond the achievement of independence and continued to vex American politics for many years, even for generations. Meanwhile, however, the colonies had valuable, though less visible, sources of strength. Practically all farmers had their own arms and could form into militia companies overnight. More fundamentally, Americans had for many years been receiving basically the same information, mainly from the English press, reprinted in identical form in colonial newspapers. The effect of this was to form a singularly wide body of agreed opinion about major public issues. Another force of incalculable importance was the fact that for several generations Americans had to a large extent been governing themselves through elected assemblies, which in turn had developed sophisticated experience in committee politics.

Tradition of self-government

This factor of "institutional memory" was of great importance in the forming of a mentality of self-government. Men became attached to their habitual ways, especially when these were habitual ways of running their own affairs; and these habits formed the basis of an ideology just as pervasive and important to the people concerned as republican theories published in Britain and the European continent. Moreover, colonial self-government seemed, from a colonial point of view, to be continuous and consistent with the principles of English government—principles for which Parliament had fought the Civil Wars in the mid-17th century and which colonists believed to have been reestablished by the Glorious Revolution of 1688–89. It was equally important that experience of self-government had taught colonial leaders how to get things done. When the Continental Congress met in 1774, members did not have to debate procedure (except on voting); they already knew it. Finally, the Congress' authority was rooted in traditions of legitimacy. The old election laws were used. Voters could transfer their allegiance with minimal difficulty from the dying colonial assemblies to the new assemblies and conventions of the states.

Problems before the Second Continental Congress. When the Second Continental Congress assembled in Philadelphia in May 1775, revolution was not a certainty. The Congress had to prepare for that contingency nevertheless and thus was confronted by two parallel sets of problems. The first was how to organize for war; the second, which proved less urgent but could not be set aside forever, was how to define the legal relationship between the Congress and the states.

In June 1775 the Congress provided for the enlistment of an army and appointed Colonel George Washington (who had made a point of turning up in uniform) commander in chief. It then turned to the vexatious problems of finance. An aversion to taxation being one of the unities of American sentiment, the Congress began by trying to raise a domestic loan. It did not have much success, however, for the excellent reason that the outcome of the operation appeared highly dubious. At the same time, authority was taken for issuing a paper currency. This proved to be the most important method of domestic war finance, and, as the war years passed, Congress resorted to issuing more and more Continental currency, which depreciated rapidly and had to compete with currencies issued by state governments. (People were inclined to prefer local currencies.) The Continental Army was a further source of a form of currency because its commission agents issued certificates in exchange for goods; these certificates bore an official promise of redemption and could be used in personal transactions. Loans raised overseas, notably in France and the Netherlands, were another important source of revenue.

In 1780 Congress decided to call in all former issues of currency and replace them with a new issue on a 40-to-1 ratio. The Philadelphia merchant Robert Morris, who was appointed superintendent of finance in 1781 and came to be known as the Financier, guided the United States through its complex fiscal difficulties. Morris' personal finances were inextricably tangled up with those of the country, and he became the object of much hostile comment, but he also used his own resources to secure urgently needed loans from abroad. In 1781 Morris secured a charter for the first Bank of North America, an institution which owed much to the example of the Bank of England. Although the bank was attacked by radical egalitarians as an unrepublican manifestation of privilege, it gave the United States a firmer financial foundation.

The problem of financing and organizing the war sometimes overlapped with Congress' other major problem, that of defining its relations with the states. The Congress, being only an association of states, had no power to tax individuals. The Articles of Confederation, a plan of government organization adopted and put into practice by Congress in 1777, although not officially ratified by all the states until 1781, gave Congress the right to make requisitions on the states proportionate to their ability to pay. The states in turn had to raise these sums by their own domestic powers to tax, a method that state legislators looking for reelection were reluctant to employ. The result was that many states were constantly in heavy arrears, and, particularly after the urgency of the war years had subsided, the Congress' ability to meet expenses and repay its war debts was crippled.

The Articles of Confederation

The Congress lacked power to enforce its requisitions and fell badly behind in repaying its wartime creditors. When individual states (Maryland as early as 1782, Pennsylvania in 1785) passed legislation providing for repayment of the debt owed to their own citizens by the Continental Congress, one of the reasons for the Congress' existence had begun to crumble. Two attempts were made to get the states to agree to grant the Congress the power it needed to raise revenue by levying an impost on imports. Each failed for want of unanimous consent. Essentially, an impost would have been collected at ports, which belonged to individual states—there was no "national" territory—and therefore cut across the concept of state sovereignty. Agreement was nearly obtained on each occasion, and, if it had been, the Constitutional Convention might never have been called. But the failure sharply pointed up the weakness of the Congress and of the union between the states under the Articles of Confederation.

The Articles of Confederation reflected strong preconceptions of state sovereignty. Article II expressly reserved sovereignty to the states individually, and another article even envisaged the possibility that one state might go to war without the others. Fundamental revisions could be made only with unanimous consent, because the Articles represented a treaty between sovereigns, not the creation of a new nation-state. Other major revisions required the consent of nine states. Yet state sovereignty principles rested on artificial foundations. The states could never have achieved independence on their own, and in fact the Congress had taken the first step both in recommending that the states form their own governments and in declaring their collective independence. Most important of its domestic responsibilities, by 1787 the Congress had enacted several ordinances establishing procedures for incorporating new territories. (It had been conflicts over western land claims that had held up ratification of the Articles. Eventually the states with western claims, principally New York and Virginia, ceded them to the United States.) The Northwest Ordinance of 1787 provided for the phased settlement and government of territories in the Ohio valley, leading to eventual admission as new states. It also excluded the introduction of slavery—though it did not exclude the retention of existing slaves.

The Northwest Ordinance

The states had constantly looked to the Congress for leadership in the difficulties of war; now that the danger was past, however, disunity began to threaten to turn into disintegration. The Congress was largely discredited in the eyes of a wide range of influential men, representing both

old and new interests. The states were setting up their own tariff barriers against each other and quarreling among themselves; virtual war had broken out between competing settlers from Pennsylvania and Connecticut claiming the same lands. By 1786, well-informed men were discussing a probable breakup of the Confederation into three or more new groups, which could have led to wars among the American republics.

State politics. The problems of forming a new government affected the states individually as well as in confederation. Most of them established their own constitutions—formulated either in conventions or in the existing assemblies. The most democratic of these constitutions was the product of a virtual revolution in Pennsylvania, where a highly organized radical party seized the opportunity of the revolutionary crisis to gain power. Suffrage was put on a taxpayer basis, with nearly all adult males paying some tax; representation was reformed to bring in the populations of western counties; and a single-chamber legislature was established. An oath of loyalty to the constitution for some time excluded political opponents and particularly Quakers (who could not take oaths) from participation. For the rest, the state constitutions reflected the firm political ascendancy of the traditional ruling elite. Power ascended from a broad base in the elective franchise and representation through a narrowing hierarchy of offices restricted by property qualifications. State governors had in some cases to be men of great wealth. Senators were either wealthy or elected by the wealthy sector of the electorate. (These conditions were not invariable; Virginia, which had a powerful landed elite, dispensed with such restrictions.) Several states retained religious qualifications for office; the separation of church and state was not a popular concept, and minorities such as Baptists and Quakers were subjected to indignities that amounted in some places (notably Massachusetts and Connecticut) to forms of persecution.

Elite power provided a lever for one of the most significant transformations of the era, one that took place almost without being either noticed or intended. This was the acceptance of the principle of proportional representation as the determining rule of political action. It was made not only possible but attractive when the larger aggregations of population broadly coincided with the highest concentrations of property: great merchants and landowners from populous areas could continue to exert political ascendancy so long as they retained some sort of hold on the political process. (This would hardly have been possible if American politics had been ruled by class war, but this was not the case.) The principle reemerged to dominate the distribution of voters in the House of Representatives and in the electoral college under the new federal Constitution.

The principle of proportional representation

Relatively conservative constitutions did little to stem a tide of increasingly democratic politics. The old elites had to wrestle with new political forces (and in the process they learned how to organize in the new regime). Executive power was weakened. Many elections were held annually, and terms were limited. Legislatures quickly admitted new representatives from recent settlements, many with little previous political experience.

The new state governments, moreover, had to tackle major issues that affected all classes. The needs of public finance led to emissions of paper money. In several states these were resumed after the war, and, since they tended (though not invariably) to depreciate, they led directly to fierce controversies. The treatment of loyalists—adherents of the British cause—was also a theme of intense political dispute after the war. Despite the protests of men like Alexander Hamilton, who urged restoration of property and rights, in many states loyalists were driven out and their estates seized and redistributed in forms of auction, providing opportunities for speculation rather than personal occupation. Many states were depressed economically. In Massachusetts, which remained under orthodox control, stiff taxation under conditions of postwar depression trapped many farmers into debt. Unable to meet their obligations, they rose late in 1786 under a Revolutionary War officer, Captain Daniel Shays, in a movement to prevent the court sessions. Shays's Rebellion was crushed early in 1787 by an army raised in the state. The action caused only a few casualties, but the episode sent a shiver of fear throughout the country's propertied classes. It also seemed to justify the classical thesis that republics were unstable. It thus provided a potent stimulus to state legislatures to send delegates to the convention called (following a preliminary meeting in Annapolis) to meet at Philadelphia to revise the Articles of Confederation.

The Constitutional Convention. The Philadelphia Convention, which met in May 1787, was officially called for by the old Congress solely to remedy defects in the Articles of Confederation. But the "Virginia Plan" presented by the Virginia delegates went beyond revision and boldly proposed to introduce a new, national government in place of the existing confederation. The Convention thus immediately faced the question of whether the United States was to be a nation. This decision was effectively made when the plan for a bicameral legislature was approved. The alternative, based on the old single chamber representing autonomous states, was passed over when the hastily drafted New Jersey Plan was defeated in mid-June.

The Virginia Plan

The Constitution as it emerged after a summer of debate embodied a much stronger principle of separation of powers than was generally to be found in the state constitutions. The chief executive was to be a single figure (a composite executive was discussed) and was to be elected by an electoral college, meeting in the states. This followed much debate over the Virginia Plan's preference for legislative election. The principal control on the president was the rather remote threat of impeachment (to which James Madison attached great importance). The Virginia Plan's proposal that representation be proportional to population in both houses was severely modified by the retention of equal representation for each state in the Senate. After some contention, antislavery forces gave way to a compromise by which three-fifths of the slaves would be counted as population for purposes of representation (and direct taxation). Slave states would thus be perpetually overrepresented in national politics; provision was also added for a law permitting the recapture of fugitive slaves, though in deference to republican scruples the word "slaves" was not used.

Contemporary theory expected the legislature to be the most powerful branch of government. Thus, to balance the system, the executive was given a veto, and a judicial system with powers of review was established. It was also implicit in the structure that the new, federal judiciary would have power to veto any state laws that conflicted with either the Constitution or with federal statutes. States were forbidden to pass laws impairing obligations of contract—a measure aimed to encourage capital—and the Congress could pass no ex post facto law. But the Congress was endowed with the basic powers of a modern—and sovereign—government. This was a republic, and the United States could confer no aristocratic titles of honour. The prospect of eventual enlargement of federal power appeared in the clause giving the Congress powers to pass legislation "necessary and proper" for implementing the general purposes of the Constitution.

The states retained their civil jurisdiction; but there was an emphatic shift of the political centre of gravity to the federal government, of which the most fundamental indication was the universal understanding that this government would act directly on citizens, as individuals, throughout all the states, regardless of state authority. The language of the Constitution told of the new style: it began, "We the people of the United States," rather than "We the people of New Hampshire, Massachusetts, etc."

The draft Constitution aroused widespread opposition. Anti-Federalists—so-called because their opponents deftly seized the appellation of "Federalists," though they were really nationalists—were strong in states such as Virginia, New York, and Massachusetts, where the economy was relatively successful and many people saw little need for such extreme remedies. Anti-Federalists also expressed fears—here touches of class conflict certainly arose—that the new government would fall into the hands of merchants and men of money. Many good republicans detected oligarchy

Anti-Federalist opposition

in the structure of the Senate, with its six-year terms. The absence of a bill of rights aroused deep fears of central power. The Federalists, however, had the advantages of communications, the press, organization, and, generally, the better of the argument. Anti-Federalists also suffered the disadvantage of having no internal coherence or unified purpose.

The debate gave rise to a very intensive literature, much of it at a higher intellectual level than can be found in today's public debates. The most sustained pro-Federalist argument, written mainly by Alexander Hamilton and James Madison (assisted by John Jay) under the pseudonym of Publius, appeared in the newspapers as *The Federalist*. These essays attacked the feebleness of the Confederation and claimed that the new Constitution would have advantages for all sectors of society while threatening none. In the course of the debate, they passed from a strongly nationalist standpoint to one that showed more respect for the idea of a mixed form of government that would safeguard the states. Madison contributed assurances that a multiplicity of interests would counteract each other, preventing the consolidation of power continually charged by their enemies.

The Bill of Rights

The Bill of Rights, steered through the first Congress by Madison's diplomacy, mollified much of the latent opposition. These first 10 amendments, ratified in 1791, adopted into the Constitution the basic English common-law rights that Americans had fought for. But they did more. Unlike Britain, the United States secured a guarantee of freedom for the press and the right of (peaceable) assembly. Also unlike Britain, church and state were formally separated in a clause that seemed to set equal value on nonestablishment of religion and its free exercise. (This left the states free to maintain their own establishments.)

In state conventions held through the winter of 1787 to the summer of 1788, the Constitution was ratified by the necessary minimum of nine states. But the vote was desperately close in Virginia and New York, respectively the 10th and 11th states to ratify, and without them the whole scheme would have been built on sand.

THE SOCIAL REVOLUTION

The American Revolution was a great social upheaval but one that was widely diffused, often gradual, and different in different regions. The principles of liberty and equality stood in stark conflict with the institution of African slavery, which had built much of the country's wealth. One gradual effect of this conflict was the decline of slavery in all the Northern states; another was a spate of manumissions by liberal slave owners in Virginia. But with most slave owners, especially in South Carolina and Georgia, ideals counted for nothing. Throughout the slave states, the institution of slavery came to be reinforced by a doctrine of racial inferiority, which proved hard to dispel. Although the manumissions did result in the development of new communities of free blacks, who enjoyed considerable freedom of movement for a few years, in the 1790s the condition of free blacks deteriorated as states adopted laws restricting their activities, residences, and economic choices. They came to occupy poor neighbourhoods and grew into a permanent underclass, denied education or opportunity.

The role of women

The War of Independence also dramatized the economic importance of women. Women had always contributed indispensably to the operation of farms and often businesses, while seldom acquiring independent status; but, when war removed men from the locality, women often had to take full charge, which they proved they could do. Republican ideas spread among women, influencing discussion of women's rights, education, and role in society. Some states modified their inheritance and property laws to permit women to inherit a share of estates and to exercise limited control of property after marriage. On the whole, however, the Revolution itself had only very gradual and diffused effects on women's ultimate status. Such changes as took place amounted to a fuller recognition of the importance of women as mothers of republican citizens rather than making them into independent citizens of equal political and civil status with men.

The American Revolution was in many respects a manifestation of the Enlightenment in political, civil, and ecclesiastical action. One of its triumphs was the passage of the Virginia Statute for Religious Liberty in 1786 (which Jefferson, the original author, proudly had printed in the next edition of the French *Encyclopédie*). The state would tolerate all religions but give formal favour to none; people were free to follow the dictates of their own religious consciences. Although several states retained formal establishments, there was much competition among sects. In New England and in commercial centres of activity, and later in newer western settlements, the earlier severe Calvinism gradually gave way to a gentler and more indulgent universalism: people came to hope and then to believe that God actually wanted his creatures to be happy. Doctrinally, moreover, Unitarianism appealed to an increasing number of Congregationalists. A great new revivalist movement arose again around 1798, mainly in the new West, and this frequently renewed revival spirit appealed directly to the senses and away from the moderate intellectualism of the Enlightenment.

Americans had fought for independence to protect common-law rights; they had no program for legal reform. Gradually, however, some customary practices came to seem out of keeping with republican principles. The outstanding example was the law of inheritance. The new states took steps, where necessary, to remove the old rule of primogeniture in favour of equal partition of intestate estates; this conformed both to the egalitarian and the individualist principles preferred by American society. Humanization of the penal codes, however, occurred only gradually, in the 19th century, inspired as much by European example as by American sentiment.

Postwar Indian relations

As for the problem of the indigenous population, Americans had no clear or consistent solution. Indians were not taxed; they were not citizens; and yet they often lived, traded, and earned a living in and around Euro-American centres and settlements. In the west, Indian and Euro-American cultures interacted and constantly learned from one another, but their essentially incompatible aims often broke into hostility. The new government of the United States thus found itself involved at once in a war on its northwestern frontiers with a formidable enemy. A temporary peace was achieved after Anthony Wayne's victory in 1794 at the Battle of Fallen Timbers. The following year 12 Indian tribes signed the Treaty of Greenville, opening the northwest for U.S. settlement.

THE UNITED STATES FROM 1789 TO 1816

The Federalist administration and the formation of parties. The first elections under the new Constitution were held in 1789. Washington was unanimously voted the nation's first president. His secretary of the treasury, Alexander Hamilton, formed a clear-cut program that soon gave substance to the old fears of the Anti-Federalists. Hamilton, who had believed since the early 1780s that a national debt would be "a national blessing," both for economic reasons and because it would act as a "cement" to the Union, used his new power base to realize the ambitions of the nationalists. He recommended that the federal government pay off the old Continental Congress' debts at par rather than at a depreciated value and that it assume state debts, drawing the interests of the creditors toward the central rather than state governments. This plan met strong opposition from the many who had sold their securities at great discount during the postwar depression and from Southern states, which had repudiated their debts and did not want to be taxed to pay other states' debts. A compromise in Congress was reached—thanks to the efforts of Secretary of State Jefferson—whereby Southern states approved Hamilton's plan in return for Northern agreement to fix the location of the new national capital on the banks of the Potomac, closer to the South. When Hamilton next introduced his plan to found a Bank of the United States, modeled on the Bank of England, opposition began to harden. Many argued that the Constitution did not confide this power to Congress. Hamilton, however, persuaded Washington that anything not expressly forbidden by the Constitution

was permitted under implied powers—the beginning of "loose" as opposed to "strict" constructionist interpretations of the Constitution. The Bank Act passed in 1791. Hamilton also advocated plans for the support of nascent industry, which proved premature, and he imposed a revenue-raising whiskey excise that led to a minor rebellion in western Pennsylvania in 1794.

A party opposed to Hamilton's fiscal policies began to form in Congress. With Madison at its centre and with support from Jefferson, it soon extended its appeal beyond Congress to the constituencies. Meanwhile, the French Revolution and France's subsequent declaration of war against Great Britain, Spain, and Holland further divided American loyalties. Democratic-Republican societies sprang up to express support for France, while Hamilton and his supporters, known as Federalists, backed Britain for economic reasons. Washington pronounced American neutrality in Europe, but to prevent a war with Britain he sent Chief Justice John Jay to London to negotiate a treaty (1794). The United States gained only minor concessions and—humiliatingly—accepted British naval supremacy as the price of protection for American shipping.

The Jay Treaty

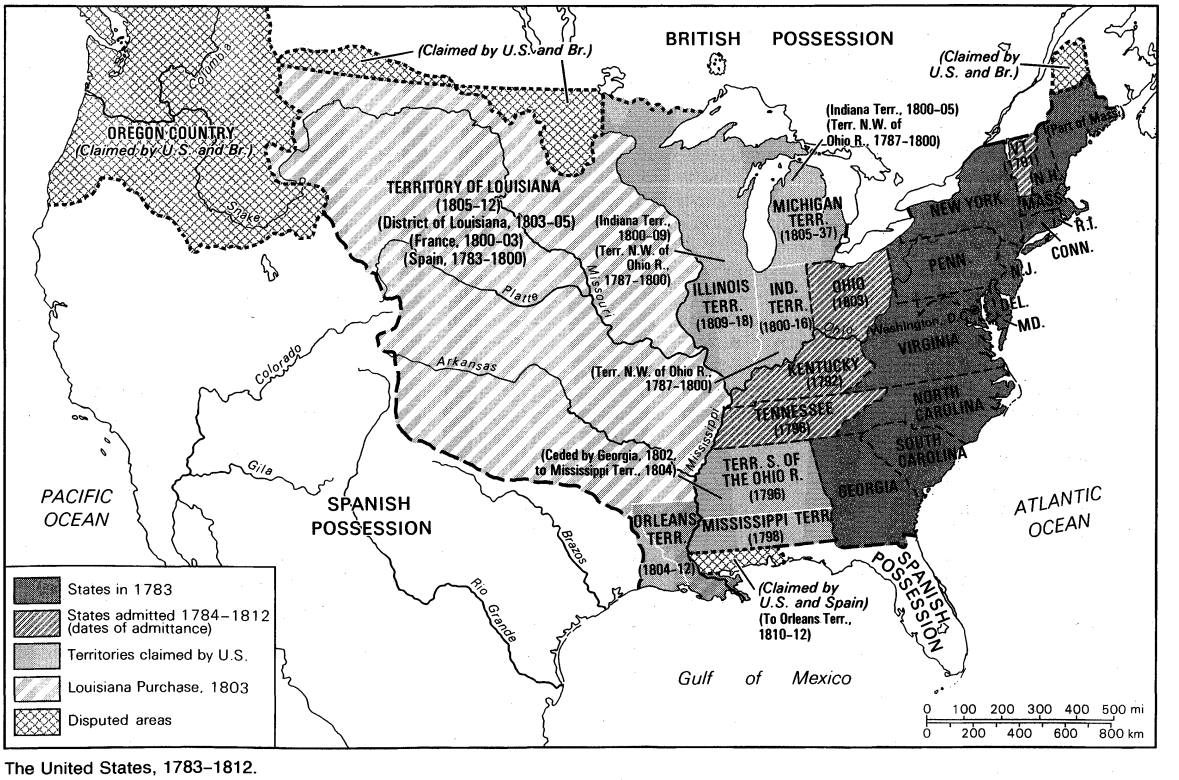
Washington, whose tolerance had been severely strained by the Whiskey Rebellion and by criticism of the Jay Treaty, chose not to run for a third presidential term. In his farewell address (in a passage drafted by Hamilton) he denounced the new party politics as divisive and dangerous. Parties did not yet aspire to national objectives, however, and, when the Federalist John Adams was elected president, the Republican Jefferson, as the presidential candidate with the second greatest number of votes, became vice president. Wars in Europe and on the high seas, together with rampant opposition at home, gave the new administration little peace. Virtual naval war with France had followed from American acceptance of British naval protection. In 1798 a French attempt to solicit bribes from American commissioners negotiating a settlement of differences (the so-called XYZ Affair) aroused a wave of anti-French feeling. Later that year the Federalist majority in Congress passed the Alien and Sedition Acts, which imposed serious civil restrictions on aliens suspected of pro-French activities and penalized U.S. citizens who criticized the government, making nonsense of the First Amendment's guarantee of free press. The acts were most often invoked to prosecute Republican editors, some of whom served jail terms. These measures in turn called forth the Virginia and Kentucky resolutions, drafted respectively by Madison and Jefferson, which invoked state sovereignty against intolerable federal powers. War with France often seemed imminent during this period, but Adams was determined to avoid issuing a formal declaration of war, and in this he succeeded.

Taxation, which had been levied to pay anticipated war costs, brought more discontent, however, including a new minor rebellion in Pennsylvania led by Jacob Fries. The rising was put down without difficulty, but widespread disagreement over issues ranging from civil liberties to taxation was polarizing American politics. A basic sense of political identity now divided Federalists from Republicans, and in the election of 1800 Jefferson drew on deep sources of Anti-Federalist opposition to challenge and defeat his old friend and colleague Adams. The result was the first contest over the presidency between political parties and the first actual change of government as a result of a general election in modern history.

The Jeffersonian Republicans in power. Jefferson began his presidency with a plea for reconciliation: "We are all Republicans, we are all Federalists." He had no plans for a permanent two-party system of government. He also began with a strong commitment to limited government and strict construction of the Constitution. All these commitments were soon to be tested by the exigencies of war, diplomacy, and political contingency.

On the American continent, Jefferson pursued a policy of expansion. He seized the opportunity when Napoleon Bonaparte decided to relinquish French ambitions in North America by offering the Louisiana territory for sale (Spain had recently ceded the territory to France). This extraordinary acquisition, purchased at a price of a few cents per acre, more than doubled the area of the United States. Jefferson had no constitutional sanction for such an exercise of executive power; he made up the rules as he went along, taking a broad construction view of the Constitution on this issue. He also sought opportunities to gain Florida from Spain; and, for scientific and political reasons, he sent Meriwether Lewis and William Clark on an expedition of exploration across the continent. This territorial expansion was not without problems. Various separatist movements periodically arose, including a plan for a Northern Confederacy formulated by New England

The Louisiana Purchase



The United States, 1783–1812.

Federalists. Aaron Burr, who had been elected Jefferson's vice president in 1800 but was replaced in 1804, led several western conspiracies. Arrested and tried for treason, he was acquitted in 1807.

As chief executive, Jefferson clashed with members of the judiciary, many of whom had been late appointments by Adams. One of his primary opponents was the late appointee Chief Justice John Marshall. In the case of *Marbury* v. *Madison* (1803), in which the Supreme Court first exercised the power of judicial review of congressional legislation, Marshall's main target was the executive, not the legislature; he dexterously succeeded in confirming but judiciously limiting the court's constitutional role without putting its actual authority at risk.

By the start of Jefferson's second term in office, Europe was engulfed in the Napoleonic Wars. The United States remained neutral, but both Britain and France imposed various orders and decrees severely restricting American trade with Europe and confiscated American ships for violating the new rules. Britain also conducted impressment raids in which U.S. citizens were sometimes seized. Unable to agree to treaty terms with Britain, Jefferson tried to coerce both Britain and France into ceasing to violate "neutral rights" with a total embargo on American exports, enacted by Congress in 1807. The results were catastrophic for American commerce and produced bitter alienation in New England, where the embargo (written backward as "O grab me") was held to be a Southern plot to destroy New England's wealth. In 1809, shortly after Madison was elected president, the embargo act was repealed.

Madison as president and the War of 1812. Madison's presidency was dominated by foreign affairs. Both Britain and France committed depredations on American shipping, but Britain was more resented, partly because with the greatest navy it was more effective and partly because Americans were extremely sensitive to British insults to national honour. Certain expansionist elements looking to both Florida and Canada began to press for war and took advantage of the issue of naval protection. Madison's own aim was to preserve the principle of freedom of the seas and to assert the ability of the United States to protect its own interests and its citizens. While striving to confront the European adversaries impartially, he was drawn into war against Britain, which was declared in June 1812 on a vote of 79–49 in the House and 19–13 in the Senate. There was almost no support for war in the Northern states.

The war began and ended in irony. The British had already rescinded the offending orders in council, but the news had not reached the United States at the time of the declaration. The Americans were poorly placed from every point of view. Ideological objections to armies and navies had been responsible for a minimal naval force. Ideological objections to banks had been responsible, in 1812, for the Senate's refusal to renew the charter of the Bank of the United States. Mercantile sentiment was hostile to the administration. Under the circumstances, it was remarkable that the United States succeeded in staggering through two years of war, eventually winning important naval successes at sea, on the Great Lakes, and on Lake Champlain. On land, a British raiding party burned public buildings in Washington, D.C., and drove President Madison to flee from his capital. The only action with long-term implications was Andrew Jackson's victory at New Orleans—won in January 1815, two weeks after peace had been signed in Ghent, Belg. Jackson's political reputation rose directly from this battle.

In historical retrospect, the most important aspect of the peace settlement was an agreement to set up a boundary commission for the Canadian border, which could thenceforth be left unguarded. It was not the end of Anglo-American hostility, but the agreement marked the advent of an era of mutual trust. The conclusion of the War of 1812, which has sometimes been called the Second War of American Independence, marked a historical cycle. It resulted in a pacification of the old feelings of pain and resentment against Great Britain and her people—still for many Americans a kind of paternal relationship. And, by freeing them of anxieties on this front, it also freed Americans to look to the west. (J.R.Po.)

The United States from 1816 to 1850

THE ERA OF MIXED FEELINGS

The years between the election to the presidency of James Monroe in 1816 and of John Quincy Adams in 1824 have long been known in American history as the Era of Good Feelings. The phrase was conceived by a Boston editor during Monroe's visit to New England early in his first term. That a representative of the heartland of Federalism could speak in such positive terms of the visit by a Southern president whose decisive election had marked not only a sweeping Republican victory but also the demise of the national Federalist Party was dramatic testimony that former foes were inclined to put aside the sectional and political differences of the past.

Effects of the War of 1812. Later scholars have questioned the strategy and tactics of the United States in the War of 1812, the war's tangible results, and even the wisdom of commencing it in the first place. To contemporary Americans, however, the striking naval victories and Andrew Jackson's victory over the British at New Orleans created a reservoir of "good feeling" on which Monroe was able to draw.

Foreign policy

Abetting the mood of nationalism was the foreign policy of the United States after the war. Florida was acquired from Spain (1819) in negotiations the success of which owed more to Andrew Jackson's indifference to such niceties as the inviolability of foreign borders and the nation's evident readiness to back him up than it did to diplomatic finesse. The Monroe Doctrine (1823), actually a few phrases inserted in a long presidential message, declared that the United States would not become involved in European affairs and would not accept European interference in the Americas; its immediate effect on other nations was slight, and that on its own citizenry was impossible to gauge, yet its self-assured tone in warning off the Old World from the New reflected well the nationalist mood that swept the nation.

Internally, the decisions of the Supreme Court under Chief Justice John Marshall in such cases as *McCulloch* v. *Maryland* (1819) and *Gibbons* v. *Ogden* (1824) promoted nationalism by strengthening Congress and national power at the expense of the states. The congressional decision to charter the second Bank of the United States (1816) was explained in part by the nation's financial weaknesses, exposed by the War of 1812, and in part by the intrigues of financial interests. The readiness of Southern Jeffersonians—former strict constructionists—to support such a measure indicates, too, an amazing degree of national feeling. Perhaps the clearest sign of a new sense of national unity was the victorious Republican Party, standing in solitary splendour on the national political horizon, its long-time foes the Federalists vanished without a trace (on the national level) and Monroe, the Republican standard-bearer, reelected so overwhelmingly in 1820 that it was long believed that the one electoral vote denied him had been held back only in order to preserve George Washington's record of unanimous selection.

National disunity. For all the signs of national unity and feelings of oneness, equally convincing evidence points in the opposite direction. The very Supreme Court decisions that delighted friends of strong national government infuriated its opponents, while Marshall's defense of the rights of private property was construed by critics as betraying a predilection for one kind of property over another. The growth of the West, encouraged by the conquest of Indian lands during the War of 1812, was by no means regarded as an unmixed blessing. Eastern conservatives sought to keep land prices high, speculative interests opposed a policy that would be advantageous to poor squatters, politicians feared a change in the sectional balance of power, and businessmen were wary of a new section with interests unlike their own. European visitors testified that, even during the so-called Era of Good Feelings, Americans characteristically expressed scorn for their countrymen in sections other than their own.

Opponents of nationalism

Economic hardship, especially the financial panic of 1819, also created disunity. The causes of the panic are complex, but its greatest effect was clearly the tendency of

its victims to blame it on one or another hostile or malevolent interest—whether the second Bank of the United States, Eastern capitalists, selfish speculators, or perfidious politicians—each charge expressing the bad feeling that existed side by side with the good.

If harmony seemed to reign on the level of national political parties, disharmony prevailed within the states. In the early 19th-century United States, local and state politics were typically waged less on behalf of great issues than for petty gain. That the goals of politics were often sordid did not mean that political contests were bland. In every section, state factions led by shrewd men waged bitter political warfare to attain or entrench themselves in power.

The most dramatic manifestation of national division was the political struggle over slavery, particularly over its spread into new territories. The Missouri Compromise of 1820 eased the threat of further disunity, at least for the time being—the sectional balance between the states was preserved; in the Louisiana Purchase, with the exception of the Missouri Territory, slavery was to be confined to the area south of the 36°30′ line. Yet this compromise did not end the crisis but only postponed it. The determination by Northern and Southern senators not to be outnumbered by one another suggests that the people continued to believe in the conflicting interests of the various great geographic sections. The weight of evidence indicates that the decade after the Battle of New Orleans was an era not of good feelings so much as it was one of mixed feelings.

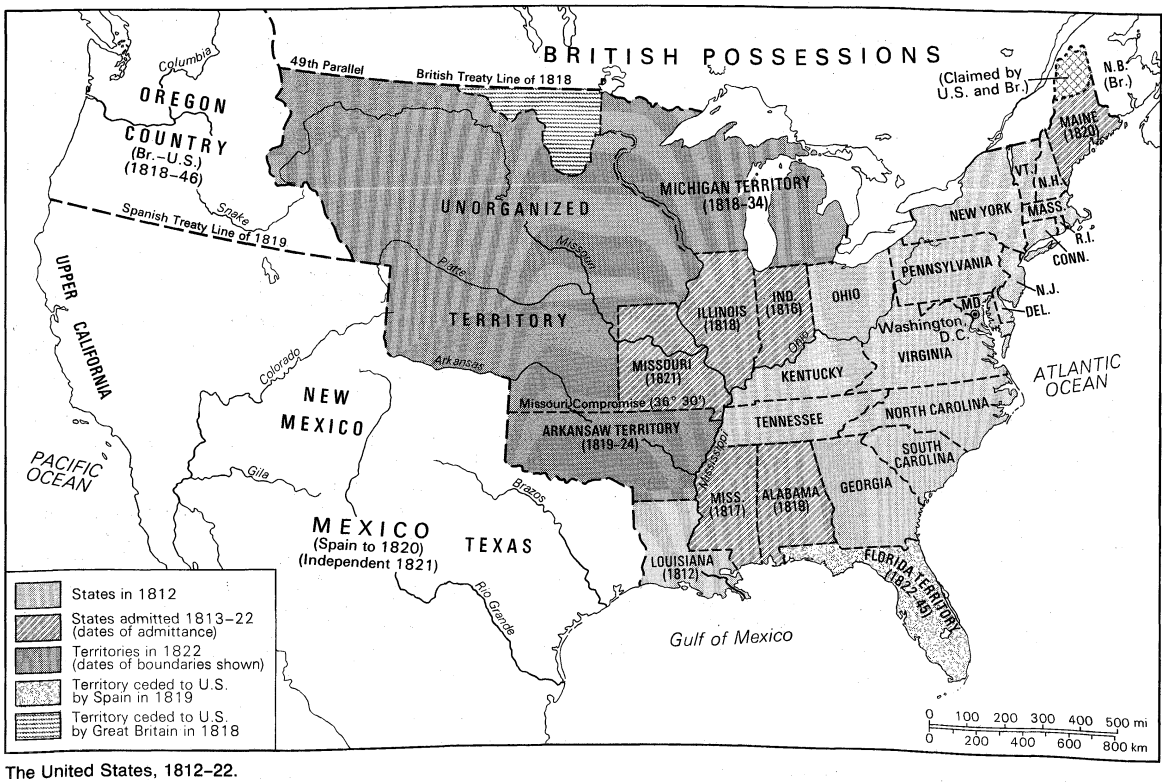
THE ECONOMY

Economic growth and maturation

The American economy expanded and matured at a remarkable rate in the decades after the War of 1812. The rapid growth of the West created a great new centre for the production of grains and pork, permitting the nation's older sections to specialize in other crops. New processes of manufacture, particularly in textiles, not only accelerated an "industrial revolution" in the Northeast but also, by drastically enlarging the Northern market for raw materials, helped account for a boom in Southern cotton production. If by midcentury the white South had come to regard slavery as a "positive good" rather than the "necessary evil" it had earlier held the system to be, it was due largely to the increasingly central role played by cotton in earning profits for the section; the cotton economy relied on slavery. Industrial workers organized the nation's first trade unions and even workingmen's political parties early in the period. The corporate form thrived in an era the booming capital requirements of which made older and simpler forms of attracting investment capital obsolete. Commerce became increasingly specialized, the division of labour in the disposal of goods for sale matching the increasingly sophisticated division of labour that had come to characterize production. Banks were created in unprecedented numbers, turning out quantities of paper money to meet the thriving economy's need for additional exchange. The fact that little coin or specie was actually kept in vaults to back up this paper explains both why many bank notes were discounted severely and why the state of the economy was typically unstable. A rage for speculation was widely noted, encouraged by an alarmingly rapid rise in real-estate values and in the production by state banks of large quantities of paper money on demand. Probably the most important changes occurred in the nation's system for moving people and goods. According to many later scholars, a "transportation revolution" was the key to almost all other economic changes in the period.

Transportation revolution. The controversial political issue of "internal improvements" focused on a simple question: would national government finance local and state transportation projects? That some presidents hesitated, in the absence of explicit language in the Constitution urging federal support, had little effect in dampening the near mania for such projects. More federal moneys were expended on them under President Andrew Jackson (served 1829–37), who on this issue was a strict constructionist, than in the administrations of all previous presidents combined. Actually, most of the capital was raised by state governments, by private citizens, and from abroad. In their turn, turnpikes (or toll roads), canals, steamboats, and railroads inspired booms, typically featured by lack of planning, business failures, shoddy construction, profits that fell far short of expectations largely because costs exceeded anticipations, rampant corruption, and, withal, an improved transportation system that was the wonder of the world. The system did well the basic job that transportation was called on to do for a swiftly expanding nation.

Waterways. Steamboats replaced rafts on the Mississippi and sharply reduced the price of Latin-American coffee in ports upstream by drastically cutting shipping time and expense. The Erie Canal, the most publicized as



The United States, 1812–22.

well as the most successful individual project constructed during the era, enabled efficient Western grain producers ultimately to undersell Eastern farmers in the distant New York City market for similar reasons. Imitators of the Erie in Pennsylvania and throughout the South and West discovered, however, that canals were poor business propositions unless cheaply and efficiently built, heavily used, and spared from drought or flooding.

Railroads. A controversy has developed among scholars over the significance of railroads in antebellum America—one interpretation regards their development as the key to the era's industrialization, while another viewpoint holds that the West would have been developed as quickly and goods moved as cheaply without them. In any case, the speed and reliability of the "iron horse" attracted investors and, at first, passengers seeking comfort in travel, rather than freight. By midcentury, however, the advantages of the new form, for all its expense, were luring shippers of industrial and agricultural products to use the nation's thousands of miles of railroad network.

Beginnings of industrialization. Agriculture remained an important industry in all sections, although a sectional division of labour became increasingly discernible at the era's end, spurred by transportation and mechanical developments that permitted more productive areas to undersell their competitors even in the home markets of the latter. Nevertheless, New York and Pennsylvania in the Northeast and Kentucky and Tennessee in the South remained important producers of corn, wheat, and livestock even in the 1840s. The trend, however, was in the direction of specialization. Cotton became king, not only in the South but in the nation as a whole, because its sale overseas brought in more money than the sale of all other products combined. Slavery accordingly became more entrenched, despite the sharp rise in the price of slaves that marked the era. Contemporary Southerners put ever more capital into the system that promised quick profits and a "social harmony" based on the total subordination of the slaves. Slavery was a complex system, particularly in cities, marked by the hiring out of tens of thousands of skilled slave artisans and by an amazing degree of free physical movement and personal behaviour displayed by slaves in such a city as New Orleans. When most of the country referred to slavery, however, they appeared to mean the plantation system based on the labour of hundreds of field hands, though, actually, a minority of Southern whites owned slaves.

Trend toward industrial specialization

By midcentury, factories accounted for most textile production in the northeastern states, while the factory system was beginning to spread across the states of the Ohio valley. The labour organizations that sprang up in the nation's cities were composed not of machine hands but primarily of skilled mechanics. A major purpose of this movement was to enable their membership to withstand the spread of a system associated by labour spokesmen with speedup, the devaluation of skill, low wages, child and female labour, and a general debasement of the working class. The spread of the factory system was not to be deflected, however, not even by a movement that at its height claimed a membership of 300,000 (a claim that was undocumented and undoubtedly exaggerated). The labour movement was crushed not by industrialization but rather by a depression that followed financial panics in 1837 and 1839.

These panics, like the earlier crisis in 1819, illustrate well the erratic course of the American economy during the era. Growth was not unbroken. Overspeculation, inflation, and governmental inaction in some instances or chicanery in others created an atmosphere of instability that was as characteristic of the era as were its tangible achievements. Hundreds of bank and business failures, large-scale unemployment, and hard times followed in the wake of these debacles, lasting in the latter case to the mid-1840s before the economy again moved forward, more productive than ever, to resume its upward thrust.

SOCIAL DEVELOPMENTS

In the decades before the Civil War (1861–65), the civilization of the United States exerted an irresistible pull on visitors, hundreds of whom were assigned to report back to European audiences that were fascinated by the new society and insatiable for information on every facet of the "fabled republic." What appeared to intrigue the travelers above all was the uniqueness of American society. In contrast to the relatively static and well-ordered civilization of the Old World, America seemed turbulent, dynamic, and in constant flux, its people crude but vital, awesomely ambitious, optimistic, and independent. Many well-bred Europeans were evidently taken aback by the self-assurance of lightly educated American common folk. Ordinary Americans seemed unwilling to defer to anyone on the basis of rank or status.

European interest in American life

The people. American society was rapidly changing. Population grew at what to Europeans was an amazing rate—although it was the normal pace of American population growth for the antebellum decades—of between three-tenths and one-third a decade. After 1820 the rate of growth was not uniform throughout the country. New England and the Southern Atlantic states languished—the former region because it was losing settlers to the superior farmlands of the Western Reserve, the latter because its economy offered too few places to newcomers.

The special feature of the population increase of the 1830s and '40s was the extent to which it was composed of immigrants. Whereas about 250,000 Europeans had come in the first three decades of the 19th century, 10 times as many arrived between 1830 and 1850. The newcomers were overwhelmingly Irish and German. Traveling in family groups rather than as individuals, they were attracted by the dazzling opportunities of American life: abundant work, land, food, and freedom on the one hand and the absence of compulsory military service on the other.

German and Irish immigrants

The German contingent did well, settling mostly on semi-improved farms and towns in the Ohio valley, their success promoted by their relatively prosperous state on arrival and by the solid aid given newcomers by the efficient network of economic and cultural organizations founded by earlier German settlers. Irish immigrants, however, fared poorly; too poor to buy land, lacking in skills, disorganized, and members of a faith considered alien and even dangerous by many native Americans, the Irish suffered various forms of ostracism and discrimination in the cities, where they tended to congregate. They provided the menial and unskilled labour needed by the expanding economy. Their low wages forced them to live in tightly packed slums, whose chief features were filth, disease, rowdyism, prostitution, drunkenness, crime, a high mortality rate, and the absence of even rudimentary toilet facilities. Adding to the woes of the first generation of Irish immigrants was the tendency of many disgruntled natives to treat the newcomers as scapegoats who allegedly threatened the future of American life and religion. In the North, only free blacks were treated worse.

Most Northern blacks possessed theoretical freedom and little else. Confined to menial occupations for the most part, they fought a losing battle against the inroads of Irish competition in northeastern cities. The struggle between the two groups erupted spasmodically into ugly street riots. The hostility shown free blacks by the general community was less violent but equally unremitting. Discrimination in politics, employment, education, housing, religion, and even in cemeteries resulted in a cruelly oppressive system. Unlike a slave, the free Northern Negro could criticize and petition against his subjugation, but this proved fruitless in preventing the continued deterioration of his situation.

Most Americans continued to live in the country. Although improved machinery had resulted in expanded farm production and had given further impetus to the commercialization of agriculture, the way of life of independent agriculturists had changed little by midcentury. The public journals put out by some farmers insisted that their efforts were unappreciated by the larger community. The actuality was complex. Many farmers led lives marked by unremitting toil, cash shortage, and little leisure. Farm workers received minuscule wages. In all sections of the nation much of the best land was concentrated in the hands of a small number of wealthy farmers. The proportion of farm families that owned their own land, however,

was far greater in the United States than in Europe, and varied evidence points to a steady improvement in the standard and style of living of agriculturalists as midcentury approached.

Urban expansion

Cities. Cities, both old and new, thrived during the era, their growth in population outstripping the spectacular growth rate of the nation as a whole and their importance and influence far transcending the relatively small proportions of citizens living in them. Whether on the "urban frontier" or in the older seaboard region, antebellum cities were the centres of wealth and political influence of their outlying hinterlands. New York City, with a population approaching 500,000 by midcentury, faced problems of a different order of magnitude from those confronting such cities as Poughkeepsie or Newark. Yet the pattern of change during the era was amazingly similar for eastern cities or western, old cities or new, great cities or small. The lifeblood of them all was commerce. Old ideals of economy in town government were grudgingly abandoned by the merchant, professional, and landowning elites that typically ruled. Taxes were increased in order to deal with pressing new problems and to enable the urban community of midcentury to realize new opportunities. Harbours were improved, police forces professionalized, services expanded, waste more reliably removed, streets improved, and welfare activities broadened, all as the result of the statesmanship and the self-interest of property owners who were convinced that amelioration was socially beneficial.

Education and religion. Cities were also centres of educational and intellectual progress. The emergence of a relatively well-financed public educational system, free of the stigma of "pauper" or "charity" schools, and the emergence of a lively "penny press," made possible by a technological revolution, were among the most important developments. An Evangelical movement that swept the Northeast and West before 1840 was largely an urban phenomenon. Cutting across Protestant denominational lines, the movement was regarded by many of its leaders as a struggle against satanic influences that thrived best in the secular atmosphere of cities. Influential merchants made generous contributions to this great "revival," which combined detailed, fiery exhortations against sin and the Devil with a social message of unabashed conservatism. The urban wealthy had reason to find such a message useful.

Wealth. The brilliant French visitor Alexis de Tocqueville, in common with most contemporary observers, believed American society to be remarkably egalitarian.

Illusions about American wealth

Most rich American men were thought to have been born poor; "self-made" was the term Henry Clay popularized for them. The society was allegedly a very fluid one, marked by the rapid rise and fall of fortunes, with room at the top accessible to all but the most humble; opportunity for success seemed freely available to all, and although material possessions were not distributed perfectly equally they were, in theory, dispersed so fairly that only a few poor and a few rich men existed at either end of the social spectrum.

The actuality, however, was far different. While the rich were inevitably not numerous, America by 1850 had more millionaires than all of Europe. New York, Boston, and Philadelphia had perhaps 1,000 individuals, each admitting to $100,000 or more, at a time when wealthy taxpayers kept secret from assessors the bulk of their wealth. Because an annual income of $4,000 or $5,000 enabled a man to live luxuriously, these were great fortunes indeed. Typically, the wealthiest 1 percent of urban citizens owned approximately one-half the wealth of the great cities of the Northeast, while the great bulk of their populations possessed little or nothing. In what has long been called the "age of the common man," rich men were almost invariably born not into humble or poor families but into wealthy and prestigious ones. In western cities, too, class lines increasingly hardened after 1830. The common man lived in the age, but he did not dominate it. It appears that contemporaries, overimpressed with the absence of a titled aristocracy and with the democratic tone and manner of American life, failed to see the extent to which money, family, and status exerted power in the New World even as they did in the Old.

JACKSONIAN DEMOCRACY

The democratization of politics. American politics became increasingly democratic during the 1820s and '30s. Local and state offices that had earlier been appointive became elective. The suffrage was expanded as property and other restrictions on voting were reduced or abandoned in most states. The freehold requirement that had denied voting to all but holders of real estate was almost everywhere discarded before 1820, while the taxpaying qualification was also removed, if more slowly and gradually. In many states a printed ballot replaced the earlier system of voice voting, while the secret ballot also grew in favour. Whereas in 1800 only two states provided for the popular choice of presidential electors, by 1832 only South Carolina still left the decision to the legislature. Conventions of elected delegates increasingly replaced legislative or congressional caucuses as the agencies for making party nominations. By the latter change, a system for nominating candidates by self-appointed cliques meeting in secret was replaced by a system of open selection of candidates by democratically elected bodies.

Extension of voting rights

These democratic changes were not engineered by Andrew Jackson and his followers, as was once believed. Most of them antedated the emergence of Jackson's Democratic Party, and in New York, Mississippi, and other states some of the reforms were accomplished over the objections of the Jacksonians. There were men in all sections who feared the spread of political democracy, but by the 1830s few were willing publicly to voice such misgivings. Jacksonians effectively sought to fix the impression that they alone were champions of democracy, engaged in mortal struggle against aristocratic opponents. The accuracy of such propaganda varied according to local circumstances. The great political reforms of the early 19th century in actuality were conceived by no one faction or party. The real question about these reforms concerns the extent to which they truly represented the victory of democracy in the United States.

Small cliques or entrenched "machines" dominated democratically elected nominating conventions as earlier they had controlled caucuses. While by the 1830s the common man—of white if not of black or red skin—had come into possession of the vote in most states, the nomination process continued to be outside his control. More importantly, the policies adopted by competing factions and parties in the states owed little to ordinary voters. The legislative programs of the "regencies" and juntos that effectively ran state politics were designed primarily to reward the party faithful and to keep them in power. State parties extolled the common people in grandiloquent terms but characteristically focused on prosaic legislation that awarded bank charters or monopoly rights to construct transportation projects to favoured insiders. That American parties would be pragmatic vote-getting coalitions, rather than organizations devoted to high political principles, was due largely to another series of reforms enacted during the era. Electoral changes that rewarded winners or plurality gatherers in small districts, in contrast to a previous system that divided a state's offices among the several leading vote getters, worked against the chances of "single issue" or "ideological" parties while strengthening parties that tried to be many things to many men.

The status of the common man

The Jacksonians. To his army of followers, Andrew Jackson was the embodiment of popular democracy. A truly self-made man of will and courage, he personified for many citizens the vast power of nature and Providence, on the one hand, and the majesty of the people, on the other. His very weaknesses, such as a nearly uncontrollable temper, were political strengths. Opponents who branded him enemy to property and order only gave credence to the claim of Jackson's supporters that he stood for the poor against the rich, the plain people against the interests.

Jackson, like most of his leading antagonists, was in fact a wealthy man of conservative social beliefs. In his many volumes of correspondence he rarely referred to labour. As a lawyer and man of affairs in Tennessee prior to his accession to the presidency, he aligned himself not with have-nots but with the influential, not with the debtor but with the creditor. His reputation was created largely

by astute men who propagated the belief that his party was the people's party and that the policies of his administrations were in the popular interest. Savage attacks on those policies by some wealthy critics only fortified the belief that the Jacksonian movement was radical as well as democratic.

Birth of the Democratic Party

At its birth in the mid-1820s, the Jacksonian, or Democratic, Party was a loose coalition of diverse men and interests united primarily by a practical vision. They held to the twin beliefs that Old Hickory, as Jackson was known, was a magnificent candidate and that his election to the presidency would benefit those who helped bring it about. His excellence as candidate derived in part from the fact that he appeared to have no known political principles of any sort. In this period there were no distinct parties on the national level. Jackson, Henry Clay, John C. Calhoun, John Quincy Adams, and William H. Crawford—the leading presidential aspirants—all portrayed themselves as "Republicans," followers of the party of the revered Jefferson. The National Republicans were the followers of Adams and Clay; the Whigs, who emerged in 1834, were, above all else, the party dedicated to the defeat of Jackson.

The major parties. The great parties of the era were thus created to attain victory for men rather than measures. Once in being, their leaders understandably sought to persuade the electorate of the primacy of principles. It is noteworthy, however, that former Federalists at first flocked to the new parties in largely equal numbers and that men on opposite sides of such issues as internal improvements or a national bank could unite behind Jackson. With the passage of time, the parties did come increasingly to be identified with distinctive, and opposing, political policies.

By the 1840s, Whig and Democratic congressmen voted as rival blocs. Whigs supported and Democrats opposed a weak executive, a new Bank of the United States, a high tariff, distribution of land revenues to the states, relief legislation to mitigate the effects of the depression, and federal reapportionment of House seats. Whigs voted against and Democrats approved an independent treasury, an aggressive foreign policy, and expansionism. These were important issues, capable of dividing the electorate just as they divided the major parties in Congress. Certainly it was significant that Jacksonians were more ready than their opponents to take punitive measures against blacks or abolitionists or to banish and use other forceful measures against the southern Indian tribes, brushing aside treaties protecting Indian rights. But these differences do not substantiate the belief that the Democrats and Whigs were divided ideologically, with only the former somehow representing the interests of the propertyless.

Party lines earlier had been more easily broken, as during the crisis that erupted over South Carolina's bitter objections to the high Tariff of 1828. Jackson's firm opposition to Calhoun's policy of nullification (*i.e.,* the right of a state to nullify a federal law, in this case the tariff) had commanded wide support within and outside the Democratic Party. Clay's solution to the crisis, a compromise tariff, represented not an ideological split with Jackson but Clay's ability to conciliate and to draw political advantage from astute tactical maneuvering.

Jackson's attack on the Bank of the United States

The Jacksonians depicted their war on the second Bank of the United States as a struggle against an alleged aristocratic monster that oppressed the West, debtor farmers, and poor people generally. Jackson's decisive reelection in 1832 was once interpreted as a sign of popular agreement with the Democratic interpretation of the bank war, but recent evidence discloses that Jackson's margin was hardly unprecedented and that Democratic success may have been due to other considerations. The second bank was evidently well-thought-of by many Westerners, many farmers, and even by Democratic politicians who admitted to opposing it primarily not to incur the wrath of Andrew Jackson.

Jackson's reasons for detesting the bank and Nicholas Biddle, its president, were complex. Anticapitalist ideology would not explain a Jacksonian policy that replaced a quasi-national bank as repository of government funds with dozens of state and private banks, equally controlled by capitalists and even more dedicated than was Biddle to profit making. The saving virtue of these "pet banks" appeared to be the Democratic political affiliations of their directors. Perhaps the pragmatism as well as the large degree of similarity between the Democrats and Whigs is best indicated by their frank adoption of the "spoils system." The Whigs, while out of office, denounced the vile Democratic policy for turning lucrative customhouse and other posts over to supporters; but once in office they resorted to similar practices. It is of interest that the Jacksonian appointees were hardly more plebeian than were their so-called aristocratic predecessors.

Minor parties. The politics of principle was represented during the era not by the major parties but by the minor ones. The Anti-Masons aimed to stamp out an alleged aristocratic conspiracy. The Workingmen's Party called for "social justice." The Locofocos (so named after the matches they used to light up their first meeting in a hall darkened by their opponents) denounced monopolists in the Democratic Party and out. The variously named nativist parties accused the Roman Catholic church of all manner of evil. The Liberty Party opposed the spread of slavery. All of these parties were ephemeral since they proved incapable of mounting a broad appeal that attracted masses of voters in addition to their original constituencies. The Democratic and Whig parties thrived not in spite of their opportunism but because of it, reflecting well the practical spirit that animated most American voters.

AN AGE OF REFORM

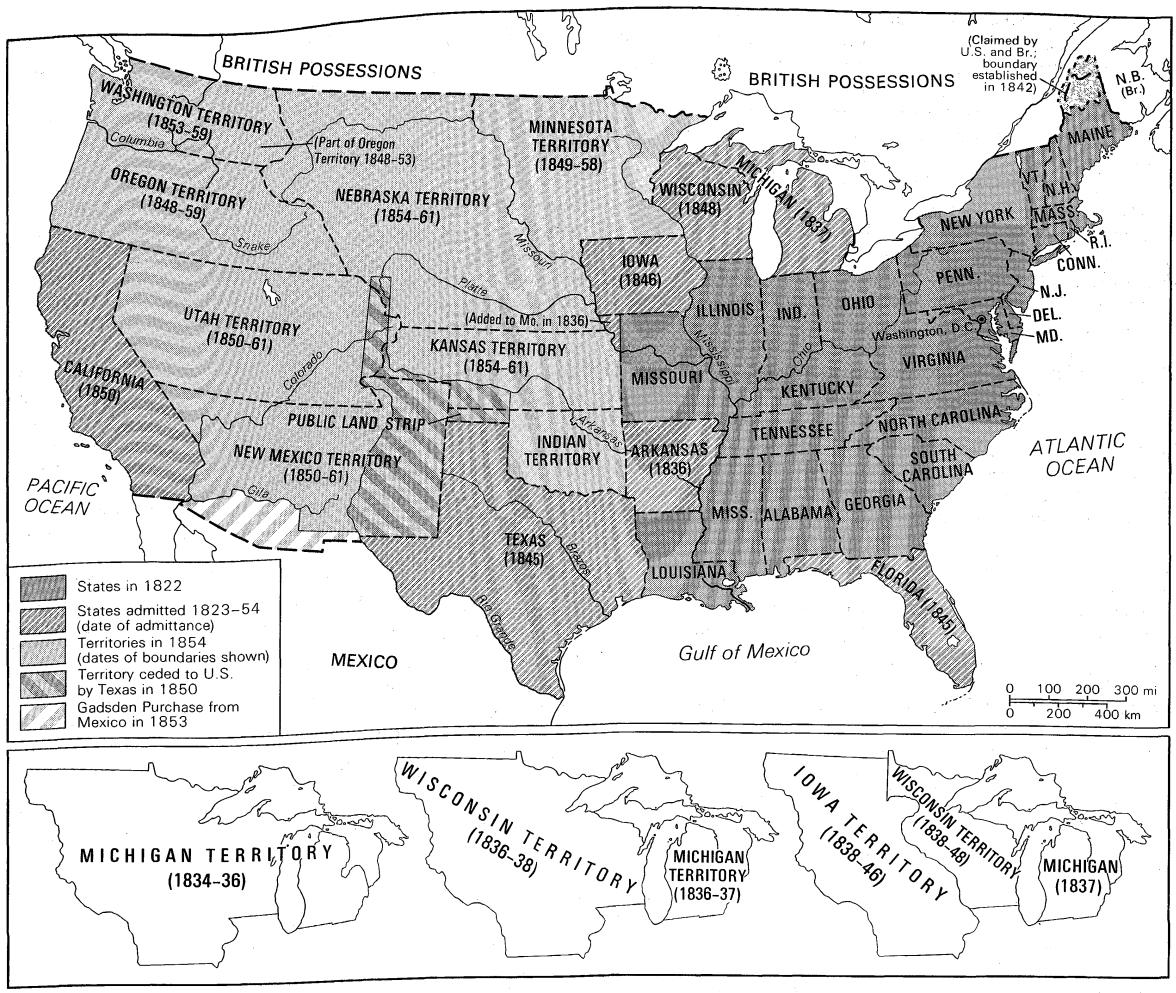
Historians have labeled the period 1830–50 an "age of reform." At the same time that the pursuit of the dollar was becoming so frenzied that some observers called it the nation's true religion, tens of thousands of Americans joined an array of movements dedicated to spiritual and secular uplift. There is not yet agreement as to why a rage for reform erupted in the antebellum decades. A few of the explanations cited, none of them conclusive, include an outburst of Protestant Evangelicalism, a reform spirit that swept across the Anglo-American community, a delayed reaction to the perfectionist teachings of the Enlightenment, and the worldwide revolution in communications that was a feature of 19th-century capitalism.

Variety of reform movements

What is not in question is the amazing variety of reform movements that flourished simultaneously in the northern states—women's rights, pacifism, temperance, prison reform, abolition of imprisonment for debt, an end to capital punishment, improving the conditions of the working classes, a system of universal education, the organization of communities that discarded private property, improving the condition of the insane and the congenitally enfeebled, and the regeneration of the individual were among the causes that inspired zealots during the era.

Abolitionism. There can be no doubt that antislavery, or "abolition" as it came to be called, was the nonpareil reform. Abolition was a diverse phenomenon. At one end of its spectrum was William Lloyd Garrison, an "immediatist," who denounced not only slavery but the Constitution of the United States for tolerating the evil. His newspaper, *The Liberator,* lived up to its promise that it would not equivocate in its war against slavery. Garrison's uncompromising tone infuriated not only the South but many Northerners as well and was long treated as though it were typical of abolitionism in general. Actually it was not. At the other end of the abolitionist spectrum and in between stood such men and women as Theodore Weld, James Birney, Gerrit Smith, Theodore Parker, Julia Ward Howe, Lewis Tappan, Salmon P. Chase, and Lydia Maria Child, all of whom represented a variety of stances, all more conciliatory than Garrison's. James Russell Lowell, whose emotional balance has been cited by a recent biographer as proof that abolitionists need not have been unstable, urged in contrast to Garrison that "the world must be healed by degrees."

Whether they were Garrisonians or not, abolitionist leaders have been scorned as cranks who were either working out their own personal maladjustments or as people using the slavery issue to restore a status that as an alleged New England elite they feared they were losing. The truth may be simpler. Few neurotics and few members of the



The United States, 1822–54.

northern socioeconomic elite became abolitionists. For all the movement's zeal and propagandistic successes, it was bitterly resented by many Northerners, and the masses of free whites were indifferent to its message. In the 1830s, urban mobs, typically led by "gentlemen of property and standing," stormed abolitionist meetings, wreaking violence on the property and persons of blacks and their white sympathizers, evidently indifferent to the niceties distinguishing one abolitionist theorist from another. The fact that abolition leaders were remarkably similar in their New England backgrounds, their Calvinist self-righteousness, their high social status, and the relative excellence of their educations is hardly evidence that their cause was either snobbish or elitist. Ordinary citizens were more inclined to loathe Negroes and to preoccupy themselves with personal advance within the system.

Resentment of abolitionists

Support of reform movements. The existence of many reform movements did not mean that a vast number of Americans supported them. Abolition did poorly at the polls. Some reforms were more popular than others, but by and large none of the major movements had mass followings. The evidence indicates that few persons actually participated in these activities. Utopian communities such as Brook Farm and those in New Harmony, Ind., and Oneida, N.Y., did not succeed in winning over many followers or in inspiring many other groups to imitate their example. The importance of these and the other movements derived neither from their size nor their achievements. Reform reflected the sensitivity of a small number of persons to imperfections in American life. In a sense, the reformers were "voices of conscience," reminding their materialistic fellow citizens that the American Dream was not yet a reality, pointing to the gulf between the ideal and the actuality.

Religious-inspired reform. A unique feature of antebellum reform was its religious character. Unlike European social critics of the same era, who were not only secular but often antireligious, American perfectionists were largely inspired by religious zeal. Not that religious enthusiasm was invariably identified with social uplift; many reformers were more concerned with saving souls than with curing social ills. The merchant princes who played active roles in—and donated large sums of money to—the Sunday school unions, home missionary societies, and Bible and tract societies did so in part out of altruism, in part because the latter organizations stressed spiritual rather than social improvement while teaching the doctrine of the "contented poor." In effect, conservatives who were strongly religious found no difficulty in using religious institutions to fortify their social predilections. Radicals, on the other hand, interpreted Christianity as a call to social action, convinced that true Christian rectitude could be achieved only in struggles that infuriated the smug and the greedy. Ralph Waldo Emerson was an example of the American reformer's insistence on the primacy of the individual. The great goal according to him was the regeneration of the human spirit, rather than a mere improvement in material conditions. Emerson and reformers like him, however, acted on the premise that a foolish consistency was indeed the hobgoblin of little minds, for they saw no contradiction in uniting with like-minded idealists to act out or argue for a new social model. The spirit was to be revived and strengthened through forthright social action undertaken by similarly independent individuals.

Conservative religion

EXPANSIONISM AND POLITICAL CRISIS AT MIDCENTURY

Throughout the 19th century, eastern settlers kept spilling over into the Mississippi valley and beyond, pushing the frontier farther westward. (In 1893 the historian Frederick Jackson Turner was to say that this ever-moving frontier was the most decisive influence on American civilization and values.) The Louisiana Purchase territory offered am-

ple room to pioneers and those who came after. American wanderlust, however, was not confined to that area. Throughout the era Americans in varying numbers moved into regions south, west, and north of the Louisiana Territory. Because Mexico and Great Britain held or claimed most of these lands, dispute inevitably broke out between these governments and the United States.

Westward expansion. The growing nationalism of the American people was effectively engaged by Democratic presidents Jackson and James K. Polk (served 1845–49) and by the expansionist Whig president John Tyler (served 1841–45) to promote their goal of enlarging the "empire for liberty." Each of these presidents performed shrewdly. Jackson waited until his last day in office to establish formal relations with the Republic of Texas, one year after his friend Sam Houston had succeeded in dissolving the ties between Mexico and the newly independent state of Texas. On the Senate's overwhelming repudiation of his proposed treaty of annexation, Tyler resorted to the use of a joint resolution so that each house could vote by a narrow margin for incorporation of Texas into the Union. Polk succeeded in getting the British to negotiate a treaty (1846) whereby the Oregon country south of the 49th parallel would revert to the United States. These were precisely the terms of his earlier proposal, which had been rejected by the British. Ready to resort to almost any means to secure the Mexican territories of New Mexico and upper California, Polk used a border incident as a pretext for commencing a war with Mexico. The war was not widely acclaimed and many congressmen disliked it, but few dared to oppose the appropriations that financed it.

The Mexican War

Although there is no evidence that these actions had anything like a public mandate, clearly they did not evoke widespread opposition. Nonetheless, the expansionists' assertion that Polk's election in 1844 could be construed as a popular clamour for the annexation of Texas was hardly a solid claim; Clay was narrowly defeated and would have won but for the defection from Whig ranks of small numbers of Liberty Party and nativist voters. The nationalistic idea, conceived in the 1840s by a Democratic editor, that it was the "manifest destiny" of the United States to expand westward to the Pacific undoubtedly prepared public opinion for the militant policies undertaken by Polk shortly thereafter. It has been said that this notion represented the mood of the American people; it is safer to say it reflected the feelings of many of the people.

Attitudes toward expansionism. Public attitudes toward expansion into Mexican territories were very much affected by the issue of slavery. Those opposed to the spread of slavery or simply not in favour of the institution joined abolitionists in discerning a proslavery policy in the Mexican War. The great political issue of the postwar years concerned slavery in the territories. Calhoun and spokesmen for the slave-owning South argued that slavery could not be constitutionally prohibited in the Mexican cession. "Free Soilers" supported the Wilmot Proviso idea—that slavery should not be permitted in the new territory. Others supported the proposal that "squatter sovereignty" should prevail—settlers in the territories should decide the issue. Still others called for the extension westward of the 36°30′ line of demarcation for slavery that had resolved the Missouri controversy in 1820. Now, 30 years later, Clay again pressed a compromise on the nation, supported dramatically by the aging Daniel Webster and by moderates in and out of the Congress. As the events in the California gold fields showed (beginning in 1849), many people had things other than political principles on their minds. The Compromise of 1850, as the separate resolutions resolving the controversy came to be known, infuriated those of high principle on both sides of the issue—Southerners resented that the compromise admitted California as a free state, abolished the slave trade in the District of Columbia, and gave territories the theoretical right to deny existence to their "peculiar institution," while antislavery men deplored the same theoretical right of territories to permit the institution and abhorred the new, more-stringent federal fugitive-slave law. That Southern political leaders ceased talking secession shortly after the enactment of the compromise indicates who truly won the political skirmish. The people probably approved the settlement—but as subsequent events were to show, the issues had not been met but only deferred. (E.Pe.)

The Compromise of 1850

Civil War

PRELUDE TO WAR, 1850–60

Before the Civil War the United States experienced a whole generation of nearly unremitting political crisis. Underlying the problem was the fact that America in the early 19th century had been a country, not a nation. The major functions of government—those relating to education, transportation, health, and public order—were performed on the state or local level, and little more than a loose allegiance to the government in Washington, a few national institutions such as churches and political parties, and a shared memory of the Founding Fathers of the republic tied the country together. Within this loosely structured society every section, every state, every locality, every group could pretty much go its own way.

Gradually, however, changes in technology and in the economy were bringing all the elements of the country into steady and close contact. Improvements in transportation—first canals, then toll roads, and especially railroads—broke down isolation and encouraged the boy from the country to wander to the city, the farmer from New Hampshire to migrate to Iowa. Improvements in the printing press, which permitted the publication of penny newspapers, and the development of the telegraph system broke through the barriers of intellectual provincialism and made everybody almost instantaneously aware of what was going on throughout the country. As the railroad network proliferated, it had to have central direction and control; and national railroad corporations—the first true "big businesses" in the United States—emerged to provide order and stability.

For many Americans the wrench from a largely rural, slow-moving, fragmented society in the early 1800s to a bustling, integrated, national social order in the mid-century was an abrupt and painful one, and they often resisted it. Sometimes resentment against change manifested itself in harsh attacks upon those who appeared to be the agents of change—especially immigrants, who seemed to personify the forces that were altering the older America. Vigorous nativist movements appeared in most cities during the 1840s; but not until the 1850s, when the huge numbers of Irish and German immigrants of the previous decade became eligible to vote, did the antiforeign fever reach its peak. Directed both against immigrants and against the Roman Catholic church, to which so many of them belonged, the so-called Know-Nothings emerged as a powerful political force in 1854 and increased the resistance to change.

Sectionalism and slavery. A more enduring manifestation of hostility toward the nationalizing tendencies in American life was the reassertion of strong feelings of sectional loyalty. New Englanders felt threatened by the West, which drained off the ablest and most vigorous members of the labour force and also, once the railroad network was complete, produced wool and grain that undersold the products of the poor New England hill country. The West, too, developed a strong sectional feeling, blending its sense of its uniqueness, its feeling of being looked down upon as raw and uncultured, and its awareness that it was being exploited by the businessmen of the East.

Reassertion of sectional loyalty

The most conspicuous and distinctive section, however, was the South—an area set apart by climate; by a plantation system designed for the production of such staple crops as cotton, tobacco, and sugar; and, especially, by the persistence of Negro slavery, which had been abolished or prohibited in all other parts of the United States. It should not be thought that all or even most white Southerners were directly involved in the section's "peculiar institution." Indeed, in 1850 there were only 347,525 slaveholders in a total white population of about 6,000,000 in the slave states. Half of these owned four slaves or fewer and could not be considered planters. In the entire South there were fewer than 1,800 persons who owned more than 100 slaves.

Nevertheless, slavery did give a distinctive tone to the whole pattern of Southern life. If the large planters were few, they were also wealthy, prestigious, and powerful; often they were the political as well as the economic leaders of their section; and their values pervaded every stratum of Southern society. Far from opposing slavery, small farmers thought only of the possibility that they too might, with hard work and good fortune, some day join the ranks of the planter class—to which they were closely connected by ties of blood, marriage, and friendship. Behind this virtually unanimous support of slavery lay the universal belief—shared by many whites in the North and West as well—that blacks were an innately inferior people who had risen only to a state of barbarism in their native Africa and who could live in a civilized society only if disciplined through slavery. Though by 1860 there were in fact about 250,000 free blacks in the South, most Southern whites resolutely refused to believe that the slaves, if freed, could ever coexist peacefully with their former masters. With shuddering horror they pointed to an insurrection of blacks that had occurred in Santo Domingo, to a brief slave rebellion led by the Negro Gabriel in Virginia in 1800, to a plot of Charleston, S.C., blacks headed by Denmark Vesey in 1822, and, especially, to a bloody and determined Virginia insurrection led by Nat Turner in 1831 as evidence that black persons had to be kept under iron control. Facing increasing opposition to slavery outside their section, Southerners developed an elaborate proslavery argument, defending the institution on biblical, economic, and sociological grounds.

Black insurrections

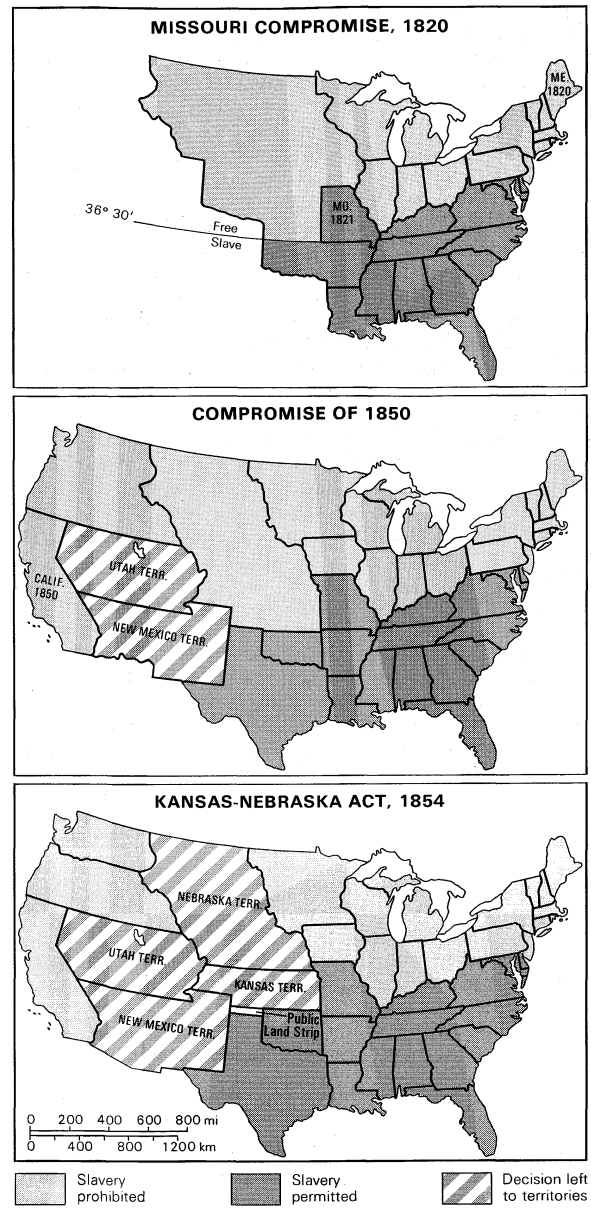
A decade of political crises. In the early years of the republic, sectional differences had existed, but it had been possible to reconcile or ignore them because distances were great, communication was difficult, and the powerless national government had almost nothing to do. The revolution in transportation and communication, however, eliminated much of the isolation, and the victory of the United States in its brief war with Mexico left the national government with problems that required action.

Popular sovereignty. The Compromise of 1850 was an uneasy patchwork of concessions to all sides that began to fall apart as soon as it was enacted. Most unsatisfactory of all in the long run would be the principle of popular sovereignty, which was bound to make of each territory a battleground where the supporters of the South would contend with the defenders of the North and West.

The seriousness of those conflicts became clear in 1854, when Stephen A. Douglas introduced his Kansas bill in Congress, establishing a territorial government for the vast region that lay between the Missouri River and the Rocky Mountains. In the Senate the bill was amended to create not one but two territories—Kansas and Nebraska—from the part of the Louisiana Purchase from which the Missouri Compromise of 1820 had forever excluded slavery.

The Kansas–Nebraska bill

Douglas, who was unconcerned over the moral issue of slavery and desirous of getting on with the settling of the West and the construction of a transcontinental railroad, knew that the Southern senators would block the organization of Kansas as a free territory. The Southerners, recognizing that the North and West had outstripped their section in population and hence in the House of Representatives, clung desperately to an equality of votes in the Senate and were not disposed to welcome any new free territories, which would inevitably become additional free states. Accordingly, Douglas thought that the doctrine of popular sovereignty, which had been applied to the territories gained from Mexico, would avoid a political contest over the Kansas territory: it would permit Southern slaveholders to move into the area, but, since the region was unsuited for plantation slavery, it would inevitably result in the formation of additional free states. His bill therefore allowed the inhabitants of the territory self-government in all matters of domestic importance, including the slavery issue. This provision in effect allowed the territorial legislatures to mandate slavery in their areas and was directly contrary to the Missouri Compromise. With the backing of President Franklin Pierce (served 1853–57), Douglas bullied, wheedled, and bluffed congressmen into passing his bill.



Compromises over extension of slavery into the territories.
By courtesy of Carnegie Institution

Polarization over slavery. Northern sensibilities were outraged. Disliking slavery, Northerners had made few efforts to change the South's "peculiar institution" so long as the republic was loosely articulated. (Indeed, when William Lloyd Garrison began his *Liberator* in 1831, urging the immediate and unconditional emancipation of all slaves, he had only a tiny following; and a few years later he had actually been mobbed in Boston.) But with the sections, perforce, being drawn closely together, Northerners could no longer profess indifference to the South and its institutions. Sectional differences, centring on the issue of slavery, began to appear in every American institution. During the 1840s the major national religious denominations, such as the Methodists and the Presbyterians, split over the slavery question. The Whig Party, which had once allied the conservative businessmen of the North and West with the planters of the South, divided and virtually disappeared after the election of 1852. When Douglas' bill opened up to slavery Kansas and Nebraska—land that had long been reserved for the westward expansion of the free states—Northerners began to organize into an antislavery political party, called in some states the Anti-Nebraska Democratic Party, in others the People's Party, but in most places, the Republican Party.

Events of 1855 and 1856 further exacerbated relations

between the sections and strengthened this new party. Kansas, once organized by Congress, became the field of battle between the free and the slave states in a contest in which concern over slavery was mixed with land speculation and office seeking. A virtual civil war broke out, with rival free- and slave-state legislatures both claiming legitimacy. Disputes between individual settlers sometimes erupted into violence. A proslavery mob sacked the town of Lawrence, an antislavery stronghold, on May 21, 1856. On May 24–25 John Brown, a free-state partisan, led a small party in a raid upon some proslavery settlers on Pottawatomie Creek, murdered five men in cold blood, and left their gashed and mutilated bodies as a warning to the slaveholders. Not even the U.S. Capitol was safe from the violence. On May 22 a South Carolina congressman brutally attacked Senator Charles Sumner of Massachusetts at his desk in the Senate chamber because he had presumably insulted the Carolinian's "honour" in a speech he had given in support of Kansas abolitionists. The 1856 presidential election made it clear that voting was becoming polarized along sectional lines. Though James Buchanan, the Democratic nominee, was elected, John C. Frémont, the Republican candidate, received a majority of the votes in the free states.

The following year the Supreme Court of the United States tried to solve the sectional conflicts that had baffled both the Congress and the president. Hearing the case of Dred Scott, a Missouri slave who claimed freedom on the ground that his master had taken him to live in free territory, the majority of the court, headed by Chief Justice Roger B. Taney, found that Negroes were not citizens of the United States and that Scott hence had no right to bring suit before the court. Taney also concluded that the U.S. laws prohibiting slavery in the territory were unconstitutional. Two Northern antislavery judges on the court bitterly attacked Taney's logic and his conclusions. Acclaimed in the South, the Dred Scott decision was condemned and repudiated throughout the North.

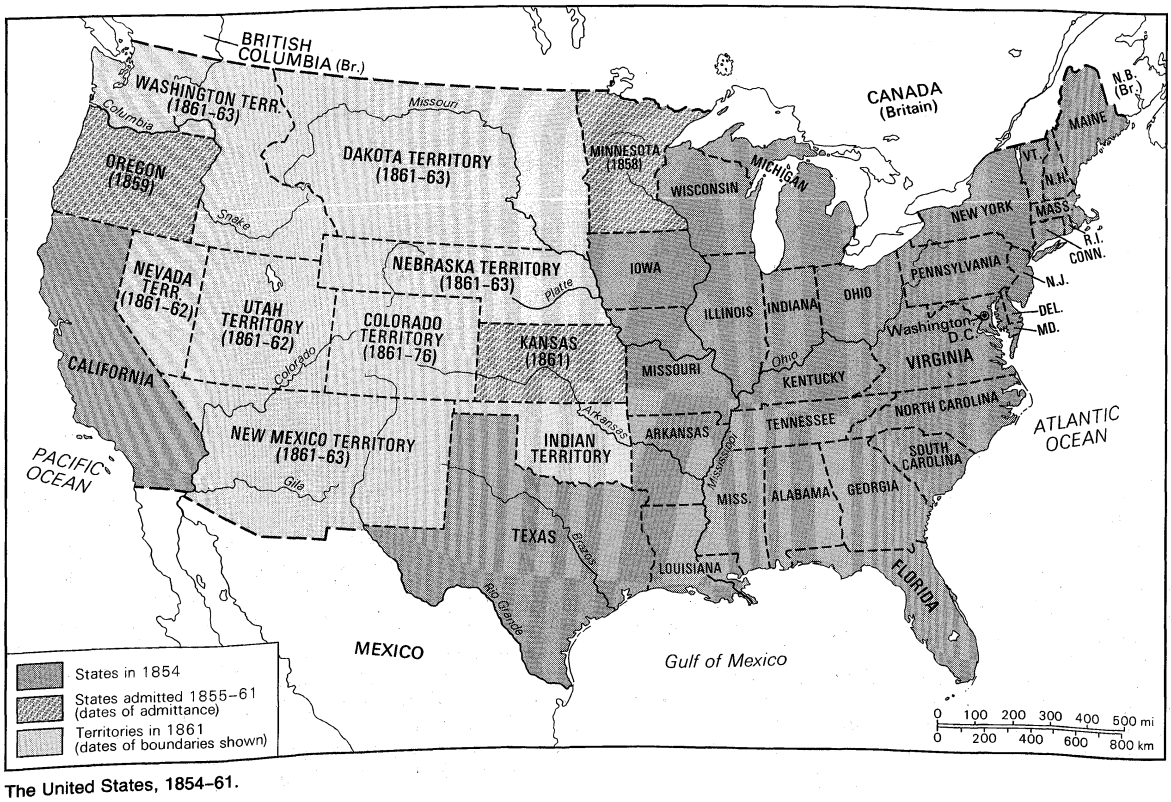
The Dred Scott case

By this point many Americans, North and South, had come to the conclusion that slavery and freedom could not much longer coexist in the United States. For Southerners the answer was withdrawal from a Union that no longer protected their rights and interests; they had talked of it as early as the Nashville Convention of 1850, when the compromise measures were under consideration, and now more and more Southerners favoured secession. For Northerners the remedy was to change the social institutions of the South; few advocated immediate or complete emancipation of the slaves, but many felt that the South's "peculiar institution" must be contained. In 1858 William H. Seward, the leading Republican of New York, spoke of an "irrepressible conflict" between freedom and slavery; and in Illinois a rising Republican politician, Abraham Lincoln, who unsuccessfully contested Douglas for a seat in the Senate, announced that "this government cannot endure, permanently half *slave* and half *free*."

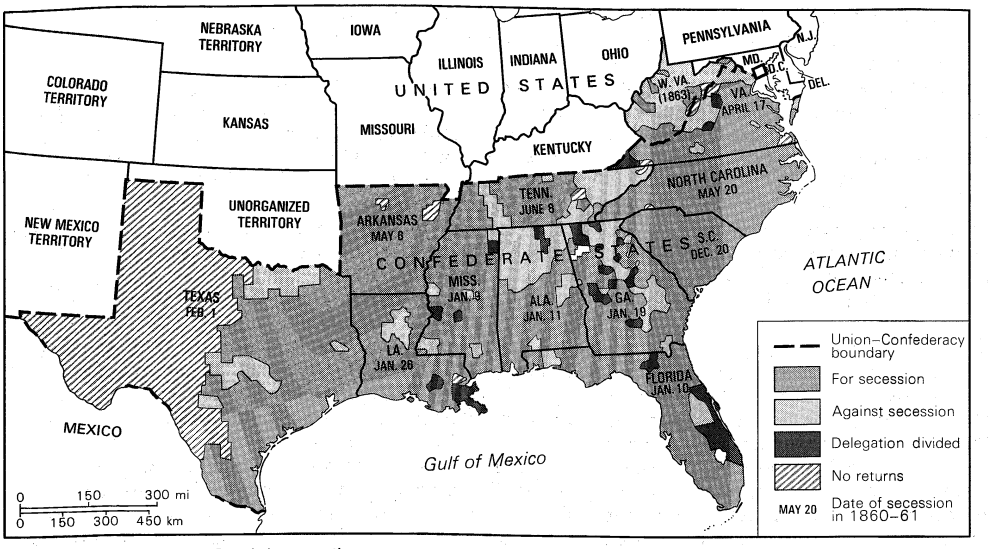
That it was not possible to end the agitation over slavery became further apparent in 1859 when on the night of October 16, John Brown, who had escaped punishment for the Pottawatomie massacre, staged a raid on Harpers Ferry, Va. (now in West Virginia), designed to free the slaves and, apparently, to help them begin a guerrilla war against the Southern whites. Even though Brown was promptly captured and Virginia slaves gave no heed to his appeals, Southerners feared that this was the beginning of organized Northern efforts to undermine their social system. The fact that Brown was a fanatic and an inept strategist whose actions were considered questionable even by abolitionists did not lessen Northern admiration for him.

The presidential election of 1860 occurred, therefore, in an atmosphere of great tension. Southerners, determined that their rights should be guaranteed by law, insisted upon a Democratic candidate willing to protect slavery in the territories; and they rejected Stephen A. Douglas, whose popular-sovereignty doctrine left the question in doubt, in favour of John C. Breckinridge. Douglas, backed by most of the Northern and border-state Democrats, ran on a separate Democratic ticket. Elderly conservatives, who deplored all agitation of the sectional questions but advanced no solutions, offered John Bell as candidate of the Constitutional Union Party. Republicans, confident of success, passed over the claims of Seward, who had accumulated too many liabilities in his long public career, and nominated Lincoln instead. Voting in the subsequent election was along markedly sectional patterns, with Republican strength confined almost completely to the North and West. Though Lincoln received only a plurality of the popular vote, he was an easy winner in the electoral college.

Election of 1860



The United States, 1854–61.



Vote on secession in the South by counties.

Adapted from R. Hofstadter, W. Miller, and D. Aaron, *The American Republic*, vol. 1, to 1865 (© 1959); by permission of Prentice-Hall, Inc.

SECESSION AND THE POLITICS OF THE CIVIL WAR, 1860–65

The coming of the war. In the South, Lincoln's election was taken as the signal for secession, and on December 20 South Carolina became the first state to withdraw from the Union. Promptly the other states of the lower South followed. Feeble efforts on the part of Buchanan's administration to check secession failed, and one by one most of the federal forts in the Southern states were taken over by secessionists. Meanwhile, strenuous efforts in Washington to work out another compromise failed. (The most promising plan was John J. Crittenden's proposal to extend the Missouri Compromise line, dividing free from slave states, to the Pacific.)

Neither extreme Southerners, now intent upon secession, nor Republicans, intent upon reaping the rewards of their hard-won election victory, were really interested in compromise. On Feb. 4, 1861—a month before Lincoln could be inaugurated in Washington—six Southern states (South Carolina, Georgia, Alabama, Florida, Mississippi, Louisiana) sent representatives to Montgomery, Ala., to set up a new independent government. Delegates from Texas soon joined them. With Jefferson Davis of Mississippi at its head, the Confederate States of America came into being, set up its own bureaus and offices, issued its own money, raised its own taxes, and flew its own flag. Not until May 1861, after hostilities had broken out and Virginia had seceded, did the new government transfer its capital to Richmond.

Formation of the Confederacy

Faced with a fait accompli, Lincoln when inaugurated was prepared to conciliate the South in every way but one: he would not recognize that the Union could be divided. The test of his determination came early in his administration, when he learned that the Federal troops under Major Robert Anderson in Fort Sumter, S.C.—then one of the few military installations in the South still in Federal hands—had to be promptly supplied or withdrawn. After agonized consultation with his Cabinet, Lincoln determined that supplies must be sent even if doing so provoked the Confederates into firing the first shot. On April 12, 1861, just before Federal supply ships could reach the beleaguered Anderson, Confederate guns in Charleston opened fire upon Fort Sumter, and the war began.

The political course of the war. For the next four years the Union and the Confederacy were locked in conflict—by far the most titanic waged in the Western Hemisphere.

The policies pursued by the governments of Abraham Lincoln and Jefferson Davis were astonishingly similar. Both presidents at first relied upon volunteers to man the armies, and both administrations were poorly prepared to arm and equip the hordes of young men who flocked to the colours in the initial stages of the war. As the fighting progressed, both governments reluctantly resorted to conscription—the Confederates first, in early 1862, and the Federal government more slowly, with an ineffective measure of late 1862 followed by a more stringent law in 1863. Both governments pursued an essentially laissez-faire policy in economic matters, with little effort to control prices, wages, or profits. Only the railroads were subject to close government regulation in both regions; and the Confederacy, in constructing some of its own powder mills, made a few experiments in "state socialism." Neither Lincoln's nor Davis' administration knew how to cope with financing the war; neither developed an effective system of taxation until late in the conflict, and both relied heavily upon borrowing. Faced with a shortage of funds, both governments were obliged to turn to the printing press and to issue fiat money; the U.S. government issued $432,000,000 in "greenbacks" (as this irredeemable, non-interest-bearing paper money was called), while the Confederacy printed over $1,554,000,000 in such paper currency. In consequence, both sections experienced runaway inflation, which was much more drastic in the South, where, by the end of the war, flour sold at $1,000 a barrel.

Manning and financing the armies

Even toward slavery, the root cause of the war, the policies of the two warring governments were surprisingly similar. The Confederate constitution, which was in most other ways similar to that of the United States, expressly guaranteed the institution of Negro slavery. Despite pressure from abolitionists, Lincoln's administration was not disposed to disturb the "peculiar institution," if only because any move toward emancipation would upset the loyalty of Delaware, Maryland, Kentucky, and Missouri—the four slave states that remained in the Union.

Moves toward emancipation. Gradually, however, under the pressure of war, both governments moved to end slavery. Lincoln came to see that emancipation of the blacks would favourably influence European opinion toward the Northern cause, would deprive the Confederates of their productive labour force on the farms, and would add much-needed recruits to the Federal armies. In September 1862 he issued his preliminary proclamation of emancipation, promising to free all slaves in rebel territory by Jan. 1, 1863, unless those states returned to the Union; and when the Confederates remained obdurate, he followed it with his promised final proclamation. A natural accompaniment of emancipation was the use of black troops, and by the end of the war the number of blacks who served in the Federal armies totaled 178,895. Uncertain of the constitutionality of his Emancipation Proclamation, Lincoln urged Congress to abolish slavery by constitutional amendment; but this was not done until Jan. 31, 1865, and the actual ratification did not take place until after the war.

Lincoln's Emancipation Proclamation

Meanwhile the Confederacy, though much more slowly, was also inevitably drifting in the direction of emancipa-

tion. The South's desperate need for troops caused many military men, including Robert E. Lee, to demand the recruitment of blacks; finally, in March 1865 the Confederate congress authorized the raising of Negro regiments. Though a few blacks were recruited for the Confederate armies, none actually served in battle because surrender was at hand. In yet another way Davis' government showed its awareness of slavery's inevitable end when, in a belated diplomatic mission to seek assistance from Europe, the Confederacy in March 1865 promised to emancipate the slaves in return for diplomatic recognition. Nothing came of the proposal, but it is further evidence that by the end of the war both North and South realized that slavery was doomed.

Sectional dissatisfaction. As war leaders, both Lincoln and Davis came under severe attack in their own sections. Both had to face problems of disloyalty. In Lincoln's case, the Irish immigrants to the eastern cities and the Southern-born settlers of the northwestern states were especially hostile to the Negro and, therefore, to emancipation, while many other Northerners became tired and disaffected as the war dragged on interminably. Residents of the Southern hill country, where slavery never had much of a foothold, were similarly hostile toward Davis. Furthermore, in order to wage war, both presidents had to strengthen the powers of central government, thus further accelerating the process of national integration that had brought on the war. Both administrations were, in consequence, vigorously attacked by state governors, who resented the encroachment upon their authority and who strongly favoured local autonomy.

Disloyalty to both presidents

The extent of Northern dissatisfaction was indicated in the congressional elections of 1862, when Lincoln and his party sustained a severe rebuff at the polls and the Republican majority in the House of Representatives was drastically reduced. Similarly in the Confederacy the congressional elections of 1863 went so strongly against the administration that Davis was able to command a majority for his measures only through the continued support of representatives and senators from the states of the upper South, which were under control of the Federal army and consequently unable to hold new elections.

As late as August 1864, Lincoln despaired of his reelection to the presidency and fully expected that the Democratic candidate, General George B. McClellan, would defeat him. Davis, at about the same time, was openly attacked by Alexander H. Stephens, the vice president of the Confederacy. But Federal military victories, especially William T. Sherman's capture of Atlanta, greatly strengthened Lincoln; and, as the war came to a triumphant close for the North, he attained new heights of popularity. Davis' administration, on the other hand, lost support with each successive defeat, and in January 1865 the Confederate congress insisted that Davis make Robert E. Lee the supreme commander of all Southern forces. (Some, it is clear, would have preferred to make the general dictator.)

(D.H.D.)

THE MILITARY BACKGROUND OF THE WAR

Comparison of North and South. At first glance it seemed that the 23 states of the Union were more than a match for the 11 seceding Southern states—South Carolina, Mississippi, Florida, Alabama, Georgia, Louisiana, Texas, Virginia, Arkansas, Tennessee, and North Carolina. There were approximately 21,000,000 people in the North compared with some 9,000,000 in the South (of whom about 3,500,000 were Negro slaves). In addition, the Federals possessed over 100,000 manufacturing plants as against 18,000 south of the Potomac River, and more than 70 percent of the railroads were in the North. Furthermore, the Union had at its command a 30-to-1 superiority in arms production, a 2-to-1 edge in available manpower, and a great preponderance in commercial and financial resources. It had a functioning government and a small but efficient regular army and navy.

The Confederacy was not predestined to defeat, however. The Southern armies had the advantage of fighting on interior lines, and their military tradition had bulked large in the history of the United States before 1860. Moreover, the long Confederate coastline of 3,500 miles (5,600 kilometres) seemed to defy blockade; and the Confederate president, Jefferson Davis, hoped to receive decisive foreign aid and intervention. Finally, the gray-clad Southern soldiers were fighting for the intangible but strong objectives of home and white supremacy. So the Southern cause was not a lost one; indeed, other nations had won independence against equally heavy odds.

The high commands. Command problems plagued both sides. Of the two rival commanders in chief, most people in 1861 thought Davis to be abler than Lincoln. Davis was a West Point graduate, a hero of the Mexican War, a capable secretary of war under President Franklin Pierce, and a U.S. representative and senator from Mississippi; whereas Lincoln—who had served in the Illinois state legislature and as an undistinguished one-term member of the U.S. House of Representatives—could boast of only a brief period of military service in the Black Hawk War, in which he did not perform well.

As president and commander in chief of the Confederate forces, Davis revealed many fine qualities, including patience, courage, dignity, restraint, firmness, energy, determination, and honesty; but he was flawed by his excessive pride, hypersensitivity to criticism, and inability to delegate minor details to his subordinates. To a large extent Davis was his own secretary of war, although five different men served in that post during the lifetime of the Confederacy. Davis himself also filled the position of general in chief of the Confederate armies until he named Lee to that position on Feb. 6, 1865, when the Confederacy was near collapse. In naval affairs—an area about which he knew little—the Confederate president seldom intervened directly, allowing the competent secretary of the navy, Stephen Mallory, to handle the Southern naval buildup and operations on the water. Although his position was onerous and perhaps could not have been filled so well by any other Southern political leader, Davis' overall performance in office left something to be desired.

Performances of Davis and Lincoln

To the astonishment of many, Lincoln grew in stature with time and experience, and by 1864 he had become a consummate war director. But he had much to learn at first, especially in strategic and tactical matters and in his choices of army commanders. With an ineffective first secretary of war—Simon Cameron—Lincoln unhesitatingly insinuated himself directly into the planning of military movements. Edwin M. Stanton, appointed to the secretaryship on Jan. 20, 1862, was equally untutored in military affairs, but he was fully as active a participant as his superior.

Winfield Scott was the Federal general in chief when Lincoln took office. The 75-year-old Scott—a hero of the War of 1812 and of the Mexican War—was a magnificent and distinguished soldier whose mind was still keen, but he was physically incapacitated and had to be retired from the service on Nov. 1, 1861. Scott was replaced by young George B. McClellan, an able and imaginative general in chief but one who had difficulty in establishing harmonious and effective relations with Lincoln. Because of this and because he had to campaign with his own Army of the Potomac, McClellan was relieved as general in chief on March 11, 1862. He was eventually succeeded on July 11 by the limited Henry W. Halleck, who held the position until replaced by Ulysses S. Grant on March 9, 1864. Halleck then became chief of staff under Grant in a long-needed streamlining of the Federal high command. Grant served efficaciously as general in chief throughout the remainder of the war.

After the initial call by Lincoln and Davis for troops and as the war lengthened indeterminately, both sides turned to raising massive armies of volunteers. Local citizens of prominence and means would organize regiments that were uniformed and accoutred at first under the aegis of the states and then mustered into the service of the Union and Confederate governments. As the war dragged on, the two governments had to resort to conscription to fill the ranks being so swiftly thinned by battle casualties.

Recruitment of troops

Strategic plans. In the area of grand strategy, Davis persistently adhered to the defensive, permitting only occasional "spoiling" forays into Northern territory. Yet per-

haps the Confederates' best chance of winning would have been an early grand offensive into the Union states before the Lincoln administration could find its ablest generals and bring the preponderant resources of the North to bear against the South.

Lincoln, on the other hand, in order to crush the rebellion and reestablish the authority of the Federal government, had to direct his blue-clad armies to invade, capture, and hold most of the vital areas of the Confederacy. His grand strategy was based on Scott's so-called Anaconda plan, a design that evolved from strategic ideas discussed in messages between Scott and McClellan on April 27, May 3, and May 21, 1861. It called for a Union blockade of the Confederacy's littoral as well as a decisive thrust down the Mississippi River and an ensuing strangulation of the South by Federal land and naval forces. But it was to take four years of grim, unrelenting warfare and enormous casualties and devastation before the Confederates could be defeated and the Union preserved.

THE LAND WAR

The war in 1861. The first military operations took place in northwestern Virginia, where nonslaveholding pro-Unionists sought to secede from the Confederacy. McClellan, in command of Federal forces in southern Ohio, advanced on his own initiative in the early summer of 1861 into western Virginia with about 20,000 men. He encountered smaller forces sent there by Lee, then in Richmond in command of all Virginia troops. Although showing signs of occasional hesitation, McClellan quickly won three small but significant battles: at Philippi on June 3, at Rich Mountain on July 11, and at Carrick's (or Corrick's) Ford on July 13. McClellan's casualties were light, and his victories went far toward eliminating Confederate resistance in northwestern Virginia, which had refused to recognize secession, and paving the way for the admittance into the Union of the new state of West Virginia in 1863.

Meanwhile, sizable armies were gathering around the Federal capital of Washington, D.C., and the Confederate capital of Richmond, Va. Federal forces abandoned Harpers Ferry on April 18, and it was quickly occupied by Southern forces, who held it for a time. The Federal naval base at Norfolk was prematurely abandoned to the enemy on April 20. On May 6 Lee ordered a Confederate force—soon to be commanded by P.G.T. Beauregard—northward to hold the rail hub of Manassas Junction, some 26 miles (42 kilometres) southwest of Washington. With Lincoln's approval, Scott appointed Irvin McDowell to command the main Federal army, being hastily collected near Washington. But political pressure and Northern public opinion impelled Lincoln, against Scott's advice, to order McDowell's still-untrained army forward to push the enemy back from Manassas. Meanwhile, Federal forces were to hold Confederate soldiers under Joseph E. Johnston in the Shenandoah valley near Winchester, thus preventing them from reinforcing Beauregard along the Bull Run near Manassas.

McDowell advanced from Washington on July 16 with some 32,000 men and moved slowly toward Bull Run. Two days later a reconnaissance in force was repulsed by the Confederates at Mitchell's and Blackburn's Fords, and

First Battle of Bull Run

when McDowell attacked on July 21 in the First Battle of Bull Run (in the South, First Manassas), he discovered that Johnston had escaped the Federals in the valley and had joined Beauregard near Manassas just in time, bringing the total Confederate force to around 28,000. McDowell's sharp attacks with green troops forced the equally untrained Southerners back a bit, but a strong defensive stand by Thomas Jonathan Jackson (who thereby gained the nickname "Stonewall") enabled the Confederates to check and finally throw back the Federals in the afternoon. The Federal retreat to Washington soon became a rout. McDowell lost 2,708 men—killed, wounded, and missing (including prisoners)—against a Southern loss of 1,981. Both sides now settled down to a long war.

The war in the East in 1862. Fresh from his victories in western Virginia, McClellan was called to Washington to replace Scott. There he began to mold the Army of the Potomac into a resolute, effective shield and sword of the Union. But personality clashes and unrelenting opposition to McClellan from the Radical Republicans in Congress hampered the sometimes tactless, conservative, Democratic general. It took time to drill, discipline, and equip this force of considerably more than 100,000 men, but as fall blended into winter loud demands arose that McClellan advance against Johnston's Confederate forces at Centreville and Manassas. McClellan, however, fell seriously ill with typhoid fever in December, and when he had recovered weeks later he found that Lincoln, desperately eager for action, had ordered him to advance on Feb. 22, 1862. Long debates ensued between president and commander. When in March McClellan finally began his Peninsular Campaign, he discovered that Lincoln and Stanton had withheld large numbers of his command in front of Washington for the defense of the capital—forces that were actually not needed there. Upon taking command of the army in the field, McClellan was relieved of his duties as general in chief.

The Peninsular Campaign. Advancing up the historic peninsula between the York and James rivers, McClellan began a month-long siege of Yorktown and captured that stronghold on May 4, 1862. A Confederate rearguard action at Williamsburg the next day delayed the blue-clads,

Federal threat to Richmond

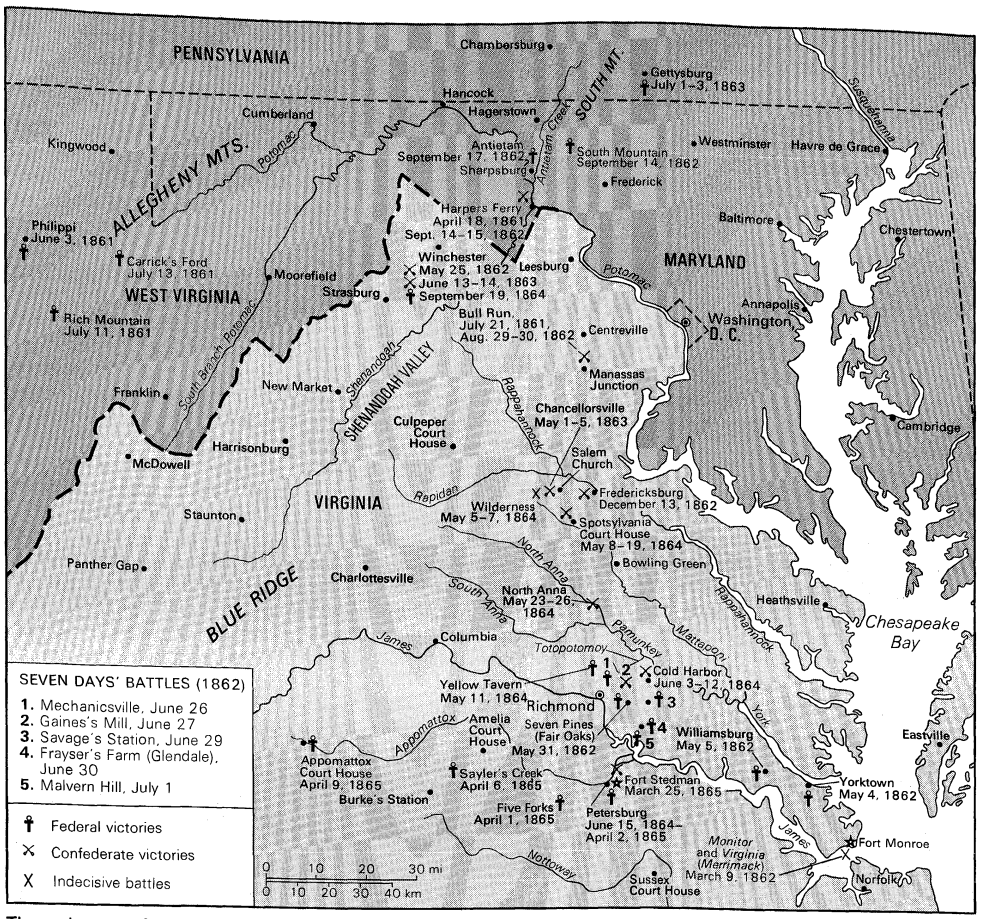
who then slowly moved up through heavy rain to within four miles of Richmond. Striving to seize the initiative, Johnston attacked McClellan's left wing at Seven Pines (Fair Oaks) on May 31 and, after scoring initial gains, was checked; Johnston was severely wounded, and Lee, who had been serving as Davis' military adviser, succeeded Johnston in command of the Army of Northern Virginia. McClellan counterattacked on June 1 and forced the Southerners back into the environs of Richmond. The Federals suffered a total of 5,031 casualties out of a force of nearly 100,000, while the Confederates lost 6,134 of about 74,000 men.

As McClellan inched forward toward Richmond in June, Lee prepared a counterstroke. He recalled from the Shenandoah valley Jackson's forces—which had threatened Harpers Ferry and had brilliantly defeated several scattered Federal armies—and, with about 90,000 soldiers, attacked McClellan on June 26 to begin the fighting of the Seven Days' Battles (usually dated June 25–July 1). In the ensuing days at Mechanicsville, Gaines's Mill, Savage's Station, Frayser's Farm (Glendale), and Malvern Hill, Lee tried unsuccessfully to crush the Army of the Potomac, which McClellan was moving to another base on the James River; but the Confederate chieftain had at least saved Richmond. McClellan inflicted 20,614 casualties on Lee while suffering 15,849 himself. McClellan felt he could not move upon Richmond without considerable reinforcement, and against his protests his army was withdrawn from the peninsula to Washington by Lincoln and the new general in chief, Halleck. Many of McClellan's units were given to a new Federal Army commander, John Pope, who was directed to move overland against Richmond.

Second Battle of Bull Run (Manassas) and Antietam. Pope advanced confidently toward the Rappahannock River with his Army of Virginia, while Lee, once McClellan had been pulled back from near Richmond, moved northward to confront Pope before the latter could be joined by all of McClellan's troops. Daringly splitting his army, Lee sent Jackson to destroy Pope's base at Manassas, while he himself advanced via another route with James Longstreet's half of the army. Pope opened the Second Battle of Bull Run (in the South, Second Manassas) on August 29 with heavy but futile attacks on Jackson. The next day Lee arrived and crushed the Federal left with a massive flank assault by Longstreet, which, combined with Jackson's counterattacks, drove the Northerners back in rout upon Washington. Pope lost 16,054 men out of a force of about 70,000, while Lee lost 9,197 out of about 55,000. With the Federal soldiers now lacking confidence in Pope, Lincoln relieved him and merged his forces into McClellan's Army of the Potomac.

Lee's first invasion of the North

Lee followed up his advantage with his first invasion of the North, pushing as far as Frederick, Md. McClellan had to reorganize his army on the march, a task that he performed capably. But he was beset by contradictory orders:



The main area of the eastern campaigns, 1861–65.

Lincoln urged him to pursue Lee more swiftly; Halleck directed him to slow down and to stay closer to Washington. Biding his time, McClellan pressed forward and wrested the initiative from Lee by attacking and defeating a Confederate force at three gaps of the South Mountain between Frederick and Hagerstown on September 14. Lee fell back into a cramped defensive position along the Antietam Creek, near Sharpsburg, Md., where he was reinforced by Jackson, who had just captured about 11,500 Federals at Harpers Ferry. After a delay, McClellan struck the Confederates on September 17 in the bloodiest single-day's battle of the war. Although gaining some ground, the Federals were unable to drive the Confederate army into the Potomac; but Lee was compelled to retreat back into Virginia. At Antietam, McClellan lost 12,410 of some 69,000 engaged, while Lee lost 13,724 of perhaps 52,000 effectives. When McClellan did not pursue Lee as quickly as Lincoln and Halleck thought he should, he was replaced in command by Ambrose E. Burnside, who had been an ineffective corps commander at Antietam.

Fredericksburg. Burnside delayed for a number of weeks before marching his reinforced army of 120,281 men to a point across the Rappahannock River from Fredericksburg, Va. On December 13 he ordered a series of 16 hopeless, piecemeal, frontal assaults across open ground against Lee's army of 78,513 troops, drawn up in an impregnable position atop high ground and behind a stone wall. The Federals were repelled with staggering losses; Burnside had lost 12,653 men, compared to Lee's 5,309. The plunging Federal morale was reflected in an increasing number of desertions. Therefore, on Jan. 25, 1863, Lincoln replaced Burnside with a proficient corps commander, Joseph ("Fighting Joe") Hooker, who was a harsh critic of other generals and even of the president. Both armies went into winter quarters near Fredericksburg.

The war in the West in 1862. Military events, meanwhile, were transpiring in other arenas.

Trans-Mississippi theatre and Missouri. In the Trans-Mississippi theatre covetous Confederate eyes were cast on California, where ports for privateers could be seized, as could gold and silver to buttress a sagging treasury. Led by Henry Sibley, a Confederate force of some 2,600 invaded the Union's Department of New Mexico, where the Federal commander, Edward Canby, had but 3,810 men to defend the entire vast territory. Although plagued by pneumonia and smallpox, Sibley bettered a Federal force on Feb. 21, 1862, at Valverde and captured Albuquerque and Santa Fe on March 23. But at the crucial engagement of La Glorieta Pass (known also as Apache Canyon, Johnson's Ranch, or Pigeon's Ranch) a few days later, Sibley was checked and lost most of his wagon train. He had to retreat into Texas, where he reached safety in April but with only 900 men and seven of 337 supply wagons left.

Fighting in New Mexico

Farther eastward, in the more vital Mississippi valley, operations were unfolding as large and as important as those on the Atlantic seaboard. Missouri and Kentucky were key border states that Lincoln had to retain within the Union orbit. Commanders there—especially on the Federal side—had greater autonomy than those in Virginia. Affairs began inauspiciously for the Federals in Missouri when Union general Nathaniel Lyon's 5,000 troops were defeated at Wilson's Creek on Aug. 10, 1861, by a Confederate force of more than 10,000 under Sterling Price and Benjamin McCulloch, each side losing some 1,200 men. But the Federals under Samuel Curtis decisively set back a gray-clad army under Earl Van Dorn at Pea Ridge (Elkhorn Tavern), Ark., on March 7–8, 1862, saving Missouri for the Union and threatening Arkansas.

Operations in Kentucky and Tennessee. The Confederates to the east of Missouri had established a unified command under Albert Sidney Johnston, who manned, with only 40,000 men, a long line in Kentucky running from near Cumberland Gap on the east through Bowling Green, to Columbus on the Mississippi. Numerically superior Federal forces cracked this line in early 1862. First, George H. Thomas smashed Johnston's right flank at Mill Springs (Somerset) on January 19. Then, in February, Grant, assisted by Federal gunboats commanded by

Andrew H. Foote and acting under Halleck's orders, ruptured the centre of the Southern line in Kentucky by capturing Fort Henry on the Tennessee River and Fort Donelson, 11 miles (18 kilometres) to the east on the Cumberland River. The Confederates suffered more than 16,000 casualties at the latter stronghold—most of them taken prisoner—as against Federal losses of less than 3,000. Johnston's left anchor fell when Pope seized New Madrid, Mo., and Island Number Ten in the Mississippi in March and April. This forced Johnston to withdraw his remnants quickly from Kentucky through Tennessee and to reorganize them for a counterstroke. This seemingly impossible task he performed splendidly.

The Battle of Shiloh

The Confederate onslaught came at Shiloh, Tenn., near Pittsburg Landing, to which point on the west bank of the Tennessee River Grant and William T. Sherman had incautiously advanced. In a herculean effort, Johnston had pulled his forces together and, with 40,000 men, suddenly struck a like number of unsuspecting Federals on April 6. Johnston hoped to crush Grant before the arrival of Don Carlos Buell's 20,000 Federal troops, approaching from Nashville. A desperate combat ensued, with Confederate assaults driving the Unionists perilously close to the river. But at the height of success, Johnston was mortally wounded; the Southern attack then lost momentum, and Grant held on until reinforced by Buell. On the following day the Federals counterattacked and drove the Confederates, now under Beauregard, steadily from the field, forcing them to fall back to Corinth, in northern Mississippi. Grant's victory cost him 13,047 casualties, compared to Southern losses of 10,694. Halleck then assumed personal command of the combined forces of Grant, Buell, and Pope and inched forward to Corinth, which the Confederates evacuated on May 30.

Beauregard, never popular with Davis, was superseded by Braxton Bragg, one of the president's favourites. Bragg was an imaginative strategist and an effective drillmaster and organizer; but he was also a weak tactician and a martinet who was disliked by a number of his principal subordinates. Leaving 22,000 men in Mississippi under Price and Van Dorn, Bragg moved through Chattanooga with 30,000, hoping to reconquer Tennessee and carry the war into Kentucky. Some 18,000 other Confederate soldiers under Edmund Kirby Smith were at Knoxville. Buell led his Federal force northward to save Louisville and force Bragg to fight. Occupying Frankfort, Bragg failed to move promptly against Louisville. In the ensuing Battle of Perryville on October 8, Bragg, after an early advantage, was halted by Buell and impelled to fall back to a point south of Nashville. Meanwhile, the Federal general William S. Rosecrans had checked Price and Van Dorn at Iuka on September 19 and had repelled their attack on Corinth on October 3–4.

Buell—like McClellan a cautious, conservative, Democratic general—was slow in his pursuit of the retreating Confederates and, despite his success at Perryville, was relieved of his command by Lincoln on October 24. His successor, Rosecrans, was able to safeguard Nashville and then to move southeastward against Bragg's army at Murfreesboro. He scored a partial success by bringing on the bloody Battle of Stones River (or Murfreesboro, Dec. 31, 1862–Jan. 2, 1863). Again, after first having the better of the combat, Bragg was finally contained and forced to retreat. Of some 41,400 men, Rosecrans lost 12,906, while Bragg suffered 11,739 casualties out of about 34,700 effectives. Although it was a strategic victory for Rosecrans, his army was so shaken that he felt unable to advance again for five months, despite the urgings of Lincoln and Halleck.

The war in the East in 1863. In the East, after both armies had spent the winter in camp, the arrival of the active 1863 campaign season was eagerly awaited—especially by Hooker. "Fighting Joe" had capably reorganized and refitted his army, the morale of which was high once again. This massive host numbered around 132,000—the largest formed during the war—and was termed by Hooker "the finest army on the planet." It was opposed by Lee with about 62,000. Hooker decided to move most of his army up the Rappahannock, cross, and come in upon the Confederate rear at Fredericksburg, while John Sedgwick's smaller force would press Lee in front.

The largest army of the war

Chancellorsville. Beginning his turning movement on April 27, 1863, Hooker masterfully swung around toward the west of the Confederate army. Thus far he had outmaneuvered Lee; but Hooker was astonished on May 1 when the Confederate commander suddenly moved the bulk of his army directly against him. "Fighting Joe" lost his nerve and pulled back to Chancellorsville in the Wilderness, where the superior Federal artillery could not be used effectively.

Lee followed up on May 2 by sending Jackson on a brilliant flanking movement against Hooker's exposed right flank. Bursting like a thunderbolt upon Oliver O. Howard's 11th Corps late in the afternoon, Jackson crushed this wing; while continuing his advance, however, Jackson was accidentally wounded by his own men and died of complications shortly thereafter. This helped stall the Confederate advance. Lee then resumed the attack on the morning of May 3 and slowly pushed Hooker back; the latter was wounded by Southern artillery fire. That afternoon Sedgwick drove Jubal Early's Southerners from Marye's Heights at Fredericksburg, but Lee countermarched his weary troops, fell upon Sedgwick at Salem Church, and forced him back to the north bank of the Rappahannock. Lee then returned to Chancellorsville to resume the main engagement; but Hooker, though he had 37,000 fresh troops available, gave up the contest on May 5 and retreated across the river to his old position opposite Fredericksburg. The Federals suffered 17,278 casualties at Chancellorsville, while the Confederates lost 12,764.

Gettysburg. While both armies were licking their wounds and reorganizing, Hooker, Lincoln, and Halleck debated Union strategy. They were thus engaged when Lee launched his second invasion of the North on June 5, 1863. His advance elements moved down the Shenandoah valley toward Harpers Ferry, brushing aside small Federal forces near Winchester. Marching through Maryland into Pennsylvania, the Confederates reached Chambersburg and turned eastward. They occupied York and Carlisle and menaced Harrisburg. Meanwhile, the dashing Confederate cavalryman, J.E.B. ("Jeb") Stuart, set off on a questionable ride around the Federal army and was unable to join Lee's main army until the second day at Gettysburg.

Lee's second invasion of the North

Hooker—on unfriendly terms with Lincoln and especially Halleck—ably moved the Federal forces northward, keeping between Lee's army and Washington. Reaching Frederick, Hooker requested that the nearly 10,000-man Federal garrison at Harpers Ferry be added to his field army. When Halleck refused, Hooker resigned his command and was succeeded by the steady George Gordon Meade, the commander of the 5th Corps. Meade was granted a greater degree of freedom of movement than Hooker had enjoyed, and he carefully felt his way northward, looking for the enemy.

Learning to his surprise on June 28 that the Federal army was north of the Potomac, Lee hastened to concentrate his far-flung legions. Hostile forces came together unexpectedly at the important crossroads town of Gettysburg, in southern Pennsylvania, bringing on the greatest battle ever fought in the Western Hemisphere. Attacking on July 1 from the west and north with 28,000 men, Confederate forces finally prevailed after nine hours of desperate fighting against 18,000 Federal soldiers under John F. Reynolds. When Reynolds was killed, Abner Doubleday ably handled the outnumbered Federal troops, and only the sheer weight of Confederate numbers forced him back through the streets of Gettysburg to strategic Cemetery Ridge south of town, where Meade assembled the rest of the army that night.

On the second day of battle Meade's 93,000 troops were ensconced in a strong, fishhook-shaped defensive position, running northward from the Round Top hills along Cemetery Ridge and thence eastward to Culp's Hill. Lee, with 75,000 troops, ordered Longstreet to attack the Federals diagonally from Little Round Top northward and Richard S. Ewell to assail Cemetery Hill and Culp's Hill. The Confederate attack, coming in the late afternoon and evening, saw Longstreet capture the positions known as the Peach

Orchard, Wheat Field, and Devil's Den on the Federal left in furious fighting but fail to seize the vital Little Round Top. Ewell's later assaults on Cemetery Hill were repulsed, and he could capture only a part of Culp's Hill.

On the morning of the third day, Meade's right wing drove the Confederates from the lower slopes of Culp's Hill and checked Stuart's cavalry sweep to the east of Gettysburg in midafternoon. Then, in what has been called the greatest infantry charge in American history, Lee—against Longstreet's advice—hurled nearly 15,000 soldiers, under the immediate command of George E. Pickett, against the centre of Meade's lines on Cemetery Ridge, following a fearful artillery duel of two hours. Despite heroic efforts, only several hundred Southerners temporarily cracked the Federal centre at the so-called High-Water Mark; the rest were shot down by Federal cannoneers and musketrymen, captured, or thrown back, suffering casualties of almost 60 percent. Meade felt unable to counterattack, and Lee conducted an adroit retreat into Virginia. The Confederates had lost 28,063 men at Gettysburg, the Federals, 23,049. After indecisive maneuvering and light actions in northern Virginia in the fall of 1863, the two armies went into winter quarters. Never again was Lee able to mount a full-scale invasion of the North with his entire army.

Pickett's charge

The war in the West in 1863. *Arkansas and Vicksburg.* In Arkansas, Federal troops under Frederick Steele moved upon the Confederates and defeated them at Prairie Grove, near Fayetteville, on Dec. 7, 1862—a victory that paved the way for Steele's eventual capture of Little Rock the next September.

More importantly, Grant, back in good graces following his undistinguished performance at Shiloh, was authorized to move against the Confederate "Gibraltar of the West"—Vicksburg, Miss. This bastion was difficult to approach: Admiral David G. Farragut, Grant, and Sherman had failed to capture it in 1862. In the early months of 1863, in the so-called Bayou Expeditions, Grant was again frustrated in his efforts to get at Vicksburg from the north. Finally, escorted by Admiral David Dixon Porter's gunboats, which ran the Confederate batteries at Vicksburg, Grant landed his army to the south at Bruinsburg on April 30, 1863, and pressed northeastward. He won small but sharp actions at Port Gibson, Raymond, and Jackson, while the circumspect Confederate defender of Vicksburg, John C. Pemberton, was unable to link up with a smaller Southern force under Joseph E. Johnston near Jackson.

Turning due westward toward the rear of Vicksburg's defenses, Grant smashed Pemberton's army at Champion's Hill and the Big Black River and invested the fortress. During his 47-day siege, Grant eventually had an army of 71,000; Pemberton's command numbered 31,000, of whom 18,500 were effectives. After a courageous stand, the outnumbered Confederates were forced to capitulate on July 4. Five days later, 6,000 Confederates yielded to Nathaniel P. Banks at Port Hudson, La., to the south of Vicksburg, and Lincoln could say, in relief, "The Father of Waters again goes unvexed to the sea."

Grant's siege of Vicksburg

Chickamauga and Chattanooga. Meanwhile, 60,000 Federal soldiers under Rosecrans sought to move southeastward from central Tennessee against the important Confederate rail and industrial centre of Chattanooga, then held by Bragg with some 43,000 troops. In a series of brilliantly conceived movements, Rosecrans maneuvered Bragg out of Chattanooga without having to fight a battle. Bragg was then bolstered by troops from Longstreet's veteran corps, sent swiftly by rail from Lee's army in Virginia. With this reinforcement, Bragg turned on Rosecrans and, in a vicious two-day battle (September 19–20) at Chickamauga Creek, Ga., just southeast of Chattanooga, gained one of the few Confederate victories in the West. Bragg lost 18,454 of his 66,326 men; Rosecrans, 16,170 out of 53,919 engaged. Rosecrans fell back into Chattanooga, where he was almost encircled by Bragg.

But the Southern success was short-lived. Instead of pressing the siege of Chattanooga, Bragg unwisely sent Longstreet off in a futile attempt to capture Knoxville, then being held by Burnside. When Rosecrans showed signs of disintegration, Lincoln replaced him with Grant and strengthened the hard-pressed Federal army at Chattanooga by sending, by rail, the remnants of the Army of the Potomac's 11th and 12th Corps, under Hooker's command. Outnumbering Bragg now 56,359 to 46,165, Grant attacked on November 23–25, capturing Lookout Mountain and Missionary Ridge, defeating Bragg's army, and driving it southward toward Dalton, Ga. Grant sustained 5,824 casualties at Chattanooga and Bragg, 6,667. Confidence having been lost in Bragg by most of his top generals, Davis replaced him with Joseph E. Johnston. Both armies remained quiescent until the following spring.

The war in 1864–65. Finally dissatisfied with Halleck as general in chief and impressed with Grant's victories, Lincoln appointed Grant to supersede Halleck and to assume the rank of lieutenant general, which Congress had re-created. Leaving Sherman in command in the West, Grant arrived in Washington on March 8, 1864. He was given largely a free hand in developing his grand strategy. He retained Meade in technical command of the Army of the Potomac but in effect assumed direct control by establishing his own headquarters with it. He sought to move this army against Lee in northern Virginia while Sherman marched against Johnston and Atlanta. Several lesser Federal armies were also to advance in May.

Grant's appointment as Federal general in chief

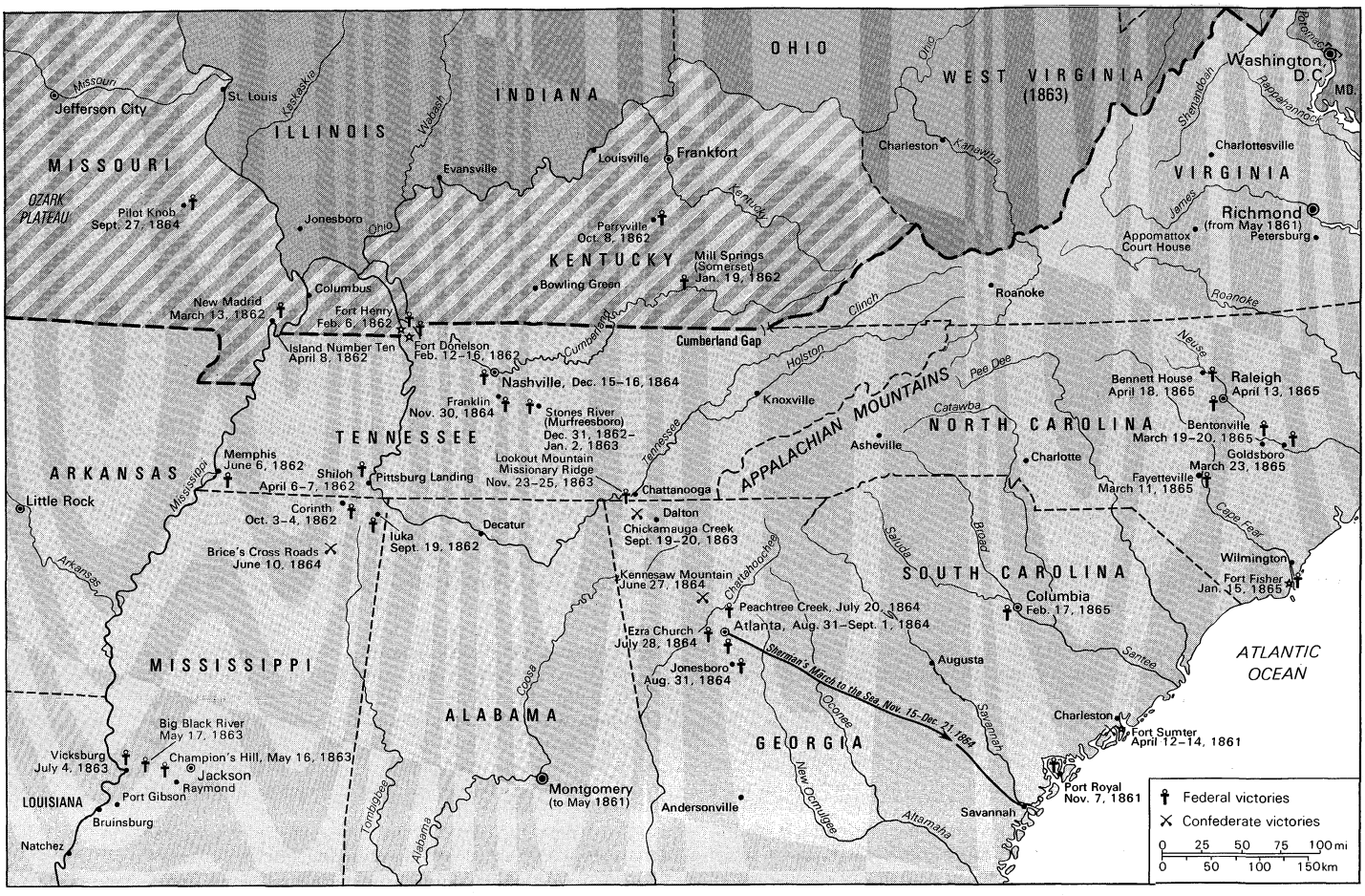
Grant's overland campaign. Grant surged across the Rapidan and Rappahannock rivers on May 4, hoping to get through the tangled Wilderness before Lee could move. But the Confederate leader reacted instantly and, on May 5, attacked Grant from the west in the Battle of the Wilderness. Two days of bitter, indecisive combat ensued. Although Grant had 115,000 men available against Lee's 62,000, he found both Federal flanks endangered. Moreover, Grant lost 17,666 soldiers compared to a probable Southern loss of about 8,000. Pulling away from the Wilderness battlefield, Grant tried to hasten southeastward to the crossroads point of Spotsylvania Court House, only to have the Confederates get there first. In savage action (May 8–19), including hand-to-hand fighting at the famous "Bloody Angle," Grant, although gaining a little ground, was essentially thrown back. He had lost 18,399 men at Spotsylvania. Lee's combined losses at the Wilderness and Spotsylvania were an estimated 17,250.

Again Grant withdrew, only to move forward in another series of attempts to get past Lee's right flank; again, at the North Anna River and at the Totopotomoy Creek, he found Lee confronting him. Finally at Cold Harbor, just northeast of Richmond, Grant launched several heavy attacks, including a frontal, near-suicidal one on June 3, only to be repelled with grievous total losses of 12,737. Lee's casualties are unknown but were much lighter.

Grant, with the vital rail centre of Petersburg—the southern key to Richmond—as his objective, made one final effort to swing around Lee's right and finally outguessed his opponent and stole a march on him. But several blunders by Federal officers, swift action by Beauregard, and Lee's belated though rapid reaction enabled the Confederates to hold Petersburg. Grant attacked on June 15 and 18, hoping to break through before Lee could consolidate the Confederate lines east of the city, but he was contained with 8,150 losses.

Unable to admit defeat but having failed to destroy Lee's army and capture Richmond, Grant settled down to a nine-month active siege of Petersburg. The summer and fall of 1864 were highlighted by the Federal failure with a mine explosion under the gray lines at Petersburg on July 30, the near capture of Washington by the Confederate Jubal Early in July, and Early's later setbacks in the Shenandoah valley at the hands of Philip H. Sheridan.

Sherman's Georgia campaigns. Meanwhile, Sherman was pushing off toward Atlanta from Dalton, Ga., on May 7, 1864, with 110,123 men against Johnston's 55,000. This masterly campaign comprised a series of cat-and-mouse moves by the rival commanders. Nine successive defensive positions were taken up by Johnston. Trying to outguess his opponent, Sherman attempted to swing around the Confederate right flank twice and around the left flank the other times, but each time Johnston divined which way Sherman was moving and each time pulled back in time to thwart him. At one point Sherman's patience snapped and he frontally assaulted the Southerners



The main area of the Western and Carolinas campaigns, 1861–65.

at Kennesaw Mountain on June 27; Johnston threw him back with heavy losses. All the while Sherman's lines of communication in his rear were being menaced by audacious Confederate cavalry raids conducted by Nathan Bedford Forrest and Joseph Wheeler. Forrest administered a crushing defeat to Federal troops under Samuel D. Sturgis at Brice's Cross Roads, Miss., on June 10. But these Confederate forays were more annoying than decisive, and Sherman pressed forward.

When Johnston finally informed Davis that he could not realistically hope to annihilate Sherman's mighty army, the Confederate president replaced him with John B. Hood, who had already lost two limbs in the war. Hood inaugurated a series of premature offensive battles at Peachtree Creek, Atlanta, Ezra Church, and Jonesboro, but he was repulsed in each of them. With his communications threatened, Hood evacuated Atlanta on the night of August 31–September 1. Sherman pursued only at first. Then, on November 15, he commenced his great March to the Sea with more than 60,000 men, laying waste to the economic resources of Georgia in a 50-mile-wide swath of destruction. He captured Savannah on December 21.

Sherman's March to the Sea

Hood had sought unsuccessfully to lure Sherman out of Georgia and back into Tennessee by marching northwestward with nearly 40,000 men toward the key city of Nashville, the defense of which had been entrusted by Sherman to George H. Thomas. At Franklin, Hood was checked for a day with severe casualties by a Federal holding force under John M. Schofield. This helped Thomas to retain Nashville, where, on December 15–16, he delivered a crushing counterstroke against Hood's besieging army, cutting it up so badly that it was of little use thereafter.

Western campaigns. Sherman's force might have been larger and his Atlanta–Savannah Campaign consummated much sooner had not Lincoln approved the Red River Campaign in Louisiana led by Banks in the spring of 1864. Accompanied by Porter's warships, Banks moved up the Red River with some 40,000 men. He had two objectives: to capture cotton and to defeat Southern forces under Kirby Smith and Richard Taylor. Not only did he fail to net much cotton but he was also checked with loss on April 8 at Sabine Cross Roads and forced to retreat. Porter lost several gunboats, and the campaign amounted to a costly debacle.

That fall Kirby Smith ordered the reconquest of Missouri. Sterling Price's Confederate army advanced on a broad front into Missouri but was set back temporarily by Thomas Ewing at Pilot Knob on September 27. Resuming the advance toward St. Louis, Price was forced westward along the south bank of the Missouri River by pursuing Federal troops under A.J. Smith, Alfred Pleasonton, and Samuel Curtis. Finally, on October 23, at Westport, near Kansas City, Price was decisively defeated and forced to retreat along a circuitous route, arriving back in Arkansas on December 2. This ill-fated raid cost Price most of his artillery as well as the greater part of his army, which numbered about 12,000.

Sherman's Carolina campaigns. On Jan. 10, 1865, with Tennessee and Georgia now securely in Federal hands, Sherman's 60,000-man force began to march northward into the Carolinas. It was only lightly opposed by much smaller Confederate forces. Sherman captured Columbia on February 17 and compelled the Confederates to evacuate Charleston (including Fort Sumter). When Lee was finally named Confederate general in chief, he promptly reinstated Johnston as commander of the small forces striving to oppose the Federal advance. Nonetheless, Sherman captured Fayetteville, N.C., on March 11 and, after an initial setback, repulsed the counterattacking Johnston at Bentonville on March 19–20. Goldsboro fell to the Federals on March 23, and Raleigh on April 13. Finally, perceiving that he no longer had any reasonable chance of containing the relentless Federal advance, Johnston surrendered to Sherman at the Bennett House near Durham Station on April 18. When Sherman's generous terms proved unacceptable to Secretary of War Stanton (Lincoln

Lee's appointment as Confederate general in chief

had been assassinated on April 14), the former submitted new terms that Johnston signed on April 26.

The final land operations. Grant and Meade were continuing their siege of Petersburg and Richmond early in 1865. For months the Federals had been lengthening their left (southern) flank while operating against several important railroads supplying the two Confederate cities. This stretched Lee's dwindling forces very thin. The Southern leader briefly threatened to break the siege when he attacked and captured Fort Stedman on March 25. But an immediate Federal counterattack regained the strongpoint, and Lee, when his lines were subsequently pierced, evacuated both Petersburg and Richmond on the night of April 2–3.

An 88-mile (142-kilometre) pursuit west-southwestward along the Appomattox River ensued, with Grant and Meade straining every nerve to bring Lee to bay. The Confederates were detained at Amelia Court House, awaiting delayed food supplies, and were badly cut up at Five Forks and Sayler's Creek, with their only avenue of escape now cut off by Sheridan and George A. Custer. When Lee's final attempt to break out failed, he surrendered the remnants of his gallant Army of Northern Virginia at the McLean house at Appomattox Court House on April 9. The lamp of magnanimity was reflected in Grant's unselfish terms.

Lee's surrender at Appomattox Court House

On the periphery of the Confederacy, 43,000 gray-clad soldiers in Louisiana under Kirby Smith surrendered to Canby on May 26. The port of Galveston, Tex., yielded to the Federals on June 2, and the greatest war on American soil was over.

THE NAVAL WAR

While the Federal armies actually stamped out Confederate land resistance, the increasingly effective Federal naval effort must not be overlooked. If Union sea power did not win the war, it enabled the war to be won. When hostilities opened, the U.S. Navy numbered 90 warships, of which only 42 were in commission, and many of these were on foreign station. Fortunately for the Federals, Lincoln had, in the person of Gideon Welles, a wise secretary of the navy and one of his most competent Cabinet members. Welles was ably seconded by his assistant, Gustavus Vasa Fox.

By the time of Lee's surrender, Lincoln's navy numbered 626 warships, of which 65 were ironclads. From a tiny force of nearly 9,000 seamen in 1861, the Union navy increased by war's end to about 59,000 sailors, whereas naval appropriations per year leaped from approximately $12,000,000 to perhaps $123,000,000. The blockade of about 3,500 miles of Confederate coastline was a factor of incalculable value in the final defeat of the Davis government, although the blockade did not become truly effective before the end of 1863.

The Confederates, on the other hand, had to start from almost nothing in building a navy. That they did so well was largely because of untiring efforts by the capable secretary of the navy, Stephen Mallory. He dispatched agents to Europe to purchase warships, sought to refurbish captured or scuttled Federal vessels, and made every effort to arm and employ Southern-owned ships then in Confederate ports. Mallory's only major omission was his delay in seeing the advantage of Confederate government control of blockade runners bringing in strategic supplies; not until later in the war did the government begin closer supervision of blockade-running vessels. Eventually, the government commandeered space on all privately owned blockade runners and even built and operated some of its own late in the war.

The naval side of the Civil War was a revolutionary one. In addition to their increasing use of steam power, the screw propeller, shell guns, and rifled ordnance, both sides built and employed ironclad warships. The notable clash on March 9, 1862, between the North's *Monitor* and the South's *Virginia* (formerly the *Merrimack*) was the first battle ever waged between ironclads. Also, the first sinking of a warship by a submarine occurred on February 17, 1864, when the Confederate submersible *Hunley* sank the blockader USS *Housatonic.*

Daring Confederate sea raiders preyed upon Union commerce. Especially successful were the *Sumter,* commanded by Raphael Semmes, which captured 18 Northern merchantmen early in the war; the *Florida,* captained by John Maffit, which, in 1863, seized 37 Federal prizes in the North and South Atlantic; and the *Shenandoah,* with James Waddell as skipper, which took 38 Union merchant ships, mostly in the Pacific. But the most famous of all the Confederate cruisers was the *Alabama,* commanded by Semmes, which captured 69 Federal ships in two years; not until June 19, 1864, was the *Alabama* intercepted and sunk off Cherbourg by the Federal warship *Kearsarge,* captained by John Winslow. A great many other Federal ships were captured, and marine insurance rates were driven to a prohibitive high by these Southern depredations. This led to a serious deterioration of the American merchant marine, the effects of which have lasted into the 20th century.

Confederate sea raiders

Besides fighting efficaciously with ironclads on the inland rivers, Lincoln's navy also played an important role in a series of coastal and amphibious operations, some in conjunction with the Federal army. As early as Nov. 7, 1861, a Federal flotilla under Samuel Francis du Pont seized Port Royal, S.C., and another squadron under Louis M. Goldsborough assisted Burnside's army in capturing Roanoke Island and New Bern on the North Carolina littoral in February–March 1862. One month later, Savannah was closed to Confederate blockade runners when the Federal navy reduced Fort Pulaski guarding the city; and on April 25 David Glasgow Farragut, running the forts near the mouth of the Mississippi, took New Orleans, which was subsequently occupied by Benjamin F. Butler's army.

But in April 1863, and again in July and August, Federal warships were repelled at Fort Sumter when they descended upon Charleston, and a Federal army under Quincy A. Gillmore fared little better when it tried to assist. Farragut had better luck, however, when he rendered Mobile, Ala., useless by reducing Fort Morgan and destroying several defending Confederate ships on Aug. 5, 1864, in the hardest-fought naval action of the war. The Confederacy's last open Atlantic port, Wilmington, N.C., successfully withstood a Federal naval attack by Porter on defending Fort Fisher when Butler's army failed to coordinate its attack properly in December 1864, but it fell one month later to Porter and an ably conducted army assault led by Alfred H. Terry. Only Galveston remained open to the Confederates in the last months of the war. In short, "Uncle Sam's web feet," as Lincoln termed the Union navy, played a decisive role in helping crush the Confederacy.

FOREIGN AFFAIRS

Davis and many Confederates expected recognition of their independence and direct intervention in the war on their behalf by Great Britain and possibly France. But they were cruelly disappointed, in part through the skillful diplomacy of Lincoln, Secretary of State Seward, and the Union ambassador to England, Charles Francis Adams, and in part through Confederate military failure at a crucial stage of the war.

The Union's first trouble with Britain came when Captain Charles Wilkes halted the British steamer *Trent* on Nov. 8, 1861, and forcibly removed two Confederate envoys, James M. Mason and John Slidell, bound for Europe. Only the eventual release of the two men prevented a diplomatic rupture with Lord Palmerston's government in London. Another crisis erupted between the Union and England when the *Alabama,* built in the British Isles, was permitted upon completion to sail and join the Confederate navy, despite Adams' protestations. And when word reached the Lincoln government that two powerful ironclad rams were being constructed in Britain for the Confederacy, Adams sent his famous "this is war" note to Palmerston, and the rams were seized by the British government at the last moment.

The diplomatic crisis of the Civil War came after Lee's striking victory at the Second Battle of Bull Run in late August 1862 and subsequent invasion of Maryland. The British government was set to offer mediation of the war

The diplomatic crisis of the war

and, if this were refused by the Lincoln administration (as it would have been), forceful intervention on behalf of the Confederacy. Only a victory by Lee on Northern soil was needed, but he was stopped by McClellan in September at Antietam, the Union's most needed success. The Confederate defeats at Gettysburg and Vicksburg the following summer ensured the continuing neutrality of Britain and France, especially when Russia seemed inclined to favour the Northern cause. Even the growing British shortage of cotton from the Southern states did not force Palmerston's government into Davis' camp, particularly when British consuls in the Confederacy were more closely restricted toward the close of the war. In the final act, even the Confederate offer to abolish slavery in early 1865 in return for British recognition fell on deaf ears.

THE COST AND SIGNIFICANCE OF THE CIVIL WAR

On the positive side, the triumph of the North, above and beyond its superior naval forces, numbers, and industrial and financial resources, was due in part to the statesmanship of Lincoln, who by 1864 had become a masterful war leader; to the pervading valour of Federal soldiers; and to the increasing skill of their officers. On the negative side, the victory can be attributed in part to failures of Confederate transportation, matériel, and political leadership. Only praise can be extended to the continuing bravery of Confederate soldiers and to the strategic and tactical dexterity of such generals as Lee, Jackson, and Joseph E. Johnston.

War casualties

While there were some desertions on both sides, the personal valour and the enormous casualties—both in absolute numbers and in percentage of numbers engaged—have not yet ceased to astound scholars and military historians everywhere. Based on the three-year standard of enlistment, about 1,556,000 soldiers served in the Federal armies, which suffered a total of 634,703 casualties (359,528 dead and 275,175 wounded). There were probably about 800,000 men serving in the Confederate forces, which sustained approximately 483,000 casualties (about 258,000 deaths and perhaps 225,000 wounded).

The cost in treasure was, of course, staggering for the embattled sections. Both governments, after strenuous attempts to finance the prosecution of the war by increasing taxes and floating loans, were obliged to resort to the printing press to make fiat money. While separate Confederate figures are lacking, the war finally cost the United States more than $15,000,000,000. The South, especially, where most of the war was fought and which lost its labour system, was physically and economically devastated. In sum, although the Union was preserved and restored, the cost in physical and moral suffering was incalculable, and some spiritual wounds caused by the holocaust still have not been healed.

The American Civil War has been called by some the last of the old-fashioned wars; others have termed it the first of the modern wars of history. Actually it was a transitional war, and it had a profound impact, technologically, on the development of modern weapons and techniques. There were many innovations. It was the first war in history in which ironclad warships clashed; the first in which the telegraph and railroad played significant roles; the first to use, extensively, rifled ordnance and shell guns and to introduce a machine gun; the first to have widespread newspaper coverage, voting by servicemen in national elections, and photographic recordings; the first to organize medical care of troops systematically; and the first to use land and water mines and to employ a submarine that could sink a warship. It was also the first war in which armies widely employed aerial reconnaissance (by means of balloons).

The Civil War has been written about as have few other wars in history. More than 60,000 books and articles give eloquent testimony to the accuracy of Walt Whitman's prediction that "a great literature will . . . arise out of the era of those four years." The events of the war left a rich heritage for future generations, and that legacy was summed up by the martyred Lincoln as showing that the reunited sections of the United States constituted "the last best hope of earth." (W.W.H.)

Reconstruction and the New South, 1865–1900

RECONSTRUCTION, 1865–77

Northern war aims

Reconstruction under Abraham Lincoln. The original Northern objective in the Civil War was the preservation of the Union—a war aim with which virtually everybody in the free states agreed. As the fighting progressed, the Lincoln government concluded that emancipation of the slaves was necessary in order to secure military victory; and thereafter freedom became a second war aim for the members of the Republican Party. The more radical members of that party—men like Charles Sumner and Thaddeus Stevens—believed that emancipation would prove a sham unless the government guaranteed the civil and political rights of the freedmen; thus, equality of all citizens before the law became a third war aim for this powerful faction. The fierce controversies of the Reconstruction era raged over which of these objectives should be insisted upon and how these goals should be secured.

Lincoln's plan. Lincoln himself had a flexible and pragmatic approach to Reconstruction, insisting only that the Southerners, when defeated, pledge future loyalty to the Union and emancipate their slaves. As the Southern states were subdued, he appointed military governors to supervise their restoration. The most vigorous and effective of these appointees was Andrew Johnson, a War Democrat whose success in reconstituting a loyal government in Tennessee led to his nomination as vice president on the Republican ticket with Lincoln in 1864. In December 1863 Lincoln announced a general plan for the orderly Reconstruction of the Southern states, promising to recognize the government of any state that pledged to support the Constitution and the Union and to emancipate the slaves if it was backed by at least 10 percent of the number of voters in the 1860 presidential election. In Louisiana, Arkansas, and Tennessee loyal governments were formed under Lincoln's plan; and they sought readmission to the Union with the seating of their senators and representatives in Congress.

The Radicals' plan. Radical Republicans were outraged at these procedures, which savoured of executive usurpation of congressional powers, which required only minimal changes in the Southern social system, and which left political power essentially in the hands of the same Southerners who had led their states out of the Union. The Radicals put forth their own plan of Reconstruction in the Wade–Davis Bill, which Congress passed on July 2, 1864; it required not 10 percent but a majority of the white male citizens in each Southern state to participate in the reconstruction process, and it insisted upon an oath of past, not just of future, loyalty. Finding the bill too rigorous and inflexible, Lincoln pocket vetoed it; and the Radicals bitterly denounced him. During the 1864–65 session of Congress, they in turn defeated the president's proposal to recognize the Louisiana government organized under his 10 percent plan. At the time of Lincoln's assassination, therefore, the president and the Congress were at loggerheads over Reconstruction.

Conflict between Lincoln and Congress

Reconstruction under Andrew Johnson. At first it seemed that Johnson might be able to work more cooperatively with Congress in the process of Reconstruction. A former representative and a former senator, he understood congressmen. A loyal Unionist who had stood by his country even at the risk of his life when Tennessee seceded, he was certain not to compromise with secession; and his experience as military governor of that state showed him to be politically shrewd and tough toward the slaveholders. "Johnson, we have faith in you," Radical Benjamin F. Wade assured the new president on the day he took the oath of office. "By the gods, there will be no trouble running the government."

Johnson's policy. Such Radical trust in Johnson proved misplaced. The new president was, first of all, himself a Southerner. He was a Democrat who looked for the restoration of his old party partly as a step toward his own reelection to the presidency in 1868. Most important of all, Johnson shared the white Southerners' attitude toward the Negro, considering black men innately inferior and unready for equal civil or political rights. On May

29, 1865, Johnson made his policy clear when he issued a general proclamation of pardon and amnesty for most Confederates and authorized the provisional governor of North Carolina to proceed with the reorganization of that state. Shortly afterward he issued similar proclamations for the other former Confederate states. In each case a state constitutional convention was to be chosen by the voters who pledged future loyalty to the U.S. Constitution. The conventions were expected to repeal the ordinances of secession, to repudiate the Confederate debt, and to accept the Thirteenth Amendment, abolishing slavery. The president did not, however, require them to enfranchise the blacks.

"Black Codes." Given little guidance from Washington, Southern whites turned to the traditional political leaders of their section for guidance in reorganizing their governments; and the new regimes in the South were suspiciously like those of the antebellum period. To be sure, slavery was abolished; but each reconstructed Southern state government proceeded to adopt a "Black Code," regulating the rights and privileges of freedmen. Varying from state to state, these codes in general treated blacks as inferiors, relegated to a secondary and subordinate position in society. Their right to own land was restricted, they could not bear arms, and they might be bound out in servitude for vagrancy and other offenses. The conduct of white Southerners indicated that they were not prepared to guarantee even minimal protection of Negro rights. In riots in Memphis (May 1866) and New Orleans (July 1866), black persons were brutally assaulted and promiscuously killed.

Civil rights legislation. Watching these developments with forebodings, Northern Republicans during the congressional session of 1865–66 inevitably drifted into conflict with the president. Congress attempted to protect the rights of blacks by extending the life of the Freedmen's Bureau, a welfare agency established in March 1865 to ease the transition from slavery to freedom; but Johnson vetoed the bill. An act to define and guarantee the blacks' basic civil rights met a similar fate, but Republicans succeeded in passing it over the president's veto. While the president, from the porch of the White House, denounced the leaders of the Republican Party as "traitors," Republicans in Congress tried to formulate their own plan to reconstruct the South. Their first effort was the passage of the Fourteenth Amendment, which guaranteed the basic civil rights of all citizens, regardless of colour, and which tried to persuade the Southern states to enfranchise blacks by threatening to reduce their representation in Congress.

Conflict between Republicans and Johnson

The president, the Northern Democrats, and the Southern whites spurned this Republican plan of Reconstruction. Johnson tried to organize his own political party in the National Union Convention, which met in Philadelphia in August 1866; and in August and September he visited many Northern and Western cities in order to defend his policies and to attack the Republican leaders. At the president's urging, every Southern state except Tennessee overwhelmingly rejected the Fourteenth Amendment.

Victorious in the fall elections, congressional Republicans moved during the 1866–67 session to devise a second, more stringent program for reconstructing the South. After long and acrimonious quarrels between Radical and moderate Republicans, the party leaders finally produced a compromise plan in the First Reconstruction Act of 1867. Expanded and clarified in three supplementary Reconstruction acts, this legislation swept away the regimes the president had set up in the South, put the former Confederacy back under military control, called for the election of new constitutional conventions, and required the constitutions adopted by these bodies to include both Negro suffrage and the disqualification of former Confederate leaders from officeholding. Under this legislation, new governments were established in all the former Confederate states (except Tennessee, which had already been readmitted); and by July 1868 Congress agreed to seat senators and representatives from Alabama, Arkansas, Florida, Louisiana, North Carolina, and South Carolina. By July 1870 the remaining Southern states had been similarly reorganized and readmitted.

Suspicious of Andrew Johnson, Republicans in Congress did not trust the president to enforce the Reconstruction legislation they passed over his repeated vetoes, and they tried to deprive him of as much power as possible. Congress limited the president's control over the army by requiring that all his military orders be issued through the general of the army, Ulysses S. Grant, who was believed loyal to the Radical cause; and in the Tenure of Office Act (1867) they limited the president's right to remove appointive officers. When Johnson continued to do all he could to block the enforcement of Radical legislation in the South, the more extreme members of the Republican Party demanded his impeachment. The president's decision in February 1868 to remove the Radical secretary of war Edwin M. Stanton from the Cabinet, in apparent defiance of the Tenure of Office Act, provided a pretext for impeachment proceedings. The House of Representatives voted to impeach the president, and after a protracted trial the Senate acquitted him by the margin of only one vote.

Impeachment of Johnson

The South during Reconstruction. In the South the Reconstruction period was a time of readjustment accompanied by disorder. Southern whites wished to keep blacks in a condition of quasi-servitude, extending few civil rights and firmly rejecting social equality. Blacks, on the other hand, wanted full freedom and, above all, land of their own. Inevitably, there were frequent clashes. Some erupted into race riots, but acts of terrorism against individual black leaders were more common.

During this turmoil, Southern whites and blacks began to work out ways of getting their farms back into operation and of making a living. Indeed, the most important developments of the Reconstruction era were not the highly publicized political contests but the slow, almost imperceptible changes that occurred in Southern society. Blacks could now legally marry, and they set up conventional and usually stable family units; they quietly seceded from the white churches and formed their own religious organizations, which became centres for the black community. Without land or money, most freedmen had to continue working for white masters; but they were now unwilling to labour in gangs or to live in the old slave quarters under the eye of the plantation owner.

Sharecropping gradually became the accepted labour system in most of the South—planters, short of capital, favoured the system because it did not require them to pay cash wages; blacks preferred it because they could live in individual cabins on the tracts they rented and because they had a degree of independence in choosing what to plant and how to cultivate. The section as a whole, however, was desperately poor throughout the Reconstruction era; and a series of disastrously bad crops in the late 1860s, followed by the general agricultural depression of the 1870s, hurt both whites and blacks.

The governments set up in the Southern states under the congressional program of Reconstruction were, contrary to traditional clichés, fairly honest and effective. Though the period has sometimes been labeled "Black Reconstruction," the Radical governments in the South were never dominated by blacks. There were no black governors, only two black senators and a handful of congressmen, and only one legislature controlled by blacks. Those blacks who did hold office appear to have been about equal in competence and honesty to the whites. It is true that these Radical governments were expensive, but large state expenditures were necessary to rebuild after the war and to establish—for the first time in most Southern states—a system of common schools. Corruption there certainly was, though nowhere on the scale of the Tweed Ring, which at that time was busily looting New York City; but it is not possible to show that Republicans were more guilty than Democrats, or blacks than whites, in the scandals that did occur.

Southern Reconstruction governments

Though some Southern whites in the mountainous regions and some planters in the rich bottomlands were willing to cooperate with the blacks and their Northern-born "carpetbagger" allies in these new governments, there were relatively few such "scalawags"; the mass of Southern whites remained fiercely opposed to Negro political, civil, and social equality. Sometimes their hostility was expressed through such terrorist organizations as the

Ku Klux Klan, which sought to punish so-called uppity Negroes and to drive their white collaborators from the South. More frequently it was manifested through support of the Democratic Party, which gradually regained its strength in the South and waited for the time when the North would tire of supporting the Radical regimes and would withdraw federal troops from the South.

The Ulysses S. Grant administrations, 1869–77. During the two administrations of President Grant there was a gradual attrition of Republican strength. As a politician the president was passive, exhibiting none of the brilliance he had shown on the battlefield. His administration was tarnished by the dishonesty of his subordinates, whom he loyally defended. As the older Radical leaders—men like Sumner, Wade, and Stevens—died, leadership in the Republican Party fell into the hands of technicians like Roscoe Conkling and James G. Blaine, men devoid of the idealistic fervour that had marked the early Republicans. At the same time, many Northerners were growing tired of the whole Reconstruction issue and were weary of the annual outbreaks of violence in the South that required repeated use of federal force.

Weakening of the Radicals

Efforts to shore up the Radical regimes in the South grew increasingly unsuccessful. The adoption of the Fifteenth Amendment (1870), prohibiting discrimination in voting on account of race, had little effect in the South, where terrorist organizations and economic pressure from planters kept blacks from the polls. Nor were three Force Acts passed by the Republicans (1870–71), giving the president the power to suspend the writ of habeas corpus and imposing heavy penalties upon terroristic organizations, in the long run more successful. If they succeeded in dispersing the Ku Klux Klan as an organization, they also drove its members, and their tactics, more than ever into the Democratic camp.

Growing Northern disillusionment with Radical Reconstruction and with the Grant administration became evident in the Liberal Republican movement of 1872, which resulted in the nomination of the erratic Horace Greeley for president. Though Grant was overwhelmingly reelected, the true temper of the country was demonstrated in the congressional elections of 1874, which gave the Democrats control of the House of Representatives for the first time since the outbreak of the Civil War. Despite Grant's hope for a third term in office, most Republicans recognized by 1876 that it was time to change both the candidate and his Reconstruction program, and the nomination of Rutherford B. Hayes of Ohio, a moderate Republican of high principles and of deep sympathy for the South, marked the end of the Radical domination of the Republican Party.

The circumstances surrounding the disputed election of 1876 strengthened Hayes's intention to work with the Southern whites, even if it meant abandoning the few Radical regimes that remained in the South. In an election marked by widespread fraud and many irregularities, the Democratic candidate, Samuel J. Tilden, received the majority of the popular vote; but the vote in the electoral college was long in doubt. In order to resolve the impasse, Hayes's lieutenants had to enter into agreement with Southern Democratic congressmen, promising to withdraw the remaining federal troops from the South, to share the Southern patronage with Democrats, and to favour that section's demands for federal subsidies in the building of levees and railroads. Hayes's inauguration marked, for practical purposes, the restoration of "home rule" for the South—*i.e.,* that the North would no longer interfere in Southern elections to protect the blacks and that the Southern whites would again take control of their state governments.

Restoration of Southern "home rule"

THE NEW SOUTH, 1877–90

The era of conservative domination, 1877–90. The Republican regimes in the Southern states began to fall as early as 1870; by 1877 they had all collapsed. For the next 13 years the South was under the leadership of white Democrats whom their critics called Bourbons because, like the French royal family, they supposedly had learned nothing and forgotten nothing from the revolution they had experienced. For the South as a whole, the characterization is neither quite accurate nor quite fair. In most Southern states the new political leaders represented not only the planters but also the rising Southern business community, interested in railroads, cotton textiles, and urban land speculation.

Even on racial questions the new Southern political leaders were not so reactionary as the label Bourbon might suggest. Though whites were in the majority in all but two of the Southern states, the conservative regimes did not attempt to disfranchise the Negroes. Partly their restraint was caused by fear of further federal intervention; chiefly, however, it stemmed from a conviction on the part of conservative leaders that they could control the black voters, whether through fraud, intimidation, or manipulation.

Indeed, Negro votes were sometimes of great value to these regimes, which favoured the businessmen and planters of the South at the expense of the small white farmers. These "Redeemer" governments sharply reduced or even eliminated the programs of the state governments that benefited poor people. The public school system was starved for money; in 1890 the per capita expenditure in the South for public education was only 97 cents, as compared with $2.24 in the country as a whole. The care of state prisoners, the insane, and the blind was also neglected; and measures to safeguard the public health were rejected. At the same time these conservative regimes were often astonishingly corrupt, and embezzlement and defalcation on the part of public officials were even greater than during the Reconstruction years.

Reduced benefits for the poor

The small white farmers resentful of planter dominance, residents of the hill country outvoted by Black Belt constituencies, and politicians excluded from the ruling cabals tried repeatedly to overthrow the conservative regimes in the South. During the 1870s they supported Independent or Greenback Labor candidates, but without notable success. In 1879 the Readjuster Party in Virginia—so named because its supporters sought to readjust the huge funded debt of that state so as to lessen the tax burden on small farmers—gained control of the legislature and secured in 1880 the election of its leader, General William Mahone, to the U.S. Senate. Not until 1890, however, when the powerful Farmers' Alliance, hitherto devoted exclusively to the promotion of agricultural reforms, dropped its ban on politics, was there an effective challenge to conservative hegemony. In that year, with Alliance backing, Benjamin R. Tillman was chosen governor of South Carolina and James S. Hogg was elected governor of Texas; the heyday of Southern populism was at hand.

Jim Crow legislation. Negro voting in the South was a casualty of the conflict between Redeemers and Populists. Although some Populist leaders, such as Tom Watson in Georgia, saw that poor whites and poor blacks in the South had a community of interest in the struggle against the planters and the businessmen, most small white farmers exhibited vindictive hatred toward the blacks, whose votes had so often been instrumental in upholding conservative regimes. Beginning in 1890, when Mississippi held a new constitutional convention, and continuing through 1908, when Georgia amended its constitution, every state of the former Confederacy moved to disfranchise blacks. Because the U.S. Constitution forbade outright racial discrimination, the Southern states excluded Negroes by requiring that potential voters be able to read or to interpret any section of the Constitution—a requirement that local registrars waived for whites but rigorously insisted upon when an audacious black wanted to vote. Louisiana, more ingenious, added the "grandfather clause" to its constitution, which exempted from this literacy test all of those who had been entitled to vote on Jan. 1, 1867—*i.e.,* before Congress imposed Negro suffrage upon the South—together with their sons and grandsons. Other states imposed stringent property qualifications for voting or enacted complex poll taxes.

Discrimination in voting

Socially as well as politically, race relations in the South deteriorated as farmers' movements rose to challenge the conservative regimes. By 1890, with the triumph of Southern populism, the black's place was clearly defined by law; he was relegated to a subordinate and entirely segregated

position. Not only were legal sanctions (some reminiscent of the "Black Codes") being imposed upon the Negroes, but informal, extralegal, and often brutal steps were also being taken to keep them in their "place." From 1889 to 1899, lynchings in the South averaged 187.5 per year.

Booker T. Washington and the Atlanta Compromise. Faced with implacable and growing hostility from Southern whites, many blacks during the 1880s and '90s felt that their only sensible course was to avoid open conflict and to work out some pattern of accommodation. The most influential black spokesman for this policy was Booker T. Washington, the head of Tuskegee Institute in Alabama, who urged his fellow Negroes to forget about politics and college education in the classical languages and to learn how to be better farmers and artisans. With thrift, industry, and abstention from politics, he thought that Negroes could gradually win the respect of their white neighbours. In 1895, in a speech at the opening of the Atlanta Cotton States and International Exposition, Washington most fully elaborated his position, which became known as the Atlanta Compromise. Abjuring hopes of federal intervention in behalf of the Negro, Washington argued that reform in the South would have to come from within. Change could best be brought about if blacks and whites recognized that "the agitation of questions of social equality is the extremest folly"; in the social life the races in the South could be as separate as the fingers, but in economic progress as united as the hand.

Enthusiastically received by Southern whites, Washington's program also found many adherents among Southern blacks, who saw in his doctrine a way to avoid head-on, disastrous confrontations with overwhelming white force. Whether or not Washington's plan would have produced a generation of orderly, industrious, frugal blacks slowly working themselves into middle-class status is not known because of the intervention of a profound economic depression throughout the South during most of the post-Reconstruction period. Neither poor whites nor poor blacks had much opportunity to rise in a region that was desperately impoverished. By 1890 the South ranked lowest in every index that compared the sections of the United States—lowest in per capita income, lowest in public health, lowest in education. In short, by the 1890s the South, a poor and backward region, had yet to recover from the ravages of the Civil War or to reconcile itself to the readjustments required by the Reconstruction era. (D.H.D.)

Economic depression in the South

The transformation of American society, 1865–1900

NATIONAL EXPANSION

Growth of the nation. The population of the continental United States in 1880 was slightly above 50,000,000. In 1900 it was just under 76,000,000, a gain of more than 50 percent, but still the smallest rate of population increase for any 20-year period of the 19th century. The rate of growth was unevenly distributed, ranging from less than 10 percent in northern New England to more than 125 percent in the 11 states and territories of the Far West. Most of the states east of the Mississippi reported gains slightly below the national average.

Immigration. Much of the population increase was due to the more than 9,000,000 immigrants who entered the United States in the last 20 years of the century, the largest number to arrive in any comparable period up to that time. From the earliest days of the republic until 1895, the majority of immigrants had always come from northern or western Europe. Beginning in 1896, however, the great majority of the immigrants were from southern or eastern Europe. Nervous Americans, already convinced that immigrants wielded too much political power or were responsible for violence and industrial strife, found new cause for alarm, fearing that the new immigrants could not easily be assimilated into American society. Those fears gave added stimulus to agitation for legislation to limit the number of immigrants eligible for admission to the United States and led, in the early 20th century, to quota laws favouring immigrants from northern and western Europe.

Until that time, the only major restriction against immigration was the Chinese Exclusion Act, passed by Congress in 1882, prohibiting for a period of 10 years the immigration of Chinese labourers into the United States. This act was both the culmination of more than a decade of agitation on the West Coast for the exclusion of the Chinese and an early sign of the coming change in the traditional U.S. philosophy of welcoming virtually all immigrants. In response to pressure from California, Congress had passed an exclusion act in 1879, but it had been vetoed by President Hayes on the ground that it abrogated rights guaranteed to the Chinese by the Burlingame Treaty of 1868. In 1880 these treaty provisions were revised to permit the United States to suspend the immigration of Chinese. The Chinese Exclusion Act was renewed in 1892 for another 10-year period, and in 1902 the suspension of Chinese immigration was made indefinite.

The Chinese Exclusion Act

Westward migration. The United States completed its North American expansion in 1867, when Secretary of State Seward persuaded Congress to purchase Alaska from Russia for $7,200,000. Thereafter, the development of the West progressed rapidly, with the percentage of American citizens living west of the Mississippi increasing from about 22 percent in 1880 to 27 percent in 1900. New states were added to the Union throughout the century, and by 1900 there were only three territories still awaiting statehood in the continental United States: Oklahoma, Arizona, and New Mexico.

Urban growth. In 1890 the Bureau of the Census discovered that a continuous line could no longer be drawn across the West to define the farthest advance of settlement. Despite the continuing westward movement of population, the frontier had become a symbol of the past. The movement of people from farms to cities more accurately predicted the trends of the future. In 1880 about 28 percent of the American people lived in communities designated by the Bureau of the Census as urban; by 1900 that figure had risen to 40 percent. In those statistics could be read the beginning of the decline of rural power in America and the emergence of a society built upon a burgeoning industrial complex.

Disappearance of the frontier

The West. Abraham Lincoln once described the West as the "treasure house of the nation." In the 30 years after the discovery of gold in California, prospectors found gold or silver in every state and territory of the Far West.

The mineral empire. There were few truly rich "strikes" in the post-Civil War years. Of those few, the most important were the fabulously rich Comstock Lode of silver in western Nevada (first discovered in 1859 but developed more extensively later) and the discovery of gold in the Black Hills of South Dakota (1874) and at Cripple Creek, Colo. (1891).

Each new discovery of gold or silver produced an instant mining town to supply the needs and pleasures of the prospectors. If most of the ore was close to the surface, the prospectors would soon extract it and depart, leaving behind a ghost town—empty of people but a reminder of a romantic moment in the past. If the veins ran deep, organized groups with the capital to buy the needed machinery would move in to mine the subsoil wealth, and the mining town would gain some stability as the centre of a local industry. In a few instances, those towns gained permanent status as the commercial centres of agricultural areas that first developed to meet the needs of the miners but later expanded to produce a surplus that they exported to other parts of the West.

The open range. At the close of the Civil War, the price of beef in the Northern states was abnormally high. At the same time, millions of cattle grazed aimlessly on the plains of Texas. A few shrewd Texans concluded that there might be greater profits in cattle than in cotton, especially because it required little capital to enter the cattle business—only enough to employ a few cowboys to tend the cattle during the year and to drive them to market in the spring. No one owned the cattle, and they grazed without charge upon the public domain.

The one serious problem was the shipment of the cattle to market. The Kansas Pacific resolved that problem when it completed a rail line that ran as far west as Abilene,

The problem of cattle shipment

Kan., in 1867. Abilene was 200 miles (300 kilometres) from the nearest point in Texas where the cattle grazed during the year, but Texas cattlemen almost immediately instituted the annual practice of driving that portion of their herds that was ready for market overland to Abilene in the spring. There they met representatives of Eastern packinghouses, to whom they sold their cattle.

The open-range cattle industry prospered beyond expectations and even attracted capital from conservative investors in the British Isles. By the 1880s the industry had expanded along the plains as far north as the Dakotas. In the meantime, a new menace had appeared in the form of the advancing frontier of population; but the construction of the Santa Fe Railway through Dodge City, Kan., to La Junta, Colo., permitted the cattlemen to move their operations westward ahead of the settlers; Dodge City replaced Abilene as the principal centre for the annual meeting of cattlemen and buyers. Despite sporadic conflicts with settlers encroaching upon the high plains, the open range survived until a series of savage blizzards struck the plains with unprecedented fury in the winter of 1886–87, killing hundreds of thousands of cattle and forcing many owners into bankruptcy. Those who still had some cattle and some capital abandoned the open range, gained title to lands farther west, where they could provide shelter for their livestock, and revived a cattle industry on land that would be immune to further advances of the frontier of settlement. Their removal to these new lands had been made possible in part by the construction of other railroads connecting the region with Chicago and the Pacific coast.

The expansion of the railroads. In 1862 Congress authorized the construction of two railroads that together would provide the first railroad link between the Mississippi valley and the Pacific coast. One was the Union Pacific, to run westward from Council Bluffs, Iowa; the other was the Central Pacific, to run eastward from Sacramento, Calif. To encourage the rapid completion of those roads, Congress provided generous subsidies in the form of land grants and loans. Construction was slower than Congress had anticipated, but the two lines met, with elaborate ceremonies, on May 10, 1869, at Promontory, Utah.

Completion of the first transcontinental railroad

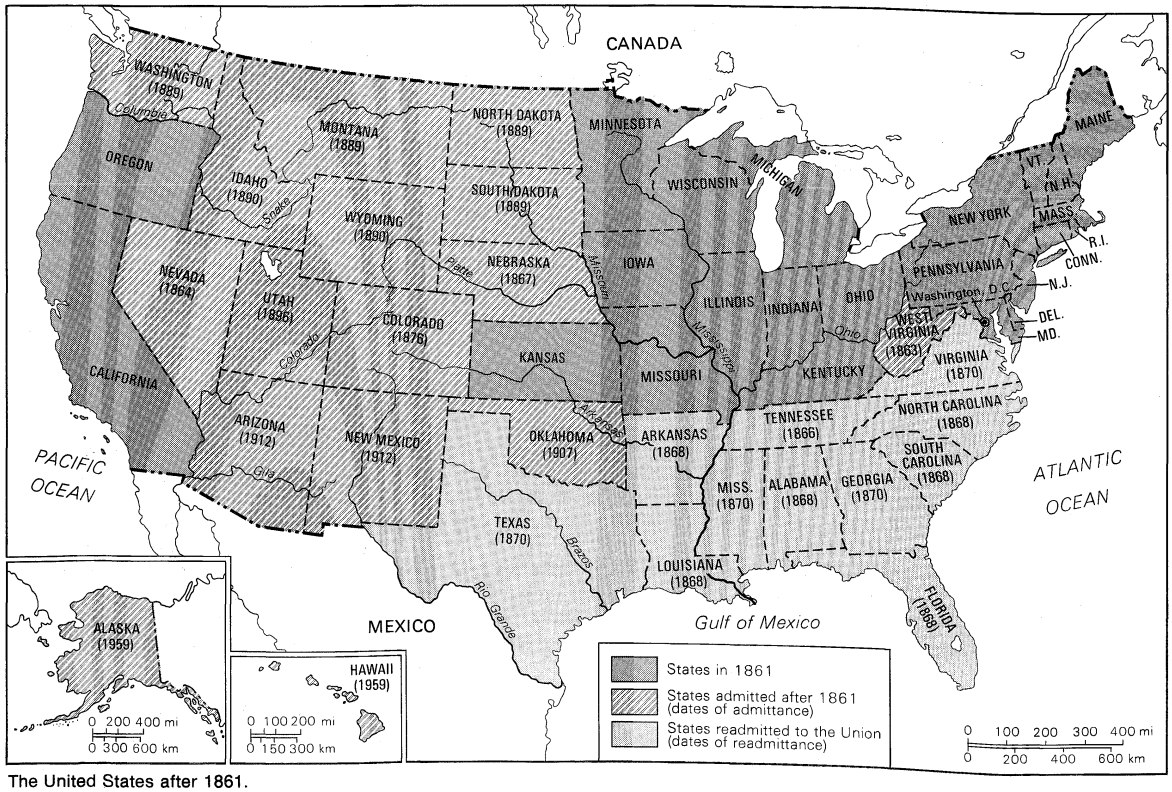
In the meantime, other railroads had begun construction westward, but the panic of 1873 and the ensuing depression halted or delayed progress on many of those lines. With the return of prosperity after 1877, some railroads resumed or accelerated construction; and by 1883 three more rail connections between the Mississippi valley and the West Coast had been completed—the Northern Pacific, from St. Paul to Portland; the Santa Fe, from Chicago to Los Angeles; and the Southern Pacific, from New Orleans to Los Angeles. The Southern Pacific had also acquired, by purchase or construction, lines from Portland to San Francisco and from San Francisco to Los Angeles.

The construction of the railroads from the Midwest to the Pacific coast was the railroad builders' most spectacular achievement in the quarter century after the Civil War. No less important, in terms of the national economy, was the development in the same period of an adequate rail network in the Southern states and the building of other railroads that connected virtually every important community west of the Mississippi with Chicago.

The West developed simultaneously with the building of the Western railroads, and in no part of the nation was the importance of railroads more generally recognized. The railroad gave vitality to the regions it served, but, by withholding service, it could doom a community to stagnation. The railroads appeared to be ruthless in exploiting their powerful position: they fixed prices to suit their convenience; they discriminated among their customers; they attempted to gain a monopoly of transportation wherever possible; and they interfered in state and local politics to elect favourites to office, to block unfriendly legislation, and even to influence the decisions of the courts.

Indian policy. Large tracts of land in the West were reserved by law for the exclusive use of specified Indian tribes. By 1870, however, the invasion of these lands by hordes of prospectors, by cattlemen and farmers, and by the transcontinental railroads had resulted in the outbreak of a series of savage Indian wars and had raised serious questions about the government's Indian policies. Many agents of the Bureau of Indian Affairs were lax in their responsibility for dealing directly with the tribes, and some were corrupt in the discharge of their duties. Most Westerners and some army officers contended that the only satisfactory resolution of the Indian question was the removal of the tribes from all lands coveted by the whites.

In the immediate postwar years, reformers advocated adoption of programs designed to prepare the Indians for ultimate assimilation into American society. In 1869 the reformers persuaded President Grant and Congress to es-



The United States after 1861.

Programs for assimilation of Indians

tablish a nonpolitical Board of Indian Commissioners to supervise the administration of relations between the government and the Indians. The board, however, encountered so much political opposition that it accomplished little. The reformers then proposed legislation to grant title for specific acreages of land to the head of each family in those tribes thought to be ready to adopt a sedentary life as farmers. Congress resisted that proposal until land-hungry Westerners discovered that, if the land were thus distributed, a vast surplus of land would result that could be added to the public domain. When land speculators joined the reformers in support of the proposed legislation, Congress in 1887 enacted the Dawes Act, which empowered the president to grant title to 160 acres (65 hectares) to the head of each family, with smaller allotments to single members of the tribe, in those tribes believed ready to accept a new way of life as farmers. With the grant of land, which could not be alienated by the Indians for 25 years, they were to be granted U.S. citizenship. Reformers rejoiced that they had finally given the Indians an opportunity to have a dignified role in U.S. society, overlooking the possibility that there might be values in Indian culture worthy of preservation. Meanwhile, the land promoters placed successive presidents under great pressure to accelerate the application of the Dawes Act in order to open more land for occupation or speculation.

INDUSTRIALIZATION OF THE U.S. ECONOMY

The growth of industry. By 1878 the United States had reentered a period of prosperity after the long depression of the mid-1870s. In the ensuing 20 years the volume of industrial production, the number of workers employed in industry, and the number of manufacturing plants all more than doubled. A more accurate index to the scope of this industrial advance may be found in the aggregate annual value of all manufactured goods, which increased from about $5,400,000,000 in 1879 to perhaps $13,000,000,000 in 1899. The expansion of the iron and steel industry, always a key factor in any industrial economy, was even more impressive: from 1880 to 1900 the annual production of steel in the United States went from about 1,400,000 to more than 11,000,000 tons. Before the end of the century, the United States surpassed Great Britain in the production of iron and steel and was providing more than one-quarter of the world's supply of pig iron.

Many factors combined to produce this burst of industrial activity. The exploitation of Western resources, including mines and lumber, stimulated a demand for improved transportation, while the gold and silver mines provided new sources of capital for investment in the East. The construction of railroads, especially in the West and South, with the resulting demand for steel rails, was a major force in the expansion of the steel industry and increased the railroad mileage in the United States from less than 93,262 miles (150,151 kilometres) in 1880 to about 190,000 miles (310,000 kilometres) in 1900. Technological advances, including the utilization of the Bessemer and open-hearth processes in the manufacture of steel, resulted in improved products and lower production costs. A series of major inventions, including the telephone, typewriter, linotype, phonograph, electric light, cash register, air brake, refrigerator car, and the automobile, became the bases for new industries, while many of them revolutionized the conduct of business. The use of petroleum products in industry as well as for domestic heating and lighting became the cornerstone of the most powerful of the new industries of the period, while the trolley car, the increased use of gas and electric power, and the telephone led to the establishment of important public utilities that were natural monopolies and could operate only on the basis of franchises granted by state or municipal governments. The widespread employment of the corporate form of business organization offered new opportunities for large-scale financing of business enterprise and attracted new capital, much of it furnished by European investors. Over all this industrial activity, there presided a colourful and energetic group of entrepreneurs, who gained the attention, if not always the commendation, of the public and who appeared to symbolize for the public the new class of leadership in the United States. Of this numerous group the best known were John D. Rockefeller in oil, Andrew Carnegie in steel, and such railroad builders and promoters as Cornelius Vanderbilt, Leland Stanford, Collis P. Huntington, Henry Villard, and James J. Hill.

New class of leadership

The dispersion of industry. The period was notable also for the wide geographic distribution of industry. The Eastern Seaboard from Massachusetts to Pennsylvania continued to be the most heavily industrialized section of the United States, but there was a substantial development of manufacturing in the states adjacent to the Great Lakes and in certain sections of the South.

The experience of the steel industry reflected this new pattern of diffusion. Two-thirds of the iron and steel industry was concentrated in the area of western Pennsylvania and eastern Ohio. After 1880, however, the development of iron mines in northern Minnesota (the Vermilion Range in 1884 and the Mesabi Range in 1892) and in Tennessee and northern Alabama was followed by the expansion of the iron and steel industry in the Chicago area and by the establishment of steel mills in northern Alabama and in Tennessee.

Most manufacturing in the Midwest was in enterprises closely associated with agriculture and represented expansion of industries that had first been established before 1860. Meat-packing, which in the years after 1875 became one of the major industries of the nation in terms of the value of its products, was almost a Midwestern monopoly, with a large part of the industry concentrated in Chicago. Flour milling, brewing, and the manufacture of farm machinery and lumber products were other important Midwestern industries.

The industrial invasion of the South was spearheaded by textiles. Cotton mills became the symbol of the New South, and mills and mill towns sprang up in the Piedmont region from Virginia to Georgia and into Alabama. By 1900 almost one-quarter of all the cotton spindles in the United States were in the South, and Southern mills were expanding their operations more rapidly than were their well-established competitors in New England. The development of lumbering in the South was even more impressive, though less publicized; by the end of the century the South led the nation in lumber production, contributing almost one-third of the annual supply.

Industrial combinations. The geographic dispersal of industry was part of a movement that was converting the United States into an industrial nation. It attracted less attention, however, than the trend toward the consolidation of competing firms into large units capable of dominating an entire industry. The movement toward consolidation received special attention in 1882 when Rockefeller and his associates organized the Standard Oil Trust under the laws of Ohio. A trust was a new type of industrial organization, in which the voting rights of a controlling number of shares of competing firms were entrusted to a small group of men, or trustees, who thus were able to prevent competition among the companies they controlled. The stockholders presumably benefited through the larger dividends they received. For a few years the trust was a popular vehicle for the creation of monopolies, and by 1890 there were trusts in whiskey, lead, cottonseed oil, and salt.

Trusts, mergers, and holding companies

In 1892 the courts of Ohio ruled that the trust violated that state's antimonopoly laws. Standard Oil then reincorporated as a holding company under the more hospitable laws of New Jersey. Thereafter, holding companies or outright mergers became the favourite forms for the creation of monopolies, though the term trust remained in the popular vocabulary as a common description of any monopoly. The best-known mergers of the period were those leading to the formation of the American Tobacco Company (1890) and the American Sugar Refining Company (1891). The latter was especially successful in stifling competition, for it quickly gained control of most of the sugar refined in the United States.

Foreign commerce. The foreign trade of the United States, if judged by the value of exports, kept pace with the growth of domestic industry. Exclusive of gold, silver, and reexports, the annual value of exports from the United States in 1877 was about $590,000,000; by 1900 it had

increased to approximately $1,371,000,000. The value of imports also rose, though at a slower rate. When gold and silver are included, there was only one year in the entire period in which the United States had an unfavourable balance of trade; and, as the century drew to a close, the excess of exports over imports increased perceptibly.

Agriculture continued to furnish the bulk of U.S. exports. Cotton, wheat, flour, and meat products were consistently the items with the greatest annual value among exports. Of the nonagricultural products sent abroad, petroleum was the most important, though by the end of the century its position on the list of exports was being challenged by machinery.

Despite the expansion of foreign trade, the U.S. merchant marine was a major casualty of the period. While the aggregate tonnage of all shipping flying the U.S. flag remained remarkably constant, the tonnage engaged in foreign trade declined sharply, dropping from more than 2,400,000 tons on the eve of the Civil War to a low point of only 726,000 tons in 1898. The decline began during the Civil War when hundreds of ships were transferred to foreign registries to avoid destruction. Later, cost disadvantages in shipbuilding and repair and the American policy of registering only American-built ships hindered growth until World War I.

Labour. The expansion of industry was accompanied by increased tensions between employers and workers and by the appearance, for the first time in the United States, of national labour unions.

Formation of unions. The first effective labour organization that was more than regional in membership and influence was the Knights of Labor, organized in 1869. The Knights believed in the unity of the interests of all producing groups and sought to enlist in their ranks not only all labourers but everyone who could be truly classified as a producer. They championed a variety of causes, many of them more political than industrial, and they hoped to gain their ends through politics and education rather than through economic coercion.

The Knights of Labor

The hardships suffered by many workers during the depression of 1873–78 and the failure of a nationwide railroad strike, which was broken when President Hayes sent federal troops to suppress disorders in Pittsburgh and St. Louis, caused much discontent in the ranks of the Knights. In 1879 Terence V. Powderly, a railroad worker and mayor of Scranton, Pa., was elected grand master workman of the national organization. He favoured cooperation over a program of aggressive action, but the effective control of the Knights shifted to regional leaders who were willing to initiate strikes or other forms of economic pressure to gain their objectives. The Knights reached the peak of their influence in 1884–85, when much-publicized strikes against the Union Pacific, Southwest System, and Wabash railroads attracted substantial public sympathy and succeeded in preventing a reduction in wages. At that time they claimed a national membership of nearly 700,000. In 1885 Congress, taking note of the apparently increasing power of labour, acceded to union demands to prohibit the entry into the United States of immigrants who had signed contracts to work for specific employers.

The year 1886 was a troubled one in labour relations. There were nearly 1,600 strikes, involving about 600,000 workers, with the eight-hour day the most prominent item in the demands of labour. About half of these strikes were called for May Day; some of them were successful, but the failure of others and internal conflicts between skilled and unskilled members led to a decline in the Knights' popularity and influence.

The Haymarket Riot. The most serious blow to the unions came from a tragic occurrence with which they were only indirectly associated. One of the strikes called for May Day in 1886 was against the McCormick Harvesting Machine Company in Chicago. Fighting broke out along the picket lines on May 3, and, when police intervened to restore order, several strikers were injured or killed. Union leaders called a protest meeting at Haymarket Square for the evening of May 4; but, as the meeting was breaking up, a group of anarchists took over and began to make inflammatory speeches. The police quickly intervened, and a bomb exploded, killing seven policemen and injuring many others. Eight of the anarchists were arrested, tried, and convicted of murder. Four of them were hanged, and one committed suicide. The remaining three were pardoned in 1893 by Governor John P. Altgeld, who was persuaded that they had been convicted in such an atmosphere of prejudice that it was impossible to be certain that they were guilty.

The public tended to blame organized labour for the Haymarket tragedy, and many persons had become convinced that the activities of unions were likely to be attended by violence. The Knights never regained the ground they lost in 1886, and, until after the turn of the century, organized labour seldom gained any measure of public sympathy. Aggregate union membership did not again reach its 1885–86 figure until 1900. Unions, however, continued to be active; and in each year from 1889 through the end of the century there were more than 1,000 strikes.

As the power of the Knights declined, the leadership in the trade union movement passed to the American Federation of Labor (AFL). This was a loose federation of local and craft unions, organized first in 1881 and reorganized in 1886. For a few years there was some nominal cooperation between the Knights and the AFL, but the basic organization and philosophy of the two groups made cooperation difficult. The AFL appealed only to skilled workers, and its objectives were those of immediate concern to its members: hours, wages, working conditions, and the recognition of the union. It relied on economic weapons, chiefly the strike and boycott, and it eschewed political activity, except for state and local election campaigns. The central figure in the AFL was Samuel Gompers, a New York cigar maker, who was its president from 1886 to his death in 1924.

The American Federation of Labor

NATIONAL POLITICS

The dominant forces in American life in the last quarter of the 19th century were economic and social rather than political. This fact was reflected in the ineffectiveness of political leadership and in the absence of deeply divisive issues in politics, except perhaps for the continuing agrarian agitation for inflation. There were colourful political personalities, but they gained their following on a personal basis rather than as spokesmen for a program of political action. No president of the period was truly the leader of his party, and none apparently aspired to that status except Grover Cleveland during his second term (1893–97). Such shrewd observers of U.S. politics as Woodrow Wilson and James Bryce agreed that great men did not become presidents; and it was clear that the nominating conventions of both major parties commonly selected candidates who were "available" in the sense that they had few enemies.

Congress had been steadily increasing in power since the Johnson administration and, in the absence of leadership from the White House, was largely responsible for formulating public policy. As a result, public policy commonly represented a compromise among the views of many congressional leaders—a situation made the more essential because of the fact that in only four of the 20 years from 1877 to 1897 did the same party control the White House, the Senate, and the House.

The Republicans appeared to be the majority party in national politics. From the Civil War to the end of the century, they won every presidential election save those of 1884 and 1892, and they had a majority in the Senate in all but three Congresses during that same period. The Democrats, however, won a majority in the House in eight of the 10 Congresses from 1875 to 1895. The success of the Republicans was achieved in the face of bitter intraparty schisms that plagued Republican leaders from 1870 until after 1890 and despite the fact that, in every election campaign after 1876, they were forced to concede the entire South to the opposition. The Republicans had the advantage of having been the party that had defended the Union against secession and had freed the slaves. When all other appeals failed, Republican leaders could salvage votes in the North and West by reviving memories of the war. A less tangible but equally valuable advantage was the widespread belief that the continued industrial

development of the nation would be more secure under a Republican than under a Democratic administration. Except in years of economic adversity, the memory of the war and confidence in the economic program of the Republican Party were normally enough to ensure Republican success in most of the Northern and Western states.

The Rutherford B. Hayes administration. President Hayes (served 1877–81) willingly carried out the commitments made by his friends to secure the disputed Southern votes needed for his election. He withdrew the federal troops still in the South, and he appointed former senator David M. Key of Tennessee to his Cabinet as postmaster general. Hayes hoped that these conciliatory gestures would encourage many Southern conservatives to support the Republican Party in the future. But the Southerners' primary concern was the maintenance of white supremacy; this, they believed, required a monopoly of political power in the South by the Democratic Party. As a result, the policies of Hayes led to the virtual extinction rather than the revival of the Republican Party in the South.

Hayes's efforts to woo the South irritated some Republicans, but his attitude toward patronage in the federal civil service was a more immediate challenge to his party. In June 1877 he issued an executive order prohibiting political activity by those who held federal appointments.

Civil service reform

When two friends of Senator Roscoe Conkling defied this order, Hayes removed them from their posts in the administration of the Port of New York. Conkling and his associates showed their contempt for Hayes by bringing about the election of one of the men (Alonzo B. Cornell) as governor of New York in 1879 and by nominating the other (Chester A. Arthur) as Republican candidate for the vice presidency in 1880.

One of the most serious issues facing Hayes was that of inflation. Hayes and many other Republicans were staunch supporters of a sound-money policy, but the issues were sectional rather than partisan. In general, sentiment in the agricultural South and West was favourable to inflation, while industrial and financial groups in the Northeast opposed any move to inflate the currency, holding that this would benefit debtors at the expense of creditors.

The silver controversy

In 1873 Congress had discontinued the minting of silver dollars, an action later stigmatized by friends of silver as the Crime of '73. As the depression deepened, inflationists began campaigns to persuade Congress to resume coinage of silver dollars and to repeal the act providing for the redemption of Civil War greenbacks in gold after Jan. 1, 1879. By 1878 the sentiment for silver and inflation was so strong that Congress passed, over the president's veto, the Bland–Allison Act, which renewed the coinage of silver dollars and, more significantly, included a mandate to the secretary of the treasury to purchase silver bullion at the market price in amounts of not less than $2,000,000 and not more than $4,000,000 each month.

Opponents of inflation were somewhat reassured by the care with which Secretary of the Treasury John Sherman was making preparation to have an adequate gold reserve to meet any demands on the Treasury for the redemption of greenbacks. Equally reassuring were indications that the nation had at last recovered from the long period of depression. These factors reestablished confidence in the financial stability of the government; and, when the date for the redemption of greenbacks arrived, there was no appreciable demand upon the Treasury to exchange them for gold.

Hayes chose not to run for reelection. Had he sought a second term, he would almost certainly have been denied renomination by the Republican leaders, many of whom he had alienated through his policies of patronage reform and Southern conciliation. Three prominent candidates contended for the Republican nomination in 1880: Grant, the choice of the "Stalwart" faction led by Senator Conkling; James G. Blaine, the leader of the rival "Half-Breed" faction; and Secretary of the Treasury Sherman. Grant had a substantial and loyal bloc of delegates in the convention, but their number was short of a majority. Neither of the other candidates could command a majority, and on the 36th ballot the weary delegates nominated a compromise candidate, Congressman James A. Garfield of Ohio. To placate the Stalwart faction, the convention nominated Chester A. Arthur of New York for vice president.

The Democrats probably would have renominated Samuel J. Tilden in 1880, hoping thereby to gain votes from those who believed Tilden had lost in 1876 through fraud. But Tilden declined to become a candidate again, and the Democratic convention nominated General Winfield S. Hancock. Hancock had been a Federal general during the Civil War, but he had no political record and little familiarity with questions of public policy.

The campaign failed to generate any unusual excitement and produced no novel issues. As in every national election of the period, the Republicans stressed their role as the party of the protective tariff and asserted that Democratic opposition to the tariff would impede the growth of domestic industry. Actually, the Democrats were badly divided on the tariff, and Hancock surprised political leaders of both parties by declaring that the tariff was an issue of only local interest.

Garfield won the election with an electoral margin of 214 to 155, but his plurality in the popular vote was a slim 9,644. The election revealed the existence of a new "solid South," for Hancock carried all the former Confederate states and three of the former slave states that had remained loyal to the Union.

The administrations of James A. Garfield and Chester A. Arthur. Garfield had not been closely identified with either the Stalwarts or the Half-Breeds, the two major factions within the Republican Party, but, upon becoming president, he upset the Stalwarts by naming the Half-Breed Blaine secretary of state. He gave even more serious offense to the Stalwart faction by appointing as collector of customs at New York a man who was unacceptable to the two senators from that state, Conkling and Thomas Platt, who showed their displeasure by resigning their Senate seats, expecting to be reelected triumphantly by the legislature of New York; but in this they were disappointed.

The tragic climax to this intraparty strife came on July 2, 1881, when Garfield was shot in Washington, D.C., by a disappointed and mentally deranged office seeker and Stalwart supporter. For two months the president lingered between life and death. He died on September 19 and was succeeded by Vice President Arthur.

The assassination of Garfield

Arthur's accession to the presidency caused widespread concern. He had held no elective office before becoming vice president, and he had been closely associated with the Stalwart wing of the party. It was assumed that, like others in that group, he would be hostile to civil service reform, and his nomination for the vice presidency had been generally regarded as a deliberate rebuke to President Hayes. The members of Garfield's Cabinet immediately tendered their resignations, but Arthur asked them to continue in office for a time. By mid-April 1882, however, all but one of the Cabinet officers had been replaced.

Arthur soon surprised his critics and the country by demonstrating an unexpected independence of his former political friends. In his first annual message to Congress, in December 1881, he announced his qualified approval of legislation that would remove appointments to the federal civil service from partisan control. In January 1883 Congress passed and Arthur signed the Pendleton Civil Service Act, which established the Civil Service Commission and provided that appointments to certain categories of offices should be made on the basis of examinations and the appointees given an indefinite tenure in their positions.

By 1884, when the next presidential election was held, Arthur's administration had won the respect of many who had viewed his accession to office with misgivings. It had not, however, gained him any strong following among the leaders of his party. The foremost candidate for the Republican nomination was the perennially powerful Blaine, who, despite opposition from those who believed he was too partisan in spirit or that he was vulnerable to charges of corrupt actions while speaker of the house many years before, was nominated on the fourth ballot.

The Democratic candidate, Governor Grover Cleveland of New York, was in many respects the antithesis of Blaine. He was a relative newcomer to politics. He had been elected mayor of Buffalo in 1881 and governor of

New York in 1882. In both positions he had earned a reputation for political independence, inflexible honesty, and an industrious and conservative administration. His record made him an attractive candidate for persons who accepted the dictum that "a public office is a public trust." This was, in 1884, a valuable asset; and it won for Cleveland the support of a few outstanding Republicans and some journals of national circulation that usually favoured Republican nominees for office.

As in 1880, the campaign was almost devoid of issues of public policy: only the perennial question of the tariff appeared to separate the two parties. Cleveland had not served in the army during the Civil War, and Republicans made an effort to use this fact, together with the power of the South in the Democratic Party, to arouse sectional prejudices against Cleveland. During the campaign it was revealed that Cleveland, a bachelor, was the father of an illegitimate son, an indiscretion that gave the Republicans a moral issue with which to counteract charges of corruption against their own candidate.

The election was very close. On the evening of the voting it was apparent that the result depended upon the vote in New York state, but not until the end of the week was it certain that Cleveland had carried New York by the narrow margin of 1,149 votes and been elected president.

Grover Cleveland's first term. Cleveland was the first Democratic president since James Buchanan a quarter of a century earlier. More than two-thirds of the electoral votes he received came from Southern or border states, so that it appeared that his election marked the close of one epoch and the beginning of a new political era in which the South could again hope to have a major voice in the conduct of national affairs. Because of his brief career in politics, Cleveland had only a limited acquaintance with leaders of his own party. He accepted literally the constitutional principle of the separation of powers, and he opened his first annual message to Congress, in December 1885, with an affirmation of his devotion to "the partitions of power between our respective departments." This appeared to be a disavowal of presidential leadership, but it quickly became apparent that Cleveland intended to defend vigorously the prerogatives that he believed belonged to the executive.

Cleveland's relations with his party and politics

During his first term (1885–89) Cleveland was confronted with a divided Congress—a Republican Senate and a Democratic House. This added to the complexities of administration, especially in the matter of appointments. Cleveland was a firm believer in a civil service based on merit rather than on partisan considerations, but, as the first Democratic president in a quarter of a century, he was under great pressure to replace Republicans in appointive offices with Democrats. He followed a line of compromise. In his first two years he removed the incumbents from about two-thirds of the offices subject to his control, but he scrutinized the qualifications of Democrats recommended for appointment and in a number of instances refused to abide by the recommendations of his party leaders. He thus offended both the reformers, who wished no partisan removals, and his fellow Democrats, whose nominees he rejected. Although his handling of the patronage alienated some powerful Democrats, he scored a personal triumph when he persuaded Congress to repeal the obsolete Tenure of Office Act of 1867, which Republican senators had threatened to revive in order to embarrass him.

Cleveland was a conservative on all matters relating to money, and he was inflexibly opposed to wasteful expenditure of public funds. This caused him to investigate as many as possible of the hundreds of private bills passed by Congress to compensate private individuals, usually Federal veterans, for claims against the federal government. When, as was frequently the case, he judged these claims to be ill-founded, he vetoed the bill. He was the first president to use the veto power extensively to block the enactment of this type of private legislation.

Cleveland's use of veto powers

The surplus and the tariff. The flurry of private pension bills had been stimulated, in part, by a growing surplus in the Treasury. In every year since the Civil War, there had been an excess of revenue over expenditures, a circumstance that encouraged suggestions for appropriations of public funds for a variety of purposes. The surplus also focused attention upon the tariff, the principal source of this excess revenue. In 1883 Congress had reviewed the tariff and made numerous changes in the rates, increasing the tariff on some items and reducing it on others, without materially decreasing the revenue received. Cleveland believed that the surplus presented a very real problem. It hoarded in the Treasury money that could have been in circulation, and it encouraged reckless spending by the government. Like many other Democrats, he disliked the high protective tariff. After waiting in vain for two years for Congress to meet this issue boldly, Cleveland adopted the extraordinary tactic of devoting his entire annual message in 1887 to a discussion of this question and to an appeal for a lowering of the tariff. The House then passed a bill generally conforming to Cleveland's views on the tariff; but the Senate rejected it, and the tariff became a leading issue in the presidential campaign of 1888.

The public domain. After 1877 hundreds of thousands of agricultural settlers went westward to the Plains, where they came into competition for control of the land with the cattlemen, who hitherto had dominated the open range. The pressure of population as it moved into the Plains called attention to the diminishing supply of good arable land still open to settlement, thus presaging the day when there would no longer be a vast reservoir of land in the West awaiting the farmer. It also drew attention to the fact that millions of acres of Western land were being held for speculative purposes and that other millions of acres had been acquired by questionable means or were still in the possession of railroads that failed to fulfill the obligations they had assumed when the land was granted to them. Upon assuming office, Cleveland was confronted with evidence that some of these claims had been fraudulently obtained by railroads, speculators, cattlemen, or lumbering interests. He ordered an investigation, and for more than a year agents of the Land Office roamed over the West uncovering evidence of irregularities and neglected obligations. Cleveland acted firmly. By executive orders and court action he succeeded in restoring more than 81,000,000 acres (33,000,000 hectares) to the public domain.

Land reform

The Interstate Commerce Act. The railroads were vital to the nation's economy, but, because in so many regions a single company enjoyed a monopoly of rail transportation, many of the railroads adopted policies that large numbers of their customers felt to be unfair and discriminatory. Before 1884 it was clear that the Granger laws of the preceding decade (state laws prohibiting various abuses by the railroads) were ineffective, and pressure groups turned to the federal government for relief. In this, Western farm organizations were joined by influential Eastern businessmen who believed that they, too, were the victims of discrimination by the railroads. This powerful political alliance persuaded both parties to include regulation of the railroads in their national platforms in 1884 and induced Congress to enact the Interstate Commerce Act in 1887.

This law, designed to prevent unjust discrimination by the railroads, prohibited the pooling of traffic and profits, made it illegal for a railroad to charge more for a short haul than for a longer one, required that the roads publicize their rates, and established the Interstate Commerce Commission to supervise the enforcement of the law. The rulings of the commission were subject to review by the federal courts, the decisions of which tended to narrow the scope of the act. The commission was less effective than the sponsors of the act had hoped, but the act in itself was an indication of the growing realization that only the federal government could cope with the new economic problems of the day.

The election of 1888. Cleveland's plea for a reduction of the tariff in his annual message of 1887 made it certain that the tariff would be the central issue in the presidential campaign of 1888. The Democrats renominated Cleveland, although it was thought that he had endangered his chances of reelection by his outspoken advocacy of tariff reduction. The Republicans had their usual difficulty in selecting a candidate. Blaine refused to enter the race, and no other person in the party commanded substantial support. From among the many who were willing to ac-

cept the nomination, the Republicans selected Benjamin Harrison of Indiana, a Federal general in the Civil War and the grandson of President William Henry Harrison.

Cleveland had won respect as a man of integrity and courage, but neither he nor Harrison aroused any great enthusiasm among the voters. One feature of the campaign noted by observers was the extensive use of money to influence the outcome; this was not a new phenomenon, but the spending of money to carry doubtful states and the apparent alliance between business and political bosses had never before been so open.

The results were again close. Cleveland had a plurality of about 100,000 popular votes, but the Republicans carried two states, New York and Indiana, which they had lost in 1884, and in the electoral college Harrison won by a margin of 233 to 168.

The Benjamin Harrison administration. The Republicans also gained control of both houses of the 51st Congress. Their margin in the House of Representatives, however, was so small that it seemed uncertain whether they could carry controversial legislation through it. This obstacle was overcome by the speaker of the House, Thomas B. Reed of Maine. Reed refused to recognize dilatory motions, and, contrary to precedent, he counted as present all members who were in the chamber. Using that tactic, he ruled, on occasion, that a quorum was present even though fewer than a majority had actually answered a roll call. His iron rule of the House earned him the sobriquet Czar Reed, but only through his firm control of the House could the Republicans pass three controversial bills in the summer and early autumn of 1890. One dealt with monopolies, another with silver, and the third with the tariff.

Public dissatisfaction with monopolies

The Sherman Anti-Trust Act. The first of these major measures declared illegal all combinations that restrained trade between states or with foreign nations. This law, known as the Sherman Anti-Trust Act, was passed by Congress early in July. It was the congressional response to evidence of growing public dissatisfaction with the development of industrial monopolies, which had been so notable a feature of the preceding decade.

More than 10 years passed before the Sherman Act was used to break up any industrial monopoly. It was invoked by the federal government in 1894 to obtain an injunction against a striking railroad union accused of restraint of interstate commerce, and the use of the injunction was upheld by the Supreme Court in 1895. Indeed, it is unlikely that the Senate would have passed the bill in 1890 had not the chairman of the Senate Judiciary Committee, George F. Edmunds of Vermont, felt certain that unions were combinations in restraint of trade within the meaning of the law. To those who hoped that the Sherman Act would inhibit the growth of monopoly, the results were disappointing. The passage of the act only three years after the Interstate Commerce Act was, however, another sign that the public was turning from state capitals to Washington for effective regulation of industrial giants.

The silver issue. Less than two weeks after Congress passed the antitrust law, it enacted the Sherman Silver Purchase Act, which required the secretary of the treasury to purchase each month 4,500,000 ounces (130,000 kilograms) of silver at the market price. This act superseded the Bland–Allison Act of 1878, effectively increasing the government's monthly purchase of silver by more than 50 percent. It was adopted in response to pressure from mine-owners, who were alarmed by the falling price of silver, and from Western farmers, who were always favourable to inflationary measures and who, in 1890, were also suffering from the depressed prices of their products.

The McKinley tariff. Most Republican leaders had been lukewarm to the proposal to increase the purchase of silver and had accepted it only to assure Western votes for the measure in which they were most interested—upward revision of the protective tariff. This was accomplished in the McKinley Tariff Act of October 1890, passed by Congress one month before the midterm elections of that year. The tariff was designed to appeal to the farmers because some agricultural products were added to the protected list. A few items, notably sugar, were placed on the free list, and domestic sugar planters were to be compensated by a subsidy of two cents a pound. The central feature of the act, however, was a general increase in tariff schedules, with many of these increases applying to items of general consumption.

The new tariff immediately became an issue in the congressional elections. It failed to halt the downward spiral of farm prices, but there was an almost immediate increase in the cost of many items purchased by the farmers. With discontent already rife in the agricultural regions of the West and South, the McKinley tariff added to the agrarian resentment. The outcome of the elections was a major defeat for the Republicans, whose strength in the House of Representatives was reduced by almost half.

Farmers' resentment of the tariff

The agrarian revolt. Political disaster befell the Republicans in the trans-Mississippi West, resulting from an economic and psychological depression that enveloped the region after widespread crop failures and the collapse of inflated land prices in the summer of 1887. The Western boom had begun in the late 1870s, when the tide of migration into the unoccupied farmlands beyond the Mississippi quickly led to the settlement of hitherto unoccupied parts of Iowa and Minnesota and to the pushing of the frontier westward across the Plains almost literally to the shadows of the Rocky Mountains.

Westward expansion was encouraged by the railroads that served the region. It was supported by the satisfactory price and encouraging foreign market for wheat, the money crop of the Plains. For 10 years, from 1877 through 1886, the farmers on the Plains had the benefit of an abnormally generous rainfall, leading many to assume that climatic conditions had changed and that the rain belt had moved westward to provide adequate rainfall for the Plains. Confidence was followed by unrestrained optimism that engendered wild speculation and a rise in land prices. Lured on by these illusions, the settlers went into debt to make improvements on their farms while small-town leaders dreamed of prodigious growth and authorized bond issues to construct the public improvements they felt certain would soon be needed.

The collapse of these dreams came in 1887. The year opened ominously when the Plains were swept by a catastrophic blizzard in January that killed thousands of head of cattle and virtually destroyed the cattle industry of the open range. The following summer was dry and hot; crops were poor; and, to compound the woes of the farmers, the price of wheat began to slide downward. The dry summer of 1887 was the beginning of a 10-year cycle of little rainfall and searingly hot summers. By the autumn of 1887 the exodus from the Plains had begun; five years later, areas of western Kansas and Nebraska that had once been thriving agricultural centres were almost depopulated. The agricultural regions east of the Plains were less directly affected, though there the farmers suffered from the general decline in farm prices.

Although the disaster on the Plains bred a sense of distress and frustration, the lure of good land was still strong. When the central portion of the present state of Oklahoma was opened to settlement in April 1889, an army of eager settlers, estimated to have numbered 100,000, rushed into the district to claim homesteads and build homes.

The Populists. The collapse of the boom and the falling prices of agricultural products forced many farmers to seek relief through political action. In 1888 and again in 1890 this discontent was expressed through local political groups, commonly known as Farmers' Alliances, which quickly spread through parts of the West and in the South, where economic problems had been aggravated by the shift following the Civil War from a plantation system to sharecrop and crop-lien systems. The alliances won some local victories and contributed to the discomfiture of the Republicans in 1890. They were not, however, an effective vehicle for concerted political action; and in 1891 the leaders of the alliances formed the People's (Populist) Party.

The People's Party

The Populists aspired to become a national party and hoped to attract support from labour and from reform groups generally. In practice, however, they continued through their brief career to be almost wholly a party of Western farmers. (Southern farmers, afraid of splitting the

white vote and thereby allowing blacks into power, largely remained loyal to the Democratic Party.) The Populists demanded an increase in the circulating currency, to be achieved by the unlimited coinage of silver, a graduated income tax, government ownership of the railroads, a tariff for revenue only, the direct election of U.S. senators, and other measures designed to strengthen political democracy and give the farmers economic parity with business and industry. In 1892 the Populists nominated General James B. Weaver of Iowa for president.

The election of 1892. The nominees of the two major parties for president in 1892 were the same as in the election of 1888: Harrison and Cleveland. The unpopularity of the McKinley tariff gave Cleveland an advantage, as did the discontent in the West, which was directed largely against the Republican Party. From the beginning of the campaign it appeared probable that the Democrats would be successful, and Cleveland carried not only the Southern states but also such key Northern states as New York and Illinois. His electoral vote was 277 to 145 for Harrison. Weaver carried only four Western states, three of them states with important silver mines, and received 22 electoral votes.

Cleveland's second term. When Cleveland was inaugurated for his second term in March 1893, the country hovered on the brink of financial panic. Six years of depression in the trans-Mississippi West, the decline of foreign trade after the enactment of the McKinley tariff, and an abnormally high burden of private debt were disquieting features of the situation. Most attention was centred, however, on the gold reserve in the federal Treasury. It was assumed that a minimum reserve of $100,000,000 was necessary to assure redemption of government obligations in gold. When on April 21, 1893, the reserve fell below that amount, the psychological impact was far-reaching. Investors hastened to convert their holdings into gold; banks and brokerage houses were hard-pressed; and many business houses and financial institutions failed. Prices dropped, employment was curtailed, and the nation entered a period of severe economic depression that continued for more than three years.

The panic of 1893

The causes of this disaster were numerous and complex, but the attention that focused on the gold reserve tended to concentrate concern upon a single factor—the restoration of the Treasury's supply of gold. It was widely believed that the principal cause of the drain on the Treasury was the obligation to purchase large amounts of silver. To those who held this view, the obvious remedy was the repeal of the Sherman Silver Purchase Act.

The issue was political as well as economic. It divided both major parties, but most of the leading advocates of existing silver policies were Democrats. Cleveland, however, had long been opposed to the silver-purchase policy, and in the crisis he resolved upon repeal as an essential step in protecting the Treasury. He therefore called Congress to meet in special session on Aug. 7, 1893.

The new Congress had Democratic majorities in both houses, and, if it had any mandate, it was to repeal the McKinley tariff. It had no mandate on the silver issue, and more than half of its Democratic members came from constituencies that favoured an increase in the coinage of silver. Cleveland faced a herculean task in forcing repeal through Congress, but, by the use of every power at his command, he gained his objective. The Sherman Silver Purchase Act was repealed at the end of October by a bill that made no compensating provision for the coinage of silver. Cleveland had won a personal triumph, but he had irrevocably divided his party; and in some sections of the nation he had become the most unpopular president of his generation.

Cleveland's splitting of the Democratic Party

The extent to which Cleveland had lost control of his party became apparent when Congress turned from silver to the tariff. The House passed a bill that would have revised tariff rates downward in accordance with the president's views. In the Senate, however, the bill was so altered that it bore little resemblance to the original measure, and on some items it imposed higher duties than had the McKinley Tariff Act. It was finally passed in August 1894, but Cleveland was so dissatisfied that he refused to sign it; and it became law without his signature. The act contained a provision for an income tax, but this feature was declared unconstitutional by the Supreme Court in 1895.

In the midterm elections of 1894 the Republicans recaptured control of both houses of Congress. This indicated the discontent produced by the continuing depression. It also guaranteed that, with a Democratic president and Republican Congress, there would be inaction in domestic legislation while both parties looked forward to the election of 1896.

At their convention in St. Louis the Republicans selected Governor William McKinley of Ohio as their presidential nominee. He had served in the Federal army during the Civil War, and his record as governor of Ohio tended to offset his association with the unpopular tariff of 1890. His most effective support in winning the nomination, however, was provided by Mark Hanna, a wealthy Cleveland businessman who was McKinley's closest friend.

The Democratic convention in Chicago was unusually exciting. It was controlled by groups hostile to Cleveland's financial policies, and it took the unprecedented step of rejecting a resolution commending the administration of a president of its own party. The debate on the party platform featured an eloquent defense of silver and agrarian interests by William Jennings Bryan, which won him not only a prolonged ovation but also his party's presidential nomination. Bryan was a former congressman from Nebraska, and at 36 he was the youngest man ever to be the nominee for president of a major party. By experience and conviction he shared the outlook of the agrarian elements that dominated the convention and whose principal spokesman he became.

Bryan conducted a vigorous campaign. For the first time a presidential candidate carried his case to the people in all parts of the country, and for a time it appeared that he might win. The worried conservatives charged that Bryan was a dangerous demagogue, and they interpreted the campaign as a conflict between defenders of a sound economic system that would produce prosperity and dishonest radicals who championed reckless innovations that would undermine the financial security of the nation. On this interpretation they succeeded in raising large campaign funds from industrialists who feared their interests were threatened. With this money, the Republicans were able to turn the tide and win a decisive victory. Outside the South, Bryan carried only the Western silver states and Kansas and Nebraska.

Economic recovery. Soon after taking office on March 4, 1897, McKinley called Congress into special session to revise the tariff once again. Congress responded by passing the Dingley Tariff Act, which eliminated many items from the free list and generally raised duties on imports to the highest level they had yet reached.

Although the preservation of the gold standard had been the chief appeal of the Republicans in 1896, it was not until March 1900 that Congress enacted the Gold Standard Act, which required the Treasury to maintain a minimum gold reserve of $150,000,000 and authorized the issuance of bonds, if necessary, to protect that minimum. In 1900 such a measure was almost anticlimactic, for an adequate gold supply had ceased to be a practical problem. Beginning in 1893, the production of gold in the United States had increased steadily; by 1899 the annual value of gold added to the American supply was double that of any year between 1881 and 1892. The chief source of the new supply of gold was the Klondike, where important deposits of gold had been discovered during the summer of 1896.

Increase in gold production

By 1898 the depression had run its course; farm prices and the volume of farm exports were again rising steadily, and Western farmers appeared to forget their recent troubles and to regain confidence in their economic prospects. In industry, the return of prosperity was marked by a resumption of the move toward more industrial combinations, despite the antitrust law; and great banking houses, such as J.P. Morgan and Company of New York, played a key role in many of the most important of these combinations by providing the necessary capital and receiving, in return, an influential voice in the management of the companies created by this capital. (H.W.Br./Ed.)

Imperialism, the Progressive era, and the rise to world power, 1896–1920

AMERICAN IMPERIALISM

The Spanish-American War. Militarily speaking, the Spanish-American War of 1898 was so brief and relatively bloodless as to have been a mere passing episode in the history of modern warfare. Its political and diplomatic consequences, however, were enormous: it catapulted the United States into the arena of world politics and set it, at least briefly, on the new road of imperialism. To be sure, specific events drove the United States to hostilities in 1898; but the stage had already been set by profound changes in thought about the nation's mission and its destiny.

The United States in world politics

Before the 1890s, roughly speaking, most Americans had adhered stubbornly to the belief, as old as the Revolution itself, that their country should remain aloof from European affairs and offer an example of democracy and peace to the rest of the world; but slowly in the 1880s, and more rapidly in the 1890s, new currents of thought eroded this historic conviction. The United States had become a great power by virtue of its prodigious economic growth since the Civil War; numerous publicists said that it ought to begin to act like one. Propagandists of sea power, above all, Captain Alfred T. Mahan, argued that future national security and greatness depended upon a large navy supported by bases throughout the world. After the disappearance of the American frontier in 1890, the conviction grew that the United States would have to find new outlets for an ever-increasing population and agricultural and industrial production; this belief was particularly rife among farmers in dire distress in the 1890s. Social Darwinists said that the world is a jungle, with international rivalries inevitable, and that only strong nations could survive. Added to these arguments were those of idealists and religious leaders that Americans had a duty to "take up the white man's burden" and to carry their assertedly superior culture and the blessings of Christianity to the backward peoples of the world.

Cuban rebellion

It was against this background that the events of 1898 propelled the United States along the road to war and empire. Cuban rebels had begun a violent revolution against Spanish rule in 1895, set off by a depression caused by a decline in U.S. sugar purchases from Cuba. Rebel violence led progressively to more repressive Spanish countermeasures. Cuban refugees in the United States spread exaggerated tales of Spanish atrocities, and these and numerous others were reprinted widely (particularly by William Randolph Hearst's New York *American* and Joseph Pulitzer's New York *World,* then engaged in a fierce battle for circulation). President Cleveland resisted the rising public demand for intervention, but by early 1898 the pressure, then on his successor, McKinley, was too great to be defied. When an explosion—caused by a submarine mine, according to a U.S. naval court of inquiry—sank the USS *Maine* with large loss of life in Havana harbour on Feb. 15, 1898, events moved beyond the president's control. Though Spain was willing to make large concessions to avoid war, it adamantly resisted what had become the minimum public and official U.S. demand—Spanish withdrawal from Cuba and recognition of the island's independence. Hence Congress in mid-April authorized McKinley to use the armed forces to expel the Spanish from Cuba.

For Americans it was, as Secretary of State John Hay put it in a letter to Theodore Roosevelt, "a splendid little war." An American expeditionary force, after quickly overcoming the Spaniards in Cuba, turned against Spain's last island in the Caribbean, Puerto Rico. Meanwhile, on May 1, 1898, the American commodore George Dewey, with his Asiatic squadron, destroyed a decrepit Spanish flotilla in the harbour of Manila in the Philippines.

The fighting was over by August 12, when the United States and Spain signed a preliminary peace treaty in Washington, D.C. Negotiators met in Paris in October to draw up a definitive agreement. Spain recognized the independence of Cuba and ceded Puerto Rico and Guam to the United States, but the disposition of the Philippines was another matter. Business interests in the United States, which had been noticeably cool about a war over Cuba, demanded the acquisition of the entire Philippine archipelago in the hope that Manila would become the entrepôt for a great Far Eastern trade; chauvinists declaimed against lowering the flag under Spanish pressure. Concluding that he had no alternative, McKinley forced the Spanish to "sell" the Philippines to the United States for $20,000,000.

Acquisition of the Philippines

But a strong reaction in the United States against acquisition of the Philippines had already set in by the time the Treaty of Paris was signed on Dec. 10, 1898; and anti-imperialists declared that the control and governance of distant alien peoples violated all American traditions of self-determination and would even threaten the very fabric of the republic. Though there were more than enough votes in the Senate to defeat the treaty, that body gave its consent to ratification largely because William Jennings Bryan, the Democratic leader, wanted Democrats to approve the treaty and then make imperialism the chief issue of the 1900 presidential campaign.

The new American empire. McKinley easily defeated Bryan in 1900. The victory, however, was hardly a mandate for imperialism, and, as events were soon to disclose, the American people were perhaps the most reluctant imperialists in history. No sooner had they acquired an overseas empire than they set in motion the process of its dissolution or transformation.

By the so-called Teller Amendment to the war resolution, Congress had declared that the United States would not annex Cuba. This pledge was kept, although Cuba was forced in 1903 to sign a treaty making it virtually a protectorate of the United States. The Hawaiian Islands, annexed by Congress on July 7, 1898, were made a territory in 1900 and were hence, technically, only briefly part of the American empire. Puerto Rico was given limited self-government in 1900; and the Jones Act of 1917 conferred full territorial status on the island, gave U.S. citizenship to its inhabitants, and limited its self-government only by the veto of a governor appointed by the president of the United States. Establishing any kind of government in the Philippines was much more difficult because a large band of Filipinos resisted American rule as bravely as they had fought the Spanish. The Philippine insurrection was over by 1901, however, and the Philippine Government Act of 1902 inaugurated the beginning of partial self-government, which was transformed into almost complete home rule by the Jones Act of 1916.

The Open Door in the Far East. Although Americans were reluctant imperialists, the United States was an important Pacific power after 1898, and American businessmen had inflated ambitions to tap what they thought was the huge Chinese market. The doors to that market were being rapidly closed in the 1890s, however, as Britain, France, Russia, and Japan carved out large so-called spheres of influence all the way from Manchuria to southern China. With Britain's support (the British stood to gain the most from equal trade opportunities), on Sept. 6, 1899, Secretary of State Hay addressed the first so-called Open Door note to the powers with interests in China; it asked them to accord equal trade and investment opportunities to all nationals in their spheres of interest and leased territories. With considerable bravado, Hay announced that all the powers had agreed to respect the Open Door, even though the Russians had declined to give any pledges. On July 3, 1900, after the Boxer Rebellion—an uprising in China against foreign influence—Hay circulated a second Open Door note announcing that it was American policy to preserve Chinese territorial and political integrity.

Such pronouncements had little effect because the United States was not prepared to support the Open Door policy with force; successive administrations to the 1940s, however, considered it the cornerstone of their Far Eastern policy. President Theodore Roosevelt reluctantly mediated the Russo-Japanese War in 1905 in part to protect the Open Door as well as to maintain a balance of power in the Far East. When Japan attempted in 1915 to force a virtual protectorate on China, President Woodrow Wilson intervened sternly and in some measure successfully

to protect Chinese independence. Victory for American policy seemed to come with the Nine-Power Treaty of Washington of 1922, when all nations with interests in China promised to respect the Open Door.

Building the Panama Canal and American domination in the Caribbean. Strategic necessity and the desire of Eastern businessmen to have easy access to Pacific markets combined in the late 1890s to convince the president, Congress, and a vast majority of Americans that an isthmian canal linking the Atlantic and Pacific oceans was vital to national security and prosperity. In the Hay–Pauncefote Treaty of 1901, the British government gave up the rights to joint construction with the United States that it had gained under the Clayton–Bulwer Treaty of 1850. A French company, which had tried unsuccessfully to dig a canal across the Isthmus of Panama, was eager to sell its right-of-way to the United States. Thus, the only obstacle to the project was the government of Colombia, which owned Panama. When Colombia was slow to cooperate, Roosevelt, in 1903, covertly supported a Panamanian revolution engineered by officials of the French company. A treaty was quickly negotiated between the United States and the new Republic of Panama; construction began, and the canal was opened to shipping on Aug. 15, 1914.

Concern over what Americans regarded increasingly as their "lifeline" increased in proportion to progress in the construction of the canal. An early manifestation of that concern came in 1902–03, when Britain, Germany, and Italy blockaded Venezuela to force the payment of debts, and particularly when the Germans bombarded and destroyed a Venezuelan town; so agitated was American opinion that Roosevelt used a veiled threat to force Germany to accept arbitration of the debt question by the Hague Court. When the Dominican Republic defaulted on its foreign debt to several European countries in 1904, Roosevelt quickly established an American receivership of the Dominican customs in order to collect the revenues to meet the country's debt payments. Moreover, in his annual message to Congress of 1904, the president announced a new Latin-American policy, soon called the Roosevelt Corollary to the Monroe Doctrine—because the Monroe Doctrine forbade European use of force in the New World, the United States would itself take whatever action necessary to guarantee that Latin-American states gave no cause for such European intervention. It was, in fact, a considerable extension of the Monroe Doctrine, not a correct historical interpretation of it; but it remained the cornerstone of American policy in the Caribbean at least until 1928.

Actually, Roosevelt was reluctant to interfere in the domestic affairs of neighbouring states; his one significant intervention after 1904—the administration of the Cuban government from 1906 to 1909—was undertaken in order to prevent civil war and at the insistence of Cuban authorities. Roosevelt's successor, however, William Howard Taft, had more ambitious plans to guarantee American hegemony in the approaches to the Panama Canal.

Dollar Diplomacy

Adopting a policy called Dollar Diplomacy, Taft hoped to persuade American private bankers to displace European creditors in the Caribbean area and thereby to increase American influence and encourage stability in countries prone to revolution. Dollar Diplomacy was a total failure; its one result was to involve the United States in a civil war in Nicaragua with the effect of perpetuating a reactionary and unpopular regime. (Similar initiatives by the Taft administration in the Far East—most notably a plan for the internationalization of the railroads of Manchuria—also failed.)

The accession of Woodrow Wilson in 1913 seemed to augur the beginning of a new era in Latin-American relations; the new president and his secretary of state, William Jennings Bryan, were idealists who had strongly condemned interventions and Dollar Diplomacy. But, although Wilson did negotiate a treaty with Colombia to make reparation for U.S. complicity in the Panamanian revolution, it was defeated by the Senate. Wilson also tried hard to promote a Pan-American nonaggression pact; but it foundered on the opposition of Chile, which had a long-standing border dispute with Peru.

When crises threatened the domestic stability of the Caribbean area, however, Wilson revealed that he was just as determined to protect American security as Roosevelt and Taft had been and that he was perhaps even more willing to use force. Frequent revolutions and the fear of European intervention led Wilson to impose a protectorate and a puppet government upon Haiti in 1915 and a military occupation of the Dominican Republic in 1916. He concluded a treaty with Nicaragua making that country a protectorate of the United States. Moreover, he purchased the Danish Virgin Islands in 1916 at the inflated price of $25,000,000 in order to prevent their possible transfer from Denmark to Germany.

THE PROGRESSIVE ERA

The character and variety of the Progressive movement. The inauguration of President McKinley in 1897 had seemed to mark the end of an era of domestic turmoil and the beginning of a new period of unparalleled tranquility. Prosperity was returning after the devastating panic of 1893. The agrarian uprising led by Bryan in the election of 1896 had been turned back, and the national government was securely in the hands of friends of big business. The Dingley Tariff Act of 1897 greatly increased tariff rates; the Gold Standard Act of 1897 dashed the hopes of advocates of the free coinage of silver; and McKinley did nothing to stop a series of industrial combinations in defiance of the Sherman Anti-Trust Act.

Origins of progressivism. Never were superficial signs more deceiving. Actually, the United States already was in the first stages of what historians came to call the Progressive movement. Generally speaking, progressivism was the response of various groups to problems raised by the rapid industrialization and urbanization that followed the Civil War. These problems included the spread of slums and poverty; the exploitation of labour; the breakdown of democratic government in the cities and states caused by the emergence of political organizations, or machines, allied with business interests; and a rapid movement toward financial and industrial concentration. Many Americans feared that their historic traditions of responsible democratic government and free economic opportunity for all were being destroyed by gigantic combinations of economic and political power.

Problems of industrialization and urbanization

Actually there was not, either in the 1890s or later, any single Progressive movement. The numerous movements for reform on the local, state, and national levels were too diverse, and sometimes too mutually antagonistic, ever to coalesce into a national crusade. But they were generally motivated by common assumptions and goals—*e.g.*, the repudiation of individualism and laissez-faire, concern for the underprivileged and downtrodden, the control of government by the rank and file, and the enlargement of governmental power in order to bring industry and finance under a measure of popular control.

The origins of progressivism were as complex and are as difficult to describe as the movement itself. In the vanguard were various agrarian crusaders, such as the Grangers and the Populists and Democrats under Bryan, with their demands for stringent railroad regulation and national control of banks and the money supply. At the same time a new generation of economists, sociologists, and political scientists was undermining the philosophical foundations of the laissez-faire state and constructing a new ideology to justify democratic collectivism; and a new school of social workers was establishing settlement houses and going into the slums to discover the extent of human degradation. Allied with them was a growing body of ministers, priests, and rabbis—proponents of what was called the social Gospel—who struggled to arouse the social concerns and consciences of their parishioners. Finally, journalists called "muckrakers" probed into all the dark corners of American life and carried their message of reform through mass-circulation newspapers and magazines.

Two specific catalytic agents set off the Progressive movement—the agrarian depression of the early 1890s and the financial and industrial depression that began in 1893. Low prices drove farmers by the hundreds of thousands into the People's Party of 1892. Widespread suffering in

Depressions of the early 1890s

the cities beginning in 1893 caused a breakdown of many social services and dramatized for the increasing number of urban middle-class Americans the gross inefficiency of most municipal governments.

Urban reforms. A movement already begun, to wrest control of city governments from corrupt political machines, was given tremendous impetus by the panic of 1893. The National Municipal League, organized in 1894, united various city reform groups throughout the country; corrupt local governments were overthrown in such cities as New York in 1894, Baltimore in 1895, and Chicago in 1896–97. And so it went all over the country well into the 20th century.

Despite initial differences among urban reformers, by the early 1900s the vast majority of them were fighting for and winning much the same objectives—more equitable taxation of railroad and corporate property, tenement house reform, better schools, and expanded social services for the poor. Even big-city machines like Tammany Hall became increasingly sensitive to the social and economic needs of their constituents. Reformers also devised new forms of city government to replace the old mayor–city-council arrangement that had proved to be so susceptible to corrupt influences. One was the commission form, which vested all responsibility in a small group of commissioners, each responsible for a single department; another was the city-manager form, which provided administration by a professionally trained expert, responsible to a popularly elected council (these two forms were in widespread use in small and medium-sized cities by 1920).

Reform in state governments. The reform movement spread almost at once to the state level, for it was in state capitals that important decisions affecting the cities were made. Entrenched and very professional political organizations, generously financed by officeholders and businessmen wanting special privileges, controlled most state governments in the late 1890s; everywhere, these organizations were challenged by a rising generation of young and idealistic antiorganization leaders, ambitious for power. They were most successful in the Midwest, under such leaders as Robert M. La Follette of Wisconsin; but they had counterparts all over the country—*e.g.*, Charles Evans Hughes of New York, Woodrow Wilson of New Jersey, Andrew J. Montague of Virginia, and Hiram W. Johnson of California.

Anti-organizational leaders

These young leaders revolutionized the art and practice of politics in the United States, not only by exercising strong leadership but also by effecting institutional changes such as the direct primary, direct election of senators (rather than by state legislatures), the initiative, referendum, and recall—which helped restore and revitalize political democracy. More important, perhaps, progressives to a large degree achieved their economic and social objectives—among them, strict regulation of intrastate railroads and public utilities, legislation to prevent child labour and to protect women workers, penal reform, expanded charitable services to the poor, and accident insurance systems to provide compensation to workers and their families.

Theodore Roosevelt and the Progressive movement. By 1901 the reform upheaval was too strong to be contained within state boundaries. Moreover, certain problems with which only the federal government was apparently competent to deal cried out for solution. McKinley might have succeeded in ignoring the rising tide of public opinion had he served out his second term; but McKinley's assassination in September 1901 brought to the presidency an entirely different kind of man—Theodore Roosevelt, at age 42 the youngest man yet to enter the White House. Roosevelt had broad democratic sympathies; moreover, thanks to his experience as police commissioner of New York City and governor of New York state, he was the first president to have an intimate knowledge of modern urban problems. Because Congress was securely controlled by a group of archconservative Republicans, the new president had to feel his way cautiously in legislative matters; but he emerged full-grown as a tribune of the people after his triumph in the presidential election of 1904. By 1906 he was the undisputed spokesman of national progressivism and by far its best publicity agent. (The White House was, he said, "a bully pulpit.") Meanwhile, by his leadership of public opinion and by acting as a spur on Congress, he had revived the presidency and made it incomparably the most powerful force in national politics.

In 1901, Americans were perhaps most alarmed about the spread of so-called trusts, or industrial combinations, which they thought were responsible for the steady price increases that had occurred each year since 1897. Ever alert to the winds of public opinion, Roosevelt responded by activating the Sherman Anti-Trust Act of 1890, which had lain dormant because of Cleveland's and McKinley's refusal to enforce it and also because of the Supreme Court's ruling of 1895 that the measure did not apply to combinations in manufacturing. Beginning in 1902 with a suit to dissolve a northwestern railroad monopoly, Roosevelt moved next against the so-called Beef Trust, then against the oil, tobacco, and other monopolies. In every case the Supreme Court supported the administration, going so far in the oil and tobacco decisions of 1911 as to reverse its 1895 decision. In addition, in 1903 Roosevelt persuaded a reluctant Congress to establish a Bureau of Corporations with sweeping power to investigate business practices; the bureau's thoroughgoing reports were of immense assistance in antitrust cases. While establishing the supremacy of the federal government in the industrial field, Roosevelt in 1902 also took action unprecedented in the history of the presidency by intervening on labour's behalf to force the arbitration of a strike by the United Mine Workers of America against the Pennsylvania anthracite coal operators.

The breaking of the trusts

Roosevelt moved much more aggressively after his 1904 election. Public demand for effective national regulation of interstate railroad rates had been growing since the Supreme Court had emasculated the Interstate Commerce Commission's (ICC) rate-making authority in the 1890s. Determined to bring the railroads—the country's single greatest private economic interest—under effective national control, Roosevelt waged an unrelenting battle with Congress in 1905–06. The outcome—the Hepburn Act of 1906—was his own personal triumph; it greatly enlarged the ICC's jurisdiction and forbade railroads to increase rates without its approval. By using the same tactics of aggressive leadership, Roosevelt in 1906 also obtained passage of a Meat Inspection Act and a Pure Food and Drug Act. Passage of the former was aided by the publication of Upton Sinclair's famous novel, *The Jungle* (1906), which revealed in gory detail the unsanitary conditions of the Chicago stockyards and meat-packing plants.

Meanwhile, almost from his accession to the presidency, Roosevelt had been carrying on a crusade, often independent of Congress, to conserve the nation's fast-dwindling natural resources and to make them available for exploitation under rigorous national supervision. He withdrew from the public domain some 148,000,000 acres of forest lands, 80,000,000 acres of mineral lands, and 1,500,000 acres of water-power sites. Moreover, adoption of the National Reclamation Act of 1902 made possible the beginning of an ambitious federal program of irrigation and hydroelectric development in the West.

Republican troubles under William Howard Taft. Roosevelt was so much the idol of the masses of 1908 that he could have easily gained the Republican nomination in that year. After his election in 1904, however, he had announced that he would not be a candidate four years later; adhering stubbornly to his pledge, he arranged the nomination of his secretary of war, William Howard Taft of Ohio, who easily defeated Bryan.

Taft might have made an ideal president during a time of domestic tranquility, but his tenure in the White House was far from peaceful. National progressivism was nearly at high tide; and a large group of Republican progressives, called "insurgents," sat in both houses of Congress.

The high tide of progressivism

The Republican insurgents. These Republicans, like a majority of Americans, demanded such reforms as tariff reductions, an income tax, the direct election of senators, and even stricter railroad and corporation regulations. Taft, who had strongly supported Roosevelt's policies, thought of himself as a progressive. Actually he was temperamentally and philosophically a conservative; more-

over, he lacked the qualities of a dynamic popular leader. In the circumstances, his ineptness, indecision, and failure to lead could only spell disaster for his party.

Taft's troubles began when he called Congress into special session in 1909 to take up the first item on his agenda—tariff reform. The measure that emerged from Congress actually increased rates. Republican insurgents and a majority of Americans were outraged, but Taft signed the bill and called it the best tariff law the Republicans had ever enacted. Conflicts and misunderstandings over conservation and legislative procedure caused the rift between Taft Republicans and the insurgents to grow. By 1910 the Republican insurgents were clearly in the ascendancy in the Congress. Taking control of the president's railroad-regulation measure, they added new provisions that greatly enlarged the ICC's authority. The following year they bitterly opposed Taft's measure for tariff reciprocity with Canada; it passed with Democratic support in Congress, only to go down to defeat at the hands of the Canadian electorate.

The 1912 election. Republican insurgents were determined to prevent Taft's renomination in 1912. They found their leader in Roosevelt, who had become increasingly alienated from Taft and who made a whirlwind campaign for the presidential nomination in the winter and spring of 1912. Roosevelt swept the presidential primaries, even in Taft's own state of Ohio; but Taft and conservative Republicans controlled the powerful state organizations and the Republican National Committee and were able to nominate Taft by a narrow margin. Convinced that the bosses had stolen the nomination from him, Roosevelt led his followers out of the Republican convention. In August they organized the Progressive ("Bull Moose") Party and named Roosevelt to lead the third-party cause.

Democrats had swept the 1910 congressional and gubernatorial elections; and, after the disruption of the Republican Party in the spring of 1912, it was obvious that almost any passable Democrat could win the presidency in that year. Woodrow Wilson, former president of Princeton University, who had made a brilliant Progressive record as governor of New Jersey, was nominated by the Democrats on the 46th ballot.

Taft's single objective in the 1912 campaign was to defeat Roosevelt. The real contest was between Roosevelt and Wilson for control of the Progressive majority. Campaigning strenuously on a platform that he called the New Nationalism, Roosevelt demanded effective control of big business through a strong federal commission, radical tax reform, and a whole series of measures to put the federal government squarely into the business of social and economic reform. By contrast Wilson seemed conservative with a program he called the New Freedom; it envisaged a concerted effort to destroy monopoly and to open the doors of economic opportunity to small businessmen through drastic tariff reduction, banking reform, and severe tightening of the antitrust laws. Roosevelt outpolled Taft in the election, but he failed to win many Democratic Progressives away from Wilson, who won by a huge majority of electoral votes, though receiving only about 42 percent of the popular vote.

The New Freedom and its transformation. A trained political scientist and historian, Wilson believed that the president should be the leader of public opinion, the chief formulator of legislative policy, and virtually sovereign in the conduct of foreign relations. With the support of an aroused public opinion and a compliant Democratic majority, he was able to put his theories of leadership into effect with spectacular success.

Wilson's domestic program

The first item in Wilson's program was tariff reform, a perennial Democratic objective since the Civil War; the president's measure, the Underwood Tariff Act of 1913, reduced average rates from 40 percent to 25 percent, greatly enlarged the free list, and included a modest income tax. Next came adoption of the president's measure for banking and monetary reform, the Federal Reserve Act of 1913, which created a federal reserve system to mobilize banking reserves and issue a flexible new currency—federal reserve notes—based on gold and commercial paper; uniting and supervising the entire system was a federal reserve board of presidential appointees.

The third, and Wilson thought the last, part of the New Freedom program was antitrust reform. In his first significant movement toward Roosevelt's New Nationalism, Wilson reversed his position that merely strengthening the Sherman Anti-Trust Act would suffice to prevent monopoly. Instead, he took up and pushed through Congress the Progressive-sponsored Federal Trade Commission Act of 1914. It established an agency—the Federal Trade Commission (FTC)—with sweeping authority to prevent business practices that would lead to monopoly. Meanwhile, Wilson had abandoned his original measure, the Clayton Anti-Trust Act passed by Congress in 1914; its severe provisions against interlocking directorates and practices tending toward monopoly had been gravely weakened by the time the president signed it. The Clayton Act included a declaration that labour unions, as such, were not to be construed as conspiracies in restraint of trade in violation of the antitrust laws; but what organized labour wanted, and did not get, was immunity from prosecution for such measures as the sympathetic strike and the secondary boycott, which the courts had proscribed as violations of the Sherman Act.

In a public letter in November 1914, the president announced that his reform program was complete. But various groups were still demanding the advanced kind of social and economic legislation that Roosevelt had advocated in 1912; also, by early 1916 the Progressive Party had largely disintegrated, and Wilson knew that he could win reelection only with the support of a substantial minority of Roosevelt's former followers. Consequently—and also because his own political thinking had been moving toward a more advanced Progressive position—Wilson struck out upon a new political course in 1916. He began by appointing Louis D. Brandeis, the leading critic of big business and finance, to the Supreme Court. Then in quick succession he obtained passage of a rural-credits measure to supply cheap long-term credit to farmers; anti-child-labour and federal workmen's-compensation legislation; the Adamson Act, establishing the eight-hour day for interstate railroad workers; and measures for federal aid to education and highway construction. With such a program behind him, Wilson was able to rally a new coalition of Democrats, former Progressives, independents, social workers, and a large minority of Socialists; and he narrowly defeated his Republican opponent, Charles Evans Hughes, in the 1916 presidential election.

THE RISE TO WORLD POWER

Woodrow Wilson and the Mexican Revolution. Although Wilson's consuming interest was in domestic politics, he had to deal primarily with foreign affairs while in the White House; and before the end of his presidency he had developed into a diplomatist of great skill as well as one of the commanding figures in world affairs. He was a "strong" president in the conduct of foreign policy, writing most of the important diplomatic correspondence of his government and making all important decisions himself. He usually worked well with his secretaries of state, Bryan and Robert Lansing, and often relied for advice upon his confidential counselor, Colonel Edward M. House of Texas.

Wilson's conduct of foreign affairs

Wilson served his apprenticeship by having to deal at the outset of his administration with an uprising in Mexico, set off when a military usurper, Victoriano Huerta, murdered liberal president Francisco Madero and seized the executive power in February 1913. It was difficult for the United States to remain aloof because Americans had invested heavily in Mexico and 40,000 U.S. citizens resided there.

If Wilson had followed conventional policy and the urgings of Americans with interests in Mexico, he would have recognized Huerta (as most European governments did), who promised to respect and protect all foreign investments and concessions. But Wilson was revolted by Huerta's bloody rise to power; moreover, he believed that the revolution begun by Madero in 1910 was a glorious episode in the history of human liberty. Wilson thus not only refused to recognize Huerta but also tried to persuade the dictator to step down from office and permit the hold-

ing of free elections for a new democratic government. When Huerta refused to cooperate, Wilson gave open support to the Constitutionalists—Huerta's opponents under Madero's successor, Venustiano Carranza—and, when it seemed that the Constitutionalists could not themselves drive Huerta from power, Wilson seized the port of Veracruz in April 1914 to cut off Huerta's supplies and revenues. This stratagem succeeded, and Carranza and his army occupied Mexico City in August.

The revolutionary forces then divided between Carranza's followers and those of his chief rival and most colorful general, Pancho Villa; and civil war raged for another year. Wilson refused to interfere. Carranza emerged victoriously by the summer of 1915, and Wilson accorded him de facto recognition in October. But Villa, seeking to provoke war between the United States and Mexico, raided Columbus, N.M., on March 9, 1916, burning the town and killing 19 inhabitants. Wilson sent a punitive expedition under General John J. Pershing into Mexico in hot pursuit of Villa; but the wily guerrilla eluded Pershing, and, the deeper the U.S. forces penetrated into Mexican territory, the more agitated the Carranza government became. There were two serious skirmishes between regular Mexican and U.S. troops in the spring, and full-scale war was averted only when Wilson withdrew Pershing's column some months later. Relations between the two governments were greatly improved when Wilson extended de jure recognition to Carranza's new Constitutional regime in April 1917. Thereafter, Wilson adamantly rejected all further foreign and American suggestions for intervention in Mexico.

Outbreak of World War I

The struggle for neutrality. The outbreak of general war in Europe in August 1914 raised grave challenges to Wilson's skill and leadership in foreign affairs. In spite of the appeals of propagandists for the rival Allies and Central Powers, the great majority of Americans were doggedly neutral and determined to avoid involvement unless American rights and interests were grossly violated. This, too, was Wilson's own feeling, and in August he issued an official proclamation of neutrality and two weeks later appealed to Americans to be "impartial in thought as well as in action."

Loans and supplies for the Allies. Difficulties arose first with the British government, which at once used its vast fleet to establish a long-range blockade of Germany. The U.S. State Department sent several strong protests to London, particularly against British suppression of American exports of food and raw materials to Germany. Anglo-American blockade controversies were not acute, however, because the British put their blockade controls into effect gradually, always paid for goods seized, argued persuasively that in a total war food and raw materials were as essential as guns and ammunition, and pointed out that they, the British, were simply following blockade precedents established by the United States itself during the American Civil War. As a result of a tacit Anglo-American agreement, the United States soon became the chief external source of supply for the food, raw materials, and munitions that fed the British and French war machines. In addition, and in accordance with the strict rules of neutrality, the Wilson administration permitted the Allied governments to borrow more than $2,000,000,000 in the United States in order to finance the war trade. At the same time, the president resisted all efforts by German Americans for an arms embargo on the ground that such a measure would be grossly unneutral toward the Allies.

German submarine warfare. There was no possibility of conflict between Germany and the United States so long as the former confined its warfare to the continent of Europe; a new situation full of potential danger arose, however, when the German authorities decided to use their new weapon, the submarine, to challenge British control of the seas. The German admiralty announced in February 1915 that all Allied vessels would be torpedoed without warning in a broad area and that even neutral vessels were not safe. Wilson replied at once that he would hold Germany to "strict accountability" (a conventional diplomatic term) if submarines destroyed American ships and lives without warning. The Germans soon gave broad guarantees concerning American ships, and their safety against illegal submarine attacks was not an issue between the two countries before 1917.

The sinking of unarmed ships

An issue much more fraught with danger was the safety of Americans traveling and working on Allied ships. A German submarine sank the unarmed British liner *Lusitania* without warning on May 7, 1915, killing, among others, 128 Americans. Wilson at first appealed to the Germans on broad grounds of humanity to abandon submarine warfare, but in the subsequent negotiations he narrowed the issue to one of safety for unarmed passenger liners against violent underseas attack. Momentary resolution came when a submarine sank the unarmed British liner *Arabic* in August. Wilson warned that he would break diplomatic relations if such attacks continued, and the Germans grudgingly promised not to attack unarmed passenger ships without warning. The controversy escalated to a more dangerous level when a submarine torpedoed the packet steamer *Sussex* in the English Channel with heavy loss of life in March 1916. In an ultimatum to Berlin, Wilson threatened to break diplomatic relations if the Germans did not cease attacking liners and merchantmen without warning; once again the Germans capitulated, but they threatened to resume unrestricted submarine warfare if the United States failed to force the British to observe international law in their blockade practices.

The Allies complicated the submarine controversy in late 1915 by arming many of their liners and merchantmen sailing to American ports. Wilson tried to arrange a compromise by which the Allies would disarm their ships in return for a German promise not to sink them without warning. When the British rejected the proposal, the president gave the impression that he would hold Germany accountable for American lives lost on armed ships, setting off a rebellion in Congress and the near passage of resolutions forbidding American citizens to travel on armed ships. Actually, the president had no intention of permitting armed ships to become a serious issue; their status was never a subject of serious controversy between the United States and Germany.

Arming for war. Meanwhile, the increasingly perilous state of relations with Germany had prompted Wilson, in December 1915, to call for a considerable expansion in the nation's armed forces. A violent controversy over preparedness ensued, both in Congress and in the country at large. The army legislation of 1916 was a compromise, with Wilson obtaining only a modest increase in the army and a strengthening of the National Guard; but the Naval Appropriations Act of 1916 provided for more ships than the administration had requested.

The United States enters the Great War. Wilson's most passionate desire, aside from avoiding belligerency, was to bring an end to the war through his personal mediation. He sent Colonel House to Europe in early 1915 to explore the possibilities of peace and again early in 1916 to press for a plan of Anglo-American cooperation for peace. The British refused to cooperate, and the president, more than ever eager to avoid a final confrontation with Germany on the submarine issue, decided to press forward with independent mediation. He was by this time also angered by the intensification of British blockade practices and convinced that both sides were fighting for world domination and spoils. On Dec. 18, 1916, Wilson asked the belligerents to state the terms upon which they would be willing to make peace. Soon afterward, in secret, high-level negotiations, he appealed to Britain and Germany to hold an early peace conference under his leadership.

Wilson's attempts to achieve peace

Break with Germany. Chances for peace were blasted by a decision of the German leaders, made at an imperial conference on Jan. 9, 1917, to inaugurate an all-out submarine war against all commerce, neutral as well as belligerent. The Germans knew that such a campaign would bring the United States into the war; but they were confident that their augmented submarine fleet could starve Britain into submission before the United States could mobilize and participate effectively.

The announcement of the new submarine blockade in January left the president no alternative but to break diplomatic relations with Germany, which he did on February 3. At the same time, and in subsequent addresses, the

president made it clear that he would accept unrestricted submarine warfare against belligerent merchantmen and would act only if American ships were sunk. In early March he put arms on American ships in the hope that this would deter submarine attacks. The Germans began to sink American ships indiscriminately in mid-March, and on April 2 Wilson asked Congress to recognize that a state of war existed between the United States and the German Empire. Congress approved the war resolution quickly, and Wilson signed it on April 6. (For U.S. military involvement in World War I, see the article WORLD WARS.)

Mobilization. Generally speaking, the efforts at mobilization went through two stages. During the first, lasting roughly from April to December 1917, the administration relied mainly on voluntary and cooperative efforts. During the second stage, after December 1917, the government moved rapidly to establish complete control over every important phase of economic life. Railroads were nationalized; a war industries board established ironclad controls over industry; food and fuel were strictly rationed; an emergency-fleet corporation began construction of a vast merchant fleet; and a war labour board used coercive measures to prevent strikes. Opposition to the war was sternly suppressed under the Espionage Act of 1917. At the same time, the Committee on Public Information, headed by the progressive journalist George Creel, mobilized publicists, scholars, and others in a vast prowar propaganda effort. By the spring of 1918, the American people and their economy had been harnessed for total war (a near miracle, considering the lack of preparedness only a year before).

Government controls during the war

America's role in the war. The American military contribution, while small compared to that of the Allies during the entire war, was in two respects decisive in the outcome. The U.S. Navy, fully prepared at the outset, provided the ships that helped the British overcome the submarine threat by the autumn of 1917. The U.S. Army, some 4,000,000 men strong, was raised mainly by conscription under the Selective Service Act of 1917; the American Expeditionary Force of more than 1,200,000 men under General Pershing reached France by September 1918, and this huge infusion of manpower tipped the balance on the Western Front and helped to end the war in November 1918, a year earlier than military planners had anticipated.

Wilson's vision of a new world order. In one of the most ambitious rhetorical efforts in modern history, President Wilson attempted to rally the people of the world in a movement for a peace settlement that would remove the causes of future wars and establish machinery to maintain peace. In an address to the Senate on Jan. 22, 1917, he called for a "peace without victory" to be enforced by a league of nations that the United States would join and strongly support. He reiterated this program in his war message, adding that the United States wanted above all else to "make the world safe for democracy." And when he failed to persuade the British and French leaders to join him in issuing a common statement of war aims, he went to Congress on Jan. 8, 1918, to make, in his Fourteen Points address, his definitive avowal to the American people and the world.

Wilson's Fourteen Points

In his general points Wilson demanded an end to the old diplomacy that had led to wars in the past. He proposed open diplomacy instead of entangling alliances, and he called for freedom of the seas, an impartial settlement of colonial claims, general disarmament, removal of artificial trade barriers, and, most important, a league of nations to promote peace and protect the territorial integrity and independence of its members. On specific issues he demanded, among other things, the restoration of a Belgium ravaged by the Germans; sympathetic treatment of the Russians, then involved in a civil war; establishment of an independent Poland; the return of Alsace-Lorraine to France; and autonomy or self-determination for the subject peoples of the Austro-Hungarian and Ottoman empires. A breathtaking pronouncement, the Fourteen Points gave new hope to millions of liberals and moderate socialists who were fighting for a new international order based upon peace and justice.

The Paris Peace Conference and the Versailles Treaty. With their armies reeling under the weight of a combined Allied and American assault, the Germans appealed to Wilson in October 1918 for an armistice based on the Fourteen Points and other presidential pronouncements. The Allies agreed to conclude peace on this basis, except that the British entered a reservation about freedom of the seas, and Wilson agreed to an Anglo-French demand that the Germans be required to make reparation for damages to civilian property.

Wilson led the U.S. delegation and a large group of experts to the peace conference, which opened in Paris in January 1919. He fought heroically for his Fourteen Points against the Allied leaders—David Lloyd George of Britain, Georges Clemenceau of France, and Vittorio Orlando of Italy—who, under heavy pressure from their own constituencies, were determined to divide the territories of the vanquished and make Germany pay the full cost of the war. Wilson made a number of compromises that violated the spirit if not the letter of the Fourteen Points, including the imposition of an indefinitely large reparations bill upon Germany. Moreover, the Allies had intervened in the Russian Civil War against the dominant revolutionary socialist faction, the Bolsheviks; and Wilson had halfheartedly cooperated with the Allies by dispatching small numbers of troops to northern Russia, to protect military supplies against the advancing Germans, and to Siberia, mainly to keep an eye on the Japanese, who had sent a large force there. But Wilson won many more of his Fourteen Points than he lost; his greatest victories were to prevent the dismemberment of Germany in the west and further intervention in Russia and, most important, to obtain the incorporation of the Covenant of the League of Nations into the Versailles Treaty. He was confident that the League, under American leadership, would soon rectify the injustices of the treaty.

The League of Nations

The fight over the treaty and the election of 1920. Public opinion in the United States seemed strongly in favour of quick ratification of the Versailles Treaty when the president presented that document to the Senate in July 1919. Traditional isolationist sentiment was beginning to revive, however, and a small minority of 16 senators, irreconcilably opposed to U.S. membership in the League, vowed to oppose the treaty to the bitter end. In addition, a crucial controversy developed between the president and a majority of the Republican senators, led by Henry Cabot Lodge of Massachusetts. Lodge insisted upon adding 14 reservations to the treaty. The second reservation declared that the United States assumed no obligations under Article X of the Covenant, which guaranteed the integrity and independence of members of the League; moreover it said that the president could not use the armed forces to support the Covenant without the explicit consent of Congress.

Calling this reservation a nullification of the treaty, Wilson in September made a long speaking tour of the West to build up public support for unconditional ratification. He suffered a breakdown at the end of his tour and a serious stroke on October 2. The president's illness, which incapacitated him for several months, increased his intransigence against the Lodge reservations; with equal stubbornness, the Massachusetts senator refused to consent to any compromise. The result was failure to obtain the necessary two-thirds majority for ratification, with or without reservations, when the Senate voted on Nov. 19, 1919, and again on March 19, 1920.

Wilson had suggested that the ensuing presidential campaign and election should be a "great and solemn referendum" on the League. The Democratic candidate, James M. Cox of Ohio, fought hard to make it the leading issue; but the Republican candidate, Warren G. Harding of Ohio, was evasive on the subject, and a group of 31 leading Republican internationalists assured the country that Harding's election would be the best guarantee of U.S. membership in the League of Nations. Harding swamped Cox, and his victory ended all hopes for U.S. membership. In his inaugural Harding announced that the United States would not be entangled in European affairs; he emphasized this determination by concluding a separate peace with Germany in 1921. (A.S.L.)

The United States from 1920 to 1945

THE POSTWAR REPUBLICAN ADMINISTRATIONS

Postwar conservatism. After the end of World War I, many Americans were left with a feeling of distrust toward foreigners and radicals, whom they held responsible for the war. The Russian Revolution of 1917 and the founding of the communists' Third International in 1919 further fanned American fears of radicalism. Race riots and labour unrest added to the tension. Thus, when a series of strikes and indiscriminate bombings began in 1919, the unrelated incidents were all assumed—incorrectly in most cases—to be communist-inspired. During the ensuing Red Scare, civil liberties were sometimes grossly violated and many innocent aliens were deported. The Red Scare was over within a year, but a general distrust of foreigners, liberal reform movements, and organized labour remained throughout the 1920s. In fact, many viewed Harding's landslide victory in 1920 as a repudiation of Wilson's internationalism and of the reforms of the Progressive era.

Peace and prosperity. Harding took office with a clear mandate to restore business as usual, a condition he termed "normalcy." Americans wished to put reminders of the Great War behind them, as well as the brutal strikes, the Red Scare, and the sharp recession of Wilson's last years in office. Peace and prosperity were what people desired, and these would be achieved under Harding.

The return to "normalcy"

As part of his policy of returning America to prewar conditions, Harding pardoned many individuals who had been convicted of antiwar activities or for being radicals. His main concern, however, was business. Reversing progressive and wartime trends, the Harding administration strove to establish probusiness policies. Attorney General Harry M. Daugherty obtained injunctions against striking workers. The Supreme Court sided with management in disputes over unions, minimum wage laws, child labour, and other issues. Secretary of Commerce Herbert Hoover expanded the size of his department fourfold during the next eight years in attempts to foster business growth and efficiency and to encourage trade associations and business–labour cooperation. Secretary of the Treasury Andrew W. Mellon, one of the nation's richest men, drastically cut taxes, especially on the wealthy; he also cut federal spending to reduce the national debt.

In foreign affairs the Harding administration tried to ensure peace by urging disarmament, and at the Washington Naval Conference in 1921 Secretary of State Charles Evans Hughes negotiated the first effective arms-reduction agreement in history. On the whole, however, the policies of the United States were narrow and nationalistic. It did not cooperate with the League of Nations. It insisted that Europeans pay their American debts but in 1922 passed the Fordney–McCumber Tariff, which raised duties so high that foreigners had great difficulty earning the necessary dollars. When immigration reached prewar levels (some 800,000 people entered the country between June 1920 and June 1921), Congress gave in to the protests of organized labour, which believed immigrants were taking jobs away from American citizens, and to the objections of business leaders and patriotic organizations, who feared that some of the immigrants might be radicals. Reversing traditional American policy, Congress passed first an emergency restriction bill and then in 1924 the National Origins Act. The act set a quota limiting the number of immigrants to 164,000 annually (150,000 after July 1, 1927); it discriminated against immigrants from southern and eastern Europe and barred Asians completely. The quota did not pertain to North Americans, however.

Harding's policies, his genial nature, and the return of prosperity made the president extremely popular. His sudden death, of a cerebral embolism, in the summer of 1923 resulted in a national outpouring of grief. Yet it soon became evident that his administration had been the most corrupt since Grant's. Harding had appointed venal mediocrities, many of them old cronies, to office, and they had betrayed his trust. The most publicized scandal was the illegal leasing of naval oil reserves at Teapot Dome, Wyo., which led to the conviction of Secretary of the Interior Albert B. Fall for accepting a bribe.

The Teapot Dome scandal

Calvin Coolidge, Harding's vice president and successor, was a taciturn, parsimonious New Englander who restored honesty to government. His administration suffered none of the stigma of the Harding scandals, and Coolidge, thanks to a buoyant economy and a divided Democratic Party, easily defeated the conservative Democrat John W. Davis in the election of 1924. Even though an independent campaign by Senator Robert M. La Follette of Wisconsin drew off insurgent Republicans, Coolidge received more popular, and electoral, votes than his opponents combined.

Coolidge followed Harding's policies, and prosperity continued for most of the decade. From 1922 to 1929, stock dividends rose by 108 percent, corporate profits by 76 percent, and wages by 33 percent. In 1929, 4,455,100 passenger cars were sold by American factories, one for every 27 members of the population, a record that was not broken until 1950. Productivity was the key to America's economic growth. Because of improvements in technology, overall labour costs declined by nearly 10 percent, even though the wages of individual workers rose.

The prosperity was not solidly based, however. The wealthy benefited most, and agriculture and several industries, such as textiles and bituminous coal mining, were seriously depressed; after 1926 construction declined.

New social trends. For millions of Americans, the sober-minded Coolidge was a more appropriate symbol for the era than the journalistic terms Jazz Age or Roaring Twenties. These terms were exaggerations, but they did have some basis in fact. Many young men and women who had been disillusioned by their experiences in World War I rebelled against what they viewed as unsuccessful, outmoded prewar conventions and attitudes. Women who had been forced to work outside the home because of labour shortages during the war were unwilling to give up their social and economic independence after the war had ended. Having won the right to vote when the Nineteenth Amendment was ratified in 1920, the new "emancipated" woman, the flapper, demanded to be recognized as man's equal in all areas. She adopted a masculine look, bobbing her hair and abandoning corsets; she drank and smoked in public; and she was more open about sex.

The Jazz Age

Social changes were not limited to the young. Productivity gains brought most Americans up to at least a modest level of comfort. People were working fewer hours a week and earning more money than ever before. New consumer goods—radios, telephones, refrigerators, and above all the motor car—made life better, and they were easier to buy thanks to a vastly expanded consumer credit system. Leisure activities became more important, professional sports boomed, and the rapid growth of tabloid newspapers, magazines, movies, and radios enabled millions to share in the exciting world of speakeasies, flappers, and jazz music, even if only vicariously.

On the darker side, antiforeign sentiment led to the revival of the racist, anti-Semitic, and anti-Catholic Ku Klux Klan, especially in rural areas. During the early 1920s the Klan achieved a membership of some 5,000,000 and gained control of, or influence over, many city and state governments. Rural areas also provided the base for a Christian fundamentalist movement, as farmers and small-town dwellers who felt threatened and alienated by the rapidly expanding, socially changing cities fought to preserve American moral standards by stressing religious orthodoxy. The movement grew steadily until 1925, when John T. Scopes, a biology teacher in Dayton, Tenn., was tried for violating a law common to many Southern states prohibiting the teaching of the theory of evolution. Although Scopes was found guilty of breaking the law, both the law itself and fundamentalist beliefs were ridiculed during the course of the trial, which attracted national attention.

One fundamentalist goal that was achieved was the passage in 1919 of the Prohibition (Eighteenth) Amendment, which prohibited the manufacture, sale, or transportation of intoxicating liquors. Millions of mostly Protestant churchgoers hailed Prohibition as a moral advance, and the liquor consumption of working people, as well as the incidence of alcohol-related diseases and deaths, does seem to have dropped during the period. On the other

Prohibition

hand, millions of otherwise law-abiding citizens drank the prohibited liquor, prompting the growth of organized crime. The illegal liquor business was so lucrative and federal prohibition enforcement machinery was so slight that gangsters were soon engaged in the large-scale smuggling, manufacture, and sale of alcoholic beverages.

As in legitimate business, the highest profits came from achieving economies of scale, so gangsters engaged in complex mergers and takeovers; but, unlike corporate warfare, the underworld used real guns to wipe out competition. In 1931 a national law-enforcement commission, formed to study the flouting of prohibition and the activities of gangsters, was to report that prohibition was virtually unenforceable; and, with the coming of the Great Depression, prohibition ceased to be a key political issue. In 1933 the Twenty-first Amendment brought its repeal.

In the meantime, prohibition and religion were the major issues of the 1928 presidential campaign between the Republican nominee, Herbert Hoover, and the Democrat, Governor Alfred E. Smith of New York. Smith was an opponent of prohibition and a Roman Catholic. His candidacy brought enthusiasm and a heavy Democratic vote in the large cities, but a landslide against him in the dry and Protestant hinterlands secured the election for Hoover.

The stock market crash of 1929

The Great Depression. In October 1929, only months after Hoover took office, the stock market crashed, the average value of 50 leading stocks falling by almost half in two months. Despite occasional rallies, the slide persisted until 1932, when stock averages were barely a fourth of what they had been in 1929. Industrial production soon followed the stock market, giving rise to the worst unemployment the country had ever seen. By 1933 at least a quarter of the work force was unemployed. Adjusted for deflation, salaries had fallen by 40 percent and industrial wages by 60 percent.

The causes of the Great Depression were many and various. Agriculture had collapsed in 1919 and was a continuing source of weakness. Because of poor regulatory policies, many banks were overextended. Wages had not kept up with profits, and by the late 1920s consumers were reaching the limits of their ability to borrow and spend. Production had already begun to decline and unemployment to rise before the crash. The crash, which was inevitable since stock prices were much in excess of real value, greatly accelerated every bad tendency, destroying the confidence of investors and consumers alike.

Hoover met the crisis energetically, in contrast to earlier administrations, which had done little to cope with panics except reduce government spending. He extracted promises from manufacturers to maintain production. He signed legislation providing generous additional sums for public works. He also signed the infamous Smoot–Hawley Tariff of 1930, which raised duties by as much as 50 percent. These steps failed to ease the depression, however, while the tariff helped to export it. International trade had never recovered from World War I. Europe still depended on American sales and investments for income and on American loans to maintain the complicated structure of debt payments and reparations erected in the 1920s. After the crash Americans stopped investing in Europe, and the tariff deprived foreigners of their American markets. Foreign nations struck back with tariffs of their own, and all suffered from the resulting anarchy.

In the 1930 elections the Democratic Party won control of the House of Representatives and, in combination with liberal Republicans, the Senate as well. Soon afterward a slight rise in production and employment made it seem that the worst of the depression was over. Then, in the spring of 1931, another crisis erupted. The weakening western European economy brought down a major bank in Vienna, and Germany defaulted on its reparations payments. Hoover proposed a one-year moratorium on reparations and war-debt payments, but, even though the moratorium was adopted, it was too little too late. In the resulting financial panic most European governments went off the gold standard and devalued their currencies, thus destroying the exchange system, with devastating effects upon trade. Europeans withdrew gold from American banks, leading the banks to call in their loans to American businesses. A cascade of bankruptcies ensued, bank customers collapsing first and after them the banks.

Hoover tried hard to stabilize the economy. He persuaded Congress to establish a Reconstruction Finance Corporation to lend funds to banks, railroads, insurance companies, and other institutions. At the same time, in January 1932, new capital was arranged for federal land banks. The Glass–Steagall Act provided gold to meet foreign withdrawals and liberalized Federal Reserve credit. The Federal Home Loan Bank Act sought to prop up threatened building and loan associations. But these measures failed to promote recovery or to arrest the rising tide of unemployment. Hoover, whose administrative abilities had masked severe political shortcomings, made things worse by offering negative leadership to the nation. His public addresses were conspicuously lacking in candor. He vetoed measures for direct federal relief, despite the fact that local governments and private charities, the traditional sources for welfare, were clearly incapable of providing adequate aid for the ever-rising numbers of homeless and hungry. When unemployed veterans refused to leave Washington after their request for immediate payment of approved bonuses was denied, Hoover sent out the army, which dispersed the protesters at bayonet point and burned down their makeshift quarters.

Anti-depression measures

Hoover's failures and mistakes guaranteed that whoever the Democrats nominated in 1932 would become the next president. Their candidate was Governor Franklin Delano Roosevelt of New York. He won the election by a large margin, and the Democrats won majorities in both branches of Congress.

THE NEW DEAL

The first New Deal. Roosevelt took office amid a terrifying bank crisis that had forced many states to suspend banking activities. He acted quickly to restore public confidence. On Inaugural Day, March 4, 1933, he declared that "the only thing we have to fear is fear itself." The next day he halted trading in gold and declared a national "bank holiday." On March 9 he submitted to Congress an Emergency Banking Bill authorizing government to strengthen, reorganize, and reopen solvent banks. The House passed the bill by acclamation, sight unseen, after only 38 minutes of debate. That night the Senate passed it unamended, 73 votes to 7. On March 12 Roosevelt announced that, on the following day, sound banks would begin to reopen. On March 13, deposits exceeded withdrawals in the first reopened banks. "Capitalism was saved in eight days," Raymond Moley, a member of the president's famous "brain trust," later observed.

In fact, the legal basis for the bank holiday was doubtful. The term itself was a misnomer, intended to give a festive air to what was actually a desperate last resort. Most of the reopened banks were not audited to establish their solvency; instead the public was asked to trust the president. Nevertheless, the bank holiday exemplified brilliant leadership at work. It restored confidence where all had been lost and saved the financial system. Roosevelt followed it up with legislation that did actually put the banking structure on a solid footing. The Glass–Steagall Act of 1933 separated commercial from investment banking and created the Federal Deposit Insurance Corporation to guarantee small deposits. The Banking Act of 1935 strengthened the Federal Reserve System, the first major improvement since its birth in 1913.

With the country enthusiastically behind him, Roosevelt kept Congress in special session and piece by piece sent it recommendations that formed the basic recovery program of his first 100 days in office. From March 9 to June 16, 1933, Congress enacted all of Roosevelt's proposals. Among the bills passed was one creating the Tennessee Valley Authority, which would build dams and power plants and in many other ways salvage a vast, impoverished region. The Securities Exchange Act gave the Federal Trade Commission broad new regulatory powers, which in 1934 were passed on to the newly created Securities and Exchange Commission. The Home Owners Loan Act established a corporation that refinanced one of every five mortgages on urban private residences. Other bills passed

Roosevelt's recovery program

during the Hundred Days, as well as subsequent legislation, provided aid for the unemployed and the working poor and attacked the problems of agriculture and business.

Relief. Nothing required more urgent attention than the masses of unemployed workers who, with their families, had soon overwhelmed the miserably underfinanced bodies that provided direct relief. On May 12, 1933, Congress established a Federal Emergency Relief Administration to distribute half a billion dollars to state and local agencies. Roosevelt also created the Civil Works Administration, which by January 1934 was employing more than 4,000,000 men and women. Alarmed by rising costs, Roosevelt dismantled the CWA in 1934, but the persistence of high unemployment led him to make another about-face. In 1935 the Emergency Relief Appropriation Act provided almost $5,000,000,000 to create work for some 3,500,000 persons. The Public Works Administration (PWA), established in 1933, provided jobs on long-term construction projects, and the Civilian Conservation Corps put 2,500,000 young men to work planting or otherwise improving huge tracts of forestland. For homeowners, the Federal Housing Administration began insuring private home-improvement loans to middle-income families in 1934; in 1938 it became a home-building agency as well.

Agricultural recovery. Hoover's Federal Farm Board had tried to end the long-standing agricultural depression by raising prices without limiting production. Roosevelt's Agricultural Adjustment Act (AAA) of 1933 was designed to correct the imbalance. Farmers who agreed to limit production would receive "parity" payments to balance prices between farm and nonfarm products, based on prewar income levels. Farmers benefited also from numerous other measures, such as the Farm Credit Act of 1933, which refinanced a fifth of all farm mortgages in a period of 18 months, and the creation in 1935 of the Rural Electrification Administration (REA), which did more to bring farmers into the 20th century than any other single act. Thanks to the REA, nine out of 10 farms were electrified by 1950, compared to one out of 10 in 1935.

Agricultural Adjustment Act

These additional measures were made all the more important by the limited success of the AAA. Production did fall as intended, aided by the severe drought of 1933–36, and prices rose in consequence; but many, perhaps a majority, of farmers did not prosper as a result. The AAA was of more value to big operators than to small family farmers, who often could not meet their expenses if they restricted their output and therefore could not qualify for parity payments. The farm corporation, however, was able to slash its labour costs by cutting acreage and could cut costs further by using government subsidies to purchase machinery. Thus, even before the Supreme Court invalidated the AAA in 1936, support for it had diminished.

Business recovery. As the economic crisis was above all an industrial depression, business recovery headed the New Deal's list of priorities. Working toward that goal, the administration drafted the National Industrial Recovery Act of 1933, which, among other things, created a National Recovery Administration to help business leaders draw up and enforce codes governing prices, wages, and other matters (coded industries would be exempt from the antitrust laws). Labour was offered protection from unfair practices and given the right to bargain collectively. A large-scale public works appropriation, administered through the PWA, was intended to pour sufficient money into the economy to increase consumer buying power while prices and wages went up.

Despite great initial enthusiasm for the NRA program, it was a failure. The codes became too numerous and complex for proper enforcement, and they were resented because they tended to favour the leading producers in each regulated industry. The protections afforded labour proved illusory, while the PWA, despite an impressive building record that included not only dams, bridges, and schools but also aircraft carriers, was too slow and too small to have much effect on the economy as a whole.

Yet, even if the NRA had overcome its technical problems, failure would probably still have resulted. What the country needed was economic growth, but the NRA assumed that the United States had a mature economic structure incapable of further expansion. Accordingly, it worked to stabilize the economy, eliminate wasteful or predatory competition, and protect the rights of labour. Encouraging growth was not on its agenda.

The second New Deal and the Supreme Court. In reaction to pressures from the left and hostility from the right, the New Deal shifted more toward reform in 1935–36. Popular leaders, promising more than Roosevelt, threatened to pull sufficient votes from him in the 1936 election to bring Republican victory. Senator Huey P. Long of Louisiana was building a national following with a "Share the Wealth" program. The poor in Northern cities were attracted to the Roman Catholic priest Charles E. Coughlin, who later switched from a program of nationalization and currency inflation to an antidemocratic, anti-Semitic emphasis. Many older people supported Francis E. Townsend's plan to provide $200 per month for everyone over age 60. At the same time, conservatives, including such groups as the American Liberty League, founded in 1934, attacked the New Deal as a threat to states' rights, free enterprise, and the open shop.

The shift toward reform

Roosevelt's response in 1935 was to propose greater aid to the underprivileged and extensive reforms. Congress created the Works Progress Administration, which replaced direct relief with work relief; between 1935 and 1941 the WPA employed an annual average of 2,100,000 workers, including artists and writers, who built or improved schools, hospitals, airports, and other facilities by the tens of thousands. The National Youth Administration created part-time jobs for millions of college students, high-school students, and other youngsters. Of long-range significance was the Social Security Act of 1935, which provided federal aid for the aged, retirement annuities, unemployment insurance, aid for persons who were blind or crippled, and aid to dependent children; the original act suffered from various inadequacies, but it was the beginning of a permanent, expanding national program. A tax reform law fell heavily upon corporations and well-to-do people. The National Labor Relations Act, or Wagner Act, gave organized labour federal protection in collective bargaining; it prohibited a number of "unfair practices" on the part of employers and created the strong National Labor Relations Board to enforce the law.

In the 1936 elections Roosevelt, aided by his reform program, formed a coalition that included liberals, urban ethnics, farmers, trade unionists, and the elderly. He easily defeated the Republican nominee for president, Governor Alfred ("Alf") M. Landon of Kansas, receiving more than 60 percent of the popular vote and the electoral votes of every state except Maine and Vermont. The Democratic majorities in the House and Senate were also strengthened. Viewing his decisive victory as an electoral mandate for continued reform, Roosevelt sought to neutralize the Supreme Court, which in 1935 had invalidated several early New Deal reform measures and now seemed about to strike down the Wagner Act and the Social Security Act. In February 1937 Roosevelt created a furor by proposing a reorganization of the court system that would have included giving him the power to appoint up to six new justices, thus giving the court a liberal majority. Some Democrats and a few liberal Republicans in Congress supported the proposal, but a strong coalition of Republicans and conservative Democrats, backed by much public support, fought the so-called court-packing plan.

Roosevelt's court-packing plan

Meanwhile the court itself in a new series of decisions began upholding as constitutional measures involving both state and federal economic regulation. These decisions, which began an extensive revision of constitutional law concerning governmental regulation, made the reorganization plan unnecessary; the Senate defeated it in July 1937 by a vote of 70 to 22. Roosevelt had suffered a stinging political defeat, even though he no longer had to fear the court. Turnover on the court was rapid as older members retired or died; by 1942 all but two of the justices were Roosevelt appointees.

The culmination of the New Deal. Roosevelt lost further prestige in the summer of 1937, when the nation plunged into a sharp recession. Economists had feared an inflationary boom as industrial production moved up to

within 7.5 percent of 1929. Other indices were high except for a lag in capital investment and continued heavy unemployment. Roosevelt, fearing a boom and eager to balance the budget, cut government spending, which most economists felt had brought the recovery. The new Social Security taxes removed an additional $2,000,000,000 from circulation. Between August 1937 and May 1938 the index of production fell from 117 to 76 (on a 1929 base of 100), and unemployment increased by perhaps 4,000,000 persons. Congress voted an emergency appropriation of $5,000,000,000 for work relief and public works, and by June 1938 recovery once more was under way, although unemployment remained higher than before the recession.

Roosevelt's loss of power became evident in 1938, when his attempts to defeat conservative congressional Democrats in the primaries failed. In the fall Republicans gained 80 seats in the House and seven in the Senate. The Democratic Party retained nominal control of Congress, but conservative Democrats and Republicans voting together defeated many of Roosevelt's proposals. A few last bills slipped through. The U.S. Housing Authority was created in 1937 to provide low-cost public housing. In 1938 the Fair Labor Standards Act established a minimum wage and a maximum work week. Otherwise, the president seldom got what he asked for.

The rise of organized labour

Apart from the New Deal itself, no development in the 1930s was more important than the rise of organized labour. This too had negative, or at least mixed, effects upon Roosevelt's political power. When the depression struck, only 5 percent of the work force was unionized, compared to 12 percent in 1920. The great change began in 1935 when the American Federation of Labor's Committee for Industrial Organization broke away from its timid parent and, as the Congress of Industrial Organizations (after 1938), began unionizing the mass production industries. The CIO had a unique tool, the sit-down strike. Instead of picketing a plant, CIO strikers closed it down from inside, taking the factory hostage and preventing management from operating with nonunion workers. This, together with the new reluctance of authorities, many of them Roosevelt Democrats, to act against labour, made sit-down strikes highly successful. On Feb. 11, 1937, after a long sit-down strike, General Motors, the country's mightiest corporation, recognized the United Auto Workers. The United States Steel Corporation caved in less than a month later, and by 1941 some 10,500,000 workers were unionized, three times as many as a decade before. The CIO became a mainstay of the New Deal coalition, yet it also aroused great resentment among middle-class Americans, who opposed strikes in general but the CIO's tactics especially. This further narrowed Roosevelt's political base.

An assessment of the New Deal. The New Deal established federal responsibility for the welfare of the economy and the American people. At the time, conservative critics charged it was bringing statism or even socialism. Left-wing critics of a later generation charged just the reverse—that it bolstered the old order and prevented significant reform. Others suggested that the New Deal was no more than the extension and culmination of progressivism. In its early stages, the New Deal did perhaps begin where progressivism left off and built upon the Hoover program for fighting the depression. But Roosevelt soon took the New Deal well beyond Hoover and progressivism, establishing a precedent for large-scale social programs and for government participation in economic activities. Despite the importance of this growth of federal responsibility, the New Deal's greatest achievement was to restore faith in American democracy at a time when many people believed that the only choice left was between communism and fascism. Its greatest failure was its inability to bring about complete economic recovery. Some economists, notably John Maynard Keynes of Great Britain, were calling for massive deficit spending to promote recovery; and by 1937 the New Deal's own experience proved that pump priming worked, whereas spending cutbacks only hurt the economy. Roosevelt remained unpersuaded, however, and the depression lingered on until U.S. entry into World War II brought full employment.

WORLD WAR II

The road to war. After World War I most Americans concluded that participating in international affairs had been a mistake. They sought peace through isolation and throughout the 1920s advocated a policy of disarmament and nonintervention. As a result, relations with Latin-American nations improved substantially under Hoover, an anti-imperialist. This enabled Roosevelt to establish what became known as the Good Neighbor Policy, which repudiated altogether the right of intervention in Latin America. By exercising restraint in the region as a whole and by withdrawing American occupation forces from the Caribbean, Roosevelt increased the prestige of the United States in Latin America to its highest level in memory.

As the European situation became more tense, the United States continued to hold to its isolationist policy. Congress, with the approval of Roosevelt and Secretary of State Cordell Hull, enacted a series of neutrality laws that legislated against the factors that supposedly had taken the United States into World War I. As Italy prepared to invade Ethiopia, Congress passed the Neutrality Act of 1935, embargoing shipment of arms to either aggressor or victim. Stronger legislation followed the outbreak of the Spanish Civil War in 1936, in effect penalizing the Spanish government, whose fascist enemies were receiving strong support from Benito Mussolini and Adolph Hitler.

The policy of neutrality

In the Pacific Roosevelt continued Hoover's policy of nonrecognition of Japan's conquests in Asia. When Japan invaded China in 1937, however, he seemed to begin moving away from isolationism. He did not invoke the Neutrality Act, which had just been revised, and in October he warned that war was like a disease and suggested that it might be desirable for peace-loving nations to "quarantine" aggressor nations. He then quickly denied that his statement had any policy implications, and by December, when Japanese aircraft sank a U.S. gunboat in the Yangtze River, thoughts of reprisal were stifled by public apathy and by Japan's offer of apologies and indemnities. With strong public opposition to foreign intervention, Roosevelt concentrated on regional defense, continuing to build up the navy and signing mutual security agreements with other governments in North and South America.

When Germany's invasion of Poland in 1939 touched off World War II, Roosevelt called Congress into special session to revise the Neutrality Act to allow belligerents (in reality only Great Britain and France, both on the Allied side) to purchase munitions on a cash-and-carry basis. With the fall of France to Germany in June 1940, Roosevelt, with heavy public support, threw the resources of the United States behind the British. He ordered the War and Navy departments to resupply British divisions that had been rescued at Dunkirk minus their weaponry, and in September he agreed to exchange 50 obsolescent destroyers for 99-year leases on eight British naval and air bases in the Western Hemisphere.

The question of how much and what type of additional aid should be given to the Allies became a major issue of the election of 1940, in which Roosevelt ran for an unprecedented third term. Public opinion polls, a new influence upon decision makers, showed that most Americans favoured Britain but still wished to stay out of war. Roosevelt's opponent, Wendell Willkie, capitalized on this and rose steadily in the polls by attacking the president as a warmonger. An alarmed Roosevelt fought back, going so far as to make what he knew was an empty promise. "Your boys," he said just before the election, "are not going to be sent into any foreign wars." In truth, both candidates realized that U.S. intervention in the war might become essential, contrary to their public statements. Roosevelt won a decisive victory.

Upon being returned to office, Roosevelt moved quickly to aid the Allies. His Lend-Lease Act, passed in March 1941 after vehement debate, committed the United States to supply the Allies on credit. When Germany, on March 25, extended its war zone to include Iceland and the Denmark Straits, Roosevelt retaliated in April by extending the American Neutrality Patrol to Iceland. In July the United States occupied Iceland, and U.S. naval vessels began escorting convoys of American and Icelandic ships. That

summer Lend-Lease was extended to the Soviet Union after it was invaded by Germany. In August Roosevelt met with the British prime minister, Winston Churchill, off the coast of Newfoundland to issue a set of war aims known as the Atlantic Charter. It called for national self-determination, larger economic opportunities, freedom from fear and want, freedom of the seas, and disarmament.

The Atlantic Charter

Although in retrospect U.S. entry into World War II seems inevitable, in 1941 it was still the subject of great debate. Isolationism was a great political force, and many influential individuals were determined that U.S. aid policy stop short of war. In fact, as late as Aug. 12, 1941, the House of Representatives extended the Selective Training and Service Act of 1940 by a vote of only 203 to 202. Despite isolationist resistance, Roosevelt pushed cautiously forward. In late August the navy added British and Allied ships to its Icelandic convoys. Its orders were to shoot German and Italian warships on sight, thus making the United States an undeclared participant in the Battle of the Atlantic. During October one U.S. destroyer was damaged by a German U-boat and another was sunk. The United States now embarked on an undeclared naval war against Germany, but Roosevelt refrained from asking for a formal declaration of war. According to public opinion polls, a majority of Americans still hoped to remain neutral.

The war question was soon resolved by events in the Pacific. As much as a distant neutral could, the United States had been supporting China in its war against Japan, yet it continued to sell Japan products and commodities essential to the Japanese war effort. Then, in July 1940, the United States applied an embargo on the sale of aviation gas, lubricants, and prime scrap metal to Japan. When Japanese armies invaded French Indochina in September with the apparent purpose of establishing bases for an attack on the East Indies, the United States struck back by embargoing all types of scrap iron and steel and by extending a loan to China. Japan promptly retaliated by signing a limited treaty of alliance, the Tripartite Pact, with Germany and Italy. Roosevelt extended a much larger loan to China and in December embargoed iron ore, pig iron, and a variety of other products.

Japan and the United States then entered into complex negotiations in the spring of 1941. Neither country would compromise on the China question, however, Japan refusing to withdraw and the United States insisting upon it. Believing that Japan intended to attack the East Indies, the United States stopped exporting oil to Japan at the end of the summer. In effect an ultimatum, since Japan had limited oil stocks and no alternative source of supply, the oil embargo confirmed Japan's decision to eliminate the U.S. Pacific Fleet and to conquer Southeast Asia, thereby becoming self-sufficient in crude oil and other vital resources. By the end of November Roosevelt and his military advisers knew (through intercepted Japanese messages) that a military attack was likely; they expected it to be against the East Indies or the Philippines. To their astonishment, on December 7 Japan directed its first blow against naval and air installations in Hawaii. In a bold surprise attack, Japanese aircraft destroyed or damaged 18 ships of war at Pearl Harbor, including the entire battleship force, and 347 planes. Total U.S. casualties amounted to 2,403 dead and 1,178 wounded.

On Dec. 8, 1941, Congress with only one dissenting vote declared war against Japan. Three days later Germany and Italy declared war against the United States; and Congress, voting unanimously, reciprocated. As a result of the attack on Pearl Harbor, the previously divided nation entered into the global struggle with virtual unanimity.

Declaration of war

The United States at war. Although isolationism died at Pearl Harbor, its legacy of unpreparedness lived on. Anticipating war, Roosevelt and his advisers had been able to develop and execute some plans for military expansion, but public opinion prohibited large-scale appropriations for armament and defense. Thus, when Pearl Harbor was attacked, the United States had some 2,200,000 men under arms, but most were ill-trained and poorly equipped. Barely a handful of army divisions even approached a state of readiness. The Army Air Corps possessed only 1,100 combat planes, many of which were outdated. The navy was better prepared, but it was too small to fight a two-ocean war and had barely been able to provide enough ships for convoy duty in the North Atlantic. Eventually more than 15,000,000 men and women would serve in the armed forces, but not until 1943 would the United States be strong enough to undertake large-scale offensive operations. (For U.S. military involvement in World War II, see the article WORLD WARS.)

War production. Roosevelt had begun establishing mobilization agencies in 1939, but none had sufficient power or authority to bring order out of the chaos generated as industry converted to war production. He therefore created the War Production Board in January 1942 to coordinate mobilization, and in 1943 an Office of War Mobilization was established to supervise the host of defense agencies that had sprung up in Washington, D.C. Gradually, a priorities system was devised to supply defense plants with raw materials; a synthetic rubber industry was developed from scratch; rationing conserved scarce resources; and the Office of Price Administration kept inflation under control.

After initial snarls and never-ending disputes, by the beginning of 1944 production was reaching astronomical totals—double those of all the enemy countries combined. Hailed at the time as a production miracle, this increase was about equal to what the country would have produced in peacetime, assuming full employment. War production might have risen even higher if regulation of civilian consumption and industry had been stricter.

Scientists, under the direction of the Office of Scientific Research and Development, played a more important role in production than in any previous war, making gains in rocketry, radar and sonar, and other areas. Among the new inventions was the proximity fuse, which contained a tiny radio that detonated an artillery shell in the vicinity of its target, making a direct hit unnecessary. Of greatest importance was the atomic bomb, developed by scientists in secrecy and first tested on July 6, 1945.

Development of the atomic bomb

Financing the war. The total cost of the war to the federal government between 1941 and 1945 was about $321,000,000,000 (10 times as much as World War I). Taxes paid 41 percent of the cost, less than Roosevelt requested but more than the World War I figure of 33 percent. The remainder was financed by borrowing from financial institutions, an expensive method but one that Congress preferred over the alternatives of raising taxes even higher or making war bond purchases compulsory. In consequence the national debt increased fivefold, amounting to $259,000,000,000 in 1945. The Revenue Act of 1942 revolutionized the tax structure by increasing the number who paid income taxes from 13,000,000 to 50,000,000. At the same time, through taxes on excess profits and other sources of income, the rich were made to bear a larger part of the burden, making this the only period in modern history when wealth was significantly redistributed.

Social consequences of the war. Despite the vast number of men and women in uniform, civilian employment rose from 46,000,000 in 1940 to more than 53,000,000 in 1945. The pool of unemployed men dried up in 1943, and further employment increases consisted of women, minorities, and over- or underage males. These were not enough to meet all needs, and by the end of the year a manpower shortage had developed.

One result of this shortage was that blacks made significant social and economic progress. Although the armed forces continued to practice segregation, as did Red Cross blood banks, Roosevelt, under pressure from blacks, who were outraged by the refusal of defense industries to integrate their labour forces, signed Executive Order 8802 on June 25, 1941. It prohibited racial discrimination in job training programs and by defense contractors and established a Fair Employment Practices Committee to insure compliance. By the end of 1944 nearly 2,000,000 blacks were at work in defense industries. As black contributions to the military and industry increased, so did their demands for equality. This sometimes led to racial hostilities, as on June 20, 1943, when mobs of whites invaded the black section of Detroit. Nevertheless, the gains offset the losses. Lynching virtually died out, several states outlawed

discriminatory voting practices, and others adopted fair employment laws.

Full employment also resulted in raised income levels, which, through a mixture of price and wage controls, were kept ahead of inflation. Despite both this increase in income and a no-strike pledge given by trade union leaders after Pearl Harbor, there were numerous labour actions. Workers resented wage ceilings because much of their increased income went to pay taxes and was earned by working overtime rather than through higher hourly rates. In consequence, there were almost 15,000 labour stoppages during the war at a cost of some 36,000,000 man-days. Strikes were greatly resented, particularly by the armed forces, but their effects were more symbolic than harmful. The time lost amounted to only one-ninth of 1 percent of all hours worked.

Because Pearl Harbor had united the nation, few people were prosecuted for disloyalty or sedition, unlike during World War I. The one glaring exception to this policy was the scandalous treatment of Japanese and Americans of Japanese descent. In 1942, on the basis of groundless racial fears and suspicions, virtually the entire Japanese-American population of the West Coast, amounting to 110,000 persons, was rounded up and imprisoned in "relocation" centres, which the inmates regarded as concentration camps. The Japanese-Americans lost their liberty, and in most cases their property as well, despite the fact that the Federal Bureau of Investigation, which had already arrested those individuals it considered security risks, had verified their loyalty.

The relocation of Japanese-Americans

The 1944 election. Roosevelt soundly defeated Governor Thomas E. Dewey of New York in the 1944 election, but his margin of victory was smaller than it had been previously. His running mate, chosen by party leaders, who disliked former vice president Henry A. Wallace for his extreme liberalism, was Senator Harry S. Truman of Missouri, a party Democrat who had distinguished himself by investigating fraud and waste among war contractors.

The new U.S. role in world affairs. The U.S. entry into World War II had brought an end to isolation, and President Roosevelt was determined to prevent a retreat into isolationism once the war was over. After a series of conferences in December 1941, Roosevelt and Prime Minister Churchill announced the formation of the United Nations, a wartime alliance of 26 nations. In 1943 Roosevelt began planning the organization of a postwar United Nations, meeting with congressional leaders to assure bipartisan support. The public supported Roosevelt's efforts, and that fall Congress passed resolutions committing the United States to membership in an international body "with power adequate to establish and to maintain a just and lasting peace." Finally, in the spring of 1945, delegates from 50 nations signed the charter for a permanent United Nations. In addition to political harmony, Roosevelt promoted economic cooperation, and, with his full support, in 1944 the World Bank and the International Monetary Fund were created to bar a return of the cutthroat economic nationalism that had prevailed before the war.

Throughout the war Roosevelt met with Churchill and Stalin to plan military strategy and postwar policy. His last great conference with them took place at Yalta in the Crimea in February 1945. There policies were agreed upon to enforce the unconditional surrender of Germany, to divide it into zones for occupation and policing by the respective Allied forces, and to provide democratic regimes in eastern European nations. A series of secret agreements were also made at Yalta; chief among these was the Soviet pledge to enter the war against Japan after the German surrender, in return for concessions in East Asia.

Roosevelt died suddenly of a cerebral hemorrhage on April 12 and was succeeded by Truman. In the following months the German armed forces collapsed, and on May 7 all German forces surrendered. In the Pacific the invasions of Iwo Jima and Okinawa in early 1945 brought Japan under a state of siege. In the summer, before an invasion could take place, the United States dropped atomic bombs on Hiroshima and Nagasaki. On September 2 the surrender of Japan was signed in Tokyo harbour on the battleship *Missouri*. (F.Fr./W.L.O'N.)

The end of the war

The United States since 1945

THE PEAK COLD WAR YEARS, 1945–60

The Truman Doctrine and containment. Truman, who had been chosen as vice president for domestic political reasons, was poorly prepared to assume the presidency. He had no experience of foreign affairs, knew little about Roosevelt's intentions, and was intimidated by the giant shoes he now had to fill. His first decisions were dictated by events or plans already laid. In July, two months after the German forces surrendered, he met at Potsdam, Ger., with Stalin and Churchill (who was succeeded at the conference by Clement Attlee) to discuss future operations against Japan and a peace settlement for Europe. Little was accomplished, and there would not be another meeting between Soviet and American heads of state for 10 years.

Hopes that good relations between the superpowers would ensure world peace soon faded as a result of the Stalinization of eastern Europe and Soviet support of communist insurgencies in various parts of the globe. Events came to a head in 1947 when Britain, weakened by a failing economy, decided to pull out of the eastern Mediterranean. This would leave both Greece, where a communist-inspired civil war was raging, and Turkey to the mercies of the Soviet Union. Truman now came into his own as a national leader, asking Congress to appropriate aid to Greece and Turkey and asserting, in effect, that henceforth the United States must help free peoples in general to resist communist aggression. This policy, known as the Truman Doctrine, has been criticized for committing the United States to the support of unworthy regimes and for taking on greater burdens than it was safe to assume. At first, however, the Truman Doctrine was narrowly applied. Congress appropriated $400,000,000 for Greece and Turkey, saving both from falling into unfriendly hands, and thereafter the United States relied mainly on economic assistance to support its foreign policy.

The keystone of this policy, and its greatest success, was the European Recovery Program, usually called the Marshall Plan. Europe's economy had failed to recover after the war, its paralysis being worsened by the exceptionally severe winter of 1946–47. Thus, in June 1947 Secretary of State George C. Marshall proposed the greatest foreign-aid program in world history in order to bring Europe back to economic health. In 1948 Congress created the Economic Cooperation Administration and over the next five years poured some $13,000,000,000 worth of aid into western Europe. (Assistance was offered to Eastern-bloc countries also, but they were forced by Stalin to decline.) The plan restored economic vitality and confidence to the region, while undermining the local communist parties. In 1949 Truman proposed extending similar aid to underdeveloped nations throughout the world, but the resulting Point Four Program was less successful than the Marshall Plan. Experience showed that it was easier to rebuild a modern industrial economy than to develop one from scratch.

The Marshall Plan

U.S. policy for limiting Soviet expansion had developed with remarkable speed. Soon after the collapse of hopes for world peace in 1945 and 1946, the Truman administration had accepted the danger posed by Soviet aggression and resolved to shore up noncommunist defenses at their most critical points. This policy, known as containment, a term suggested by its principal framer, George Kennan of the State Department, resulted in the Truman Doctrine and the Marshall Plan, as well as in the decision to make West Germany a pillar of strength. When the Soviet Union countered this development in June 1948 by blocking all surface routes into the western-occupied zones of Berlin, Britain and the United States supplied the sectors by air for almost a year until the Soviet Union called off the blockade. A logical culmination of U.S. policy was the creation in 1949 of the North Atlantic Treaty Organization (NATO), a military alliance among 12 (later 16) nations to resist Soviet aggression.

Containment worked less well in Asia. In December 1945 Truman sent General Marshall to China with instructions to work out an agreement between the Communist rebels and the Nationalist government of Chiang Kai-shek. This was an impossible task, and in the subsequent

fighting Mao Zedong's Communist forces prevailed. The Nationalist government fled to Taiwan in 1949, and the United States then decided to concentrate its East Asian policy upon strengthening occupied Japan, with much better results.

Postwar domestic reorganization. *Military and nuclear policies.* After the end of World War II the vast U.S. military establishment was dismantled, its strength falling from 12,000,000 men and women to about 1,500,000 in 1947. The navy and army air forces remained the world's strongest, however, and the U.S. monopoly of atomic weapons seemed to ensure security. In 1946 the United States formed an Atomic Energy Commission for purposes of research and development. The armed forces were reorganized under a secretary of defense by the National Security Act of 1947, which also created the U.S. Air Force as an independent service. In 1949 the services were brought together in a single Department of Defense, though each retained considerable autonomy. In that same year the Soviet Union exploded its own atomic device, opening an era of intense nuclear, and soon thermonuclear, competition.

Social and economic development. Peace brought with it new fears. Demobilizing the armed forces might result in massive unemployment and another depression. Or, conversely, the huge savings accumulated during the war could promote runaway inflation. The first anxiety proved groundless, even though government did little to ease the transition to a peacetime economy. War contracts were canceled, war agencies diminished or dissolved, and government-owned war plants sold to private parties. But, after laying off defense workers, manufacturers rapidly tooled up and began producing consumer goods in volume. The housing industry grew too, despite shortages of every kind, thanks to mass construction techniques pioneered by the firm of Levitt and Sons, Inc., and other developers. All this activity created millions of new jobs. The Serviceman's Readjustment Act of 1944, known as the G.I. Bill of Rights, also helped ease military personnel back into civilian life. It provided veterans with loans, educational subsidies, and other benefits.

Inflation was more troublesome. Congress lacked enthusiasm for wartime price controls and in June 1946 passed a bill preserving only limited controls. Truman vetoed the bill as inadequate, controls expired, and prices immediately soared. Congress then passed an even weaker price-control bill, which Truman signed. Nevertheless, by the end of the year, most price and wage controls had been lifted. In December the Office of Price Administration began to close down. As a result the consumer price index did not stabilize until 1948, when prices were more than a third above the 1945 level, while wage and salary income had risen by only about 15 percent.

Truman's difficulties with Congress had begun in September 1945 when he submitted a 21-point domestic program, including proposals for an expansion of social security and public housing and for the establishment of a permanent Fair Employment Practices Act banning discrimination. These and subsequent liberal initiatives, later known as the Fair Deal, were rejected by Congress, which passed only the Employment Act of 1946. This clearly stated the government's responsibility for maintaining full employment and established a Council of Economic Advisers to advise the president.

The Fair Deal

Truman's relations with Congress worsened after the 1946 elections. Voters, who were angered by the price-control debacle, a wave of strikes, and Truman's seeming inability to lead or govern, gave control of both houses of Congress to Republicans for the first time since 1928. The president and the extremely conservative 80th Congress battled from beginning to end, not over foreign policy, where bipartisanship prevailed, but over domestic matters. Congress passed two tax reductions over Truman's vetoes and in 1947, again over Truman's veto, passed the Taft–Hartley Act, which restricted unions while extending the rights of management. Congress also rejected various liberal measures submitted by Truman, who did not expect the proposals to pass but wanted Congress on record as having opposed important social legislation.

By 1948, Truman had won support for his foreign policy, but he was expected to lose the presidential election that year because of his poor domestic record. Polls showed him lagging behind Dewey, again the Republican nominee, and to make matters worse the Democratic Party splintered. Former vice president Henry A. Wallace headed the Progressive Party ticket, which pledged to improve Soviet-American relations whatever the cost. Southerners, known as Dixiecrats, who were alienated by the Democratic Party's strong civil rights plank, formed the States' Rights Democratic Party and nominated Governor Strom Thurmond of South Carolina for president. These defections appeared to ensure Truman's defeat. Instead Truman won handily, receiving almost as many votes as his opponents combined. His support came largely from labour, which was upset by the Republican passage of the Taft–Hartley Act, from blacks, who strongly supported the Democrats' civil rights provisions, and from farmers, who preferred the higher agricultural subsidies promised by the Democrats, especially at a time when commodity prices were falling.

The election of 1948

The Democrats regained control of Congress in 1948, but Truman's relations with that body continued to be troubled. In January 1949 he asked for a broad range of Fair Deal measures, with uneven results. Congress did approve a higher minimum wage, the extension of social security to 10,000,000 additional persons, more public works, larger sums for the TVA and for rural electrification, and the Housing Act of 1949, which authorized construction of 810,000 units for low-income families. Truman failed, however, to persuade Congress to repeal Taft–Hartley, to reform the agricultural subsidy system, to secure federal aid to education, to adopt his civil rights program, or, most importantly, to accept his proposal for national health insurance. He succeeded nevertheless in protecting the New Deal principle of federal responsibility for social welfare, and he helped form the Democratic agenda for the 1960s.

The Red Scare. Truman's last years in office were marred by charges that his administration was lax about, or even condoned, subversion and disloyalty and that communists, called "reds," had infiltrated the government. These accusations were made despite Truman's strongly anticommunist foreign policy and his creation, in 1947, of an elaborate Federal Employee Loyalty Program, which resulted in hundreds of federal workers being fired and in several thousand more being forced to resign.

Charges of communist subversion

The excessive fear of communist subversion was fed by numerous sources. China's fall to communism and the announcement of a Soviet atomic explosion in 1949 alarmed many, and fighting between communist and U.S.-supported factions in Korea heightened political emotions as well. Real cases of disloyalty and espionage also contributed, notably the theft of atomic secrets, for which Soviet agent Julius Rosenberg and his wife Ethel were convicted in 1951 and executed two years later. Republicans had much to gain from exploiting these and related issues.

Senator Joseph R. McCarthy of Wisconsin stood out among those who held that the Roosevelt and Truman administrations amounted to "20 years of treason." In February 1950 McCarthy claimed that he had a list (whose number varied) of State Department employees who were loyal only to the Soviet Union. McCarthy offered no evidence to support his charges and revealed only a single name, that of Owen Lattimore, who was not in the State Department and would never be convicted of a single offense. Nevertheless, McCarthy enjoyed a highly successful career, and won a large personal following, by making charges of disloyalty that, though mostly undocumented, badly hurt the Democrats. Many others promoted the scare in various ways, leading to few convictions but much loss of employment by government employees, teachers, scholars, and people in the mass media.

The Korean War. On June 25, 1950, a powerful invading force from the Soviet-supported Democratic People's Republic of Korea (North Korea) swept south of the 38th parallel into the Republic of Korea (South Korea). Within days, President Truman resolved to defend South Korea, even though there were few Americans in Korea and few troops ready for combat. The UN Security Council, acting

during a Soviet boycott, quickly passed a resolution calling upon UN members to resist North Korean aggression.

After almost being driven into the sea, UN forces, made up largely of U.S. troops and commanded by U.S. General Douglas MacArthur, counterattacked successfully and in September pushed the North Korean forces back across the border. Not content with this victory, the United States attempted to unify Korea by force, advancing almost to the borders of China and the Soviet Union. China, after its warnings were ignored, then entered the war, driving the UN forces back into South Korea. The battle line was soon stabilized along the 38th parallel, and armistice talks began on July 10, 1951, three months after Truman had relieved MacArthur for openly challenging U.S. policies. The talks dragged on fruitlessly, interrupted by outbreaks of fighting, until Eisenhower became president. The United States sustained some 142,000 casualties in this limited war, most of them occurring after China's entry.

In addition to militarizing the Cold War, the Korean conflict widened its field. The United States assumed responsibility for protecting Taiwan against invasion from mainland China. Additional military aid was extended to the French in Indochina. In December 1950 Truman called for a crash program of rearmament, not just to support the forces in Korea but especially to expand the U.S. presence in Europe. As a result defense expenditures rose to $53,600,000,000 in 1953, four times the pre-Korean level, and would decline only modestly after the armistice.

Peace, growth, and prosperity. The stalemated Korean War, a renewal of inflation, and the continuing Red Scare persuaded Truman not to stand for reelection in 1952 and also gravely handicapped Governor Adlai E. Stevenson of Illinois, the Democratic nominee. His opponent, General Dwight D. Eisenhower, was an immensely popular war hero with great personal charm and no political record, making him extremely hard to attack. Although he disliked their methods, Eisenhower allowed Republican campaigners, including his running mate, Senator Richard M. Nixon of California, to capitalize on the Red Scare by accusing the Truman administration of disloyalty. Eisenhower himself charged the administration with responsibility for the communist invasion of Korea and won wide acclaim when he dramatically promised that if elected he would visit Korea in person to end the war.

The election of 1952

Eisenhower won over many farmers, ethnic whites, workers, and Roman Catholics who had previously voted Democratic. He defeated Stevenson by a large margin, carrying 39 states, including three in the once solidly Democratic South. Despite Eisenhower's overwhelming victory, Republicans gained control of the House by just eight votes and managed only a tie in the Senate. Because the Republican margin was so slight, and because many right-wing Republicans in Congress disagreed with his policies, Eisenhower would increasingly depend upon Democrats to realize his objectives.

Eisenhower had promised to end the Korean War, hold the line on government spending, balance the budget, abolish inflation, and reform the Republican Party. On July 27, 1953, an armistice was signed in Korea freezing the status quo. By cutting defense spending while taxes remained fairly high, and by keeping a tight rein on credit, Eisenhower was able to avoid serious deficits, abolish inflation, and, despite several small recessions, encourage steady economic growth that made Americans more prosperous than they had ever been before. Eisenhower also supported public works and a modest expansion of government social programs. In 1954 the St. Lawrence Seaway Development Corporation was established by Congress. In 1956 Congress authorized the National System of Interstate and Defense Highways, Eisenhower's pet project and the greatest public works program in history. Amendments to the Social Security Act in 1954 and 1956 extended benefits to millions not previously covered. Thus, Eisenhower achieved all but the last of his goals, and even in that he was at least partially successful. At first Eisenhower did little to check the Red Scare, but in 1954 Senator McCarthy unwisely began to investigate the administration and the U.S. Army. This led to a full-scale investigation of McCarthy's own activities, and on December 2 the Senate, with Eisenhower playing a behind-the-scenes role, formally censured McCarthy for abusing his colleagues. McCarthy soon lost all influence, and his fall did much to remove the poison that had infected American politics. In short, Eisenhower was so successful in restoring tranquillity that, by the end of his first term, some people were complaining that life had become too dull.

Tensions eased in foreign affairs as well. On March 5, 1953, Joseph Stalin died, opening the door to better relations with the Soviet Union. In 1955 the Soviets agreed to end the four-power occupation of Austria, and in that July Eisenhower met in Geneva with the new Soviet leader, Nikita S. Khrushchev, for talks that were friendly though inconclusive.

As for military policy, Eisenhower instituted the "New Look," which entailed reducing the army from 1,500,000 men in 1953 to 900,000 in 1960. The navy experienced smaller reductions, while air force expenditures rose. Eisenhower was primarily interested in deterring a nuclear attack and to that end promoted expensive developments in nuclear weaponry and long-range missiles.

Eisenhower's second term. Despite suffering a heart attack in 1955 and a case of ileitis that required surgery the next year, Eisenhower stood for reelection in 1956. His opponent was once again Stevenson. Two world crises dominated the campaign. On October 23, Hungarians revolted against communist rule, an uprising that was swiftly crushed by Red Army tanks. On October 29, Israel invaded Egypt, supported by British and French forces looking to regain control of the Suez Canal and, perhaps, to destroy Egypt's president, Gamal Abdel Nasser, who had nationalized the canal in July. Eisenhower handled both crises deftly, forcing the invaders to withdraw from Egypt and preventing events in Hungary from triggering a confrontation between the superpowers. Owing in part to these crises, Eisenhower carried all but seven states in the election. It was a purely personal victory, however, for the Democrats retained control of both houses of Congress.

World crises in 1956

Domestic issues. Although the Eisenhower administration can, in general, be characterized as a period of growth and prosperity, some problems did begin to arise during the second term. In 1957–58 an economic recession hit and unemployment rose to its highest level since 1941. Labour problems increased in intensity, with some 500,000 steelworkers going on strike for 116 days in 1959. There was even evidence of corruption on the Eisenhower staff. The president remained personally popular, but public discontent was demonstrated in the large majorities gained by the Democrats in the congressional elections of 1958.

Problems associated with postwar population trends also began to be recognized. The U.S. population, which had grown markedly throughout the 1950s, passed 179,000,000 in 1960. Growth was concentrated in the West, and the country became increasingly urbanized as the middle class moved from the cities to new suburban developments. The migration left cities without their tax base but with responsibility for an increasing number of poor residents. It also resulted in a huge increase in commuters, which in turn led to continuing problems of traffic and pollution.

During Eisenhower's second term, race became a central national concern for the first time since Reconstruction. Some civil rights advances had been made in previous years. In 1954 the Supreme Court had ruled that racially segregated schools were unconstitutional. The decision provoked intense resistance in the South but was followed by a chain of rulings and orders that continually narrowed the right to discriminate. In 1955 Martin Luther King, Jr., led a boycott of segregated buses in Montgomery, Ala., giving rise to the nonviolent Civil Rights movement. But neither the president nor Congress became involved in the race issue until 1957, when the segregationist governor of Arkansas blocked the integration of a high school in Little Rock. Eisenhower then sent federal troops to enforce the court's order for integration. Congress was similarly prompted to pass the first civil rights law in 82 years, the Civil Rights Act of 1960, which made a serious effort to protect black voters.

The Supreme Court decision on racial segregation

World affairs. On Oct. 4, 1957, the Soviet Union orbited the first artificial satellite, arousing fears that the

United States was falling behind the Soviets technologically. This prompted Eisenhower, who generally held the line on spending, to sign the National Defense Education Act of 1958, which provided extensive aid to schools and students in order to bring American education up to what were regarded as Soviet levels of achievement. The event also strengthened demands for the acceleration of the arms and space races, which eventually led to the U.S. Moon landing on July 20, 1969, and to a remarkable expansion of scientific knowledge. In 1958, threatened and actual conflicts between governments friendly to Western powers and unfriendly or communist forces in Lebanon, the islands of Quemoy and Matsu offshore of China, Berlin, and Cuba caused additional concern. Only a minority believed that the United States was still ahead in military and space technology, though in fact this was true.

The illness of Secretary of State John Foster Dulles in March 1959, and his subsequent resignation, led the president to increase his own activity in foreign affairs. He now traveled more and met more often with heads of state. The most important meeting was to be a summit in 1960 with Khrushchev and Western leaders to discuss such matters as Berlin, German reunification, and arms control. But two weeks before the scheduled date an American U-2 spy plane was shot down deep inside the Soviet Union. Wrangling over this incident destroyed both the Paris summit and any hopes of bettering U.S.-Soviet relations.

An assessment of the postwar era. Despite great differences in style and emphasis, the administrations of Truman and Eisenhower were notable for their continuity. Both were essentially periods of reconstruction. After 15 years of depression and war, people were not interested in social reform but in rebuilding and expanding the educational and transportation systems, achieving stable economic growth, and, in the case of the younger generation whose lives had been most disrupted by World War II, in marrying and having children. Thus, the postwar era was the age of the housing boom, the television boom, and the baby boom, of high birth and comparatively low divorce rates, of proliferating suburbs and a self-conscious emphasis upon family "togetherness." Though frustrating to social reformers, this was probably a necessary phase of development. Once the country had been physically rebuilt, the practical needs of a rapidly growing population had been met, and standards of living had risen, there would come another age of reform.

Postwar social stability

The arrival of this new age was indicated in 1960 by the comparative youth of the presidential candidates chosen by the two major parties. The Democratic nominee, Senator John F. Kennedy of Massachusetts, was 43; the Republican, Vice President Nixon, was 47. They both were ardent cold warriors and political moderates. Kennedy's relative inexperience and his religion (he was the first Roman Catholic presidential nominee since Al Smith) placed him at an initial disadvantage. But the favourable impression he created during a series of televised debates with Nixon and the support he received from blacks after he helped the imprisoned black leader Martin Luther King, Jr., enabled him to defeat Nixon in the most closely contested election of the century. (E.E.R./W.L.O'N.)

THE LATE 20TH CENTURY

The New Frontier. During the campaign Kennedy had stated that America was "on the edge of a New Frontier"; in his inaugural speech he spoke of "a new generation of Americans"; and during his presidency he seemed to be taking government in a new direction, away from the easy-going Eisenhower style. His administration was headed by strong, dedicated personalities. The Kennedy staff was also predominantly young. Its energy and commitment revitalized the nation, but its competence was soon called into question.

In April 1961 Kennedy authorized a plan that had been initiated under Eisenhower for a covert invasion of Cuba to overthrow the newly installed, Soviet-supported Communist regime of Fidel Castro. The invasion was repulsed at the Bay of Pigs, embarrassing the administration and worsening relations between the United States and the Soviet Union. These deteriorated further at a private meeting between Kennedy and Khrushchev in June 1961 when the Soviet leader was perceived as attempting to bully his young American counterpart. Relations hit bottom in October 1962 when the Soviets secretly began to install long-range offensive missiles in Cuba, which threatened to tip the balance of nuclear power. Kennedy forced the removal of the missiles, gaining back the status he had lost at the Bay of Pigs and in his meeting with Khrushchev. Kennedy then began to work toward improving international relations, and in July 1963 he concluded a treaty with Britain and the Soviet Union banning atomic tests in the atmosphere and underwater. His program of aid to Latin America, the Alliance for Progress, raised inter-American relations to their highest level since the days of Franklin Roosevelt.

The Cuban missile crisis

Kennedy's domestic policies were designed to stimulate international trade, reduce unemployment, provide medical care for the aged, reduce federal income taxes, and protect the civil rights of blacks. The latter issue, which had aroused national concern in 1962 when federal troops were employed to assure the admission of a Negro at the University of Mississippi, caused further concern in 1963, when similar action was taken at the University of Alabama and mass demonstrations were held in support of desegregation. Although the Democrats controlled both houses of Congress, the administration's proposals usually encountered strong opposition from a coalition of Republicans and Southern Democrats. With Congress' support, Kennedy was able to increase military spending substantially. This led to greater readiness but also to a significant rise in the number of long-range U.S. missiles, which prompted a similar Soviet response.

On Nov. 22, 1963, President Kennedy was assassinated in Dallas, Texas, most probably by a lone gunman, though conspiracy theories abounded. Vice President Lyndon B. Johnson took the oath of office immediately.

Johnson and the Great Society. Johnson's first job in office was to secure enactment of New Frontier bills that had been languishing in Congress. By far the most important of these was the Civil Rights Act of 1964, which Johnson pushed through despite a filibuster by Southern senators that lasted 57 days. The act provided machinery to secure equal access to accommodations, to prevent discrimination in employment by federal contractors, and to cut off funds to segregated school districts. It also authorized the Justice Department to take a more active role in civil rights cases. Johnson went beyond the New Frontier in 1964 by declaring war on poverty. His Economic Opportunity Act provided funds for vocational training, created a Job Corps to train youths in conservation camps and urban centres, encouraged community action programs, extended loans to small businessmen and farmers, and established a domestic peace corps, the counterpart of a popular foreign program created by President Kennedy.

The Civil Rights Act

Johnson provided dynamic and successful leadership at a time of national trauma, and in the election of 1964 he won a landslide victory over his Republican opponent, the conservative senator Barry Goldwater of Arizona. More importantly, the Democrats gained 38 seats in the House of Representatives that year, enough to override the conservative bloc and enact a body of liberal social legislation.

With this clear mandate, Johnson submitted the most sweeping legislative program to Congress since the New Deal. He outlined his plan for achieving a "Great Society" in his 1965 state-of-the-Union address, and over the next two years he persuaded Congress to approve most of his proposals. The Appalachian Regional Development Act provided aid for that economically depressed area. The Housing and Urban Development Act of 1965 established a Cabinet-level department to coordinate federal housing programs. Johnson's Medicare bill fulfilled President Truman's dream of providing health care for the aged. The Elementary and Secondary Education Act of 1965 provided federal funding for public and private education below the college level. The Higher Education Act of 1965 provided scholarships for more than 140,000 needy students and authorized a National Teachers Corps. The Immigration Act of 1965 abolished the discriminatory national-origins quota system. The minimum wage was

raised and its coverage extended in 1966. In 1967, social security pensions were raised and coverage expanded. The Demonstration Cities and Metropolitan Area Redevelopment Act of 1966 provided aid to cities rebuilding blighted areas. Other measures dealt with mass transit, truth in packaging and lending, beautification, conservation, water and air quality, safety, and support for the arts.

Race relations. The civil rights revolution came to a head under the Johnson administration. Despite the Civil Rights Act of 1964, most Southern blacks found it difficult if not impossible to exercise their voting rights. In 1965, mass demonstrations were held to protest the violence and other means used to prevent black voter registration. After a peaceful protest march at Selma, Ala., was violently broken up by white authorities, Johnson responded with the Voting Rights Act of 1965, which abolished literacy tests and other voter restrictions and authorized federal intervention against voter discrimination. The subsequent rise in black voter registration transformed politics in the South. Johnson's administration also forced Southern school districts to comply with court orders to integrate and authorized firm action against white racial extremists, greatly curtailing the activity of the Ku Klux Klan.

Despite these gains, many blacks remained dissatisfied by the slow rate of progress. The nonviolent Civil Rights movement was challenged by advocates of "black power," who expelled or alienated whites and crippled the movement. Race riots broke out in most of the nation's large cities, notably in the Watts district of Los Angeles in 1965, leaving 34 dead, and two years later in Newark and Detroit. Four summers of violence resulted not only in many deaths but also in immense property losses that left whole neighborhoods ruined and their inhabitants more distressed than ever. A presidential commission headed by Governor Otto Kerner of Illinois blamed the rioting on poverty and "white racism" and called for programs to alleviate black rage and despair. Whatever the causes, after a final round provoked by the assassination of Martin Luther King, Jr., in April 1968, the rioting abated.

Social changes. The 1960s were marked by the greatest changes in morals and manners since the 1920s. Young people, college students in particular, rebelled against what they viewed as the repressed, conformist society of their parents. They advocated a sexual revolution, aided by the birth control pill and later by *Roe* v. *Wade* (1973), a Supreme Court ruling that legalized abortion. "Recreational" drugs such as marijuana and LSD were increasingly used. Opposition to U.S. involvement in Vietnam promoted the rise of a New Left, which was anticapitalist as well as antiwar. A "counterculture" sprang up that legitimized radical standards of taste and behaviour in the arts as well as in life. Feminism was reborn and joined the ranks of radical causes.

The 1960s counterculture

Except for feminism, most organized expressions of the counterculture and the New Left did not long survive the sixties. Nevertheless they changed American life. Drug taking, previously confined largely to ghettos, became part of middle-class life. The sexual revolution reduced government censorship, changed attitudes toward traditional sexual roles, and enabled homosexuals to organize and acknowledge their identities as never before. Unrestrained individualism played havoc with family values. People began marrying later and having fewer children. The divorce rate accelerated to the point that the number of divorces per year was roughly half the number of marriages. The number of abortions rose and so did the illegitimacy rate. By the 1980s one in six families was headed by a single woman, and more than half of all people living in poverty, including some 12,000,000 children, belonged to such families. Because inflation and recession made it hard to support even intact families on a single income, a majority of mothers entered the work force, including a majority of those with young children. Thus the stable, family-oriented society of the 1950s became a thing of the past.

The Vietnam War. U.S. involvement in Vietnam dated to the Truman administration, when economic and military aid was provided to deter a communist takeover of French Indochina. When France withdrew and Vietnam was divided in two in 1954, the United States continued to support anticommunist forces in South Vietnam. By 1964, communist insurgents were winning their struggle against the government of South Vietnam, which a decade of American aid had failed to strengthen or reform. In August, following an allegedly unprovoked attack on U.S. warships patrolling the Gulf of Tonkin, a resolution pledging complete support for American action in Vietnam was passed unanimously in the House of Representatives and with only two dissenting votes in the Senate.

The Gulf of Tonkin resolution

After the fall elections Johnson began deploying a huge force in Vietnam (more than half a million troops in 1968, together with strong air and naval units). This power was directed not only against the Viet Cong insurgents but also against North Vietnam, which increased its efforts as American participation escalated. Despite massive U.S. bombing of North Vietnam, the communists refused to yield. On Jan. 30, 1968, disregarding a truce called for the Tet (lunar new year) holiday, the communists launched an offensive against every major urban area in South Vietnam. A military failure, in that all the attacking forces were ultimately beaten back or destroyed, the Tet Offensive was a political victory for the communists because it persuaded many Americans that the war could not be ended at a bearable price. Opposition to U.S. involvement had been steadily rising. It became widespread, and the Vietnam War became the major issue of the 1968 election. After Senator Eugene McCarthy, a leading critic of the war, ran strongly against him in the New Hampshire primary, Johnson announced that he would not seek or accept renomination. He also curtailed bombing operations, opened peace talks with the North Vietnamese, and on November 1 ended the bombing of North Vietnam.

While war efforts were being reduced, violence within the United States seemed to be growing. Just two months after King's assassination, Senator Robert F. Kennedy, a leading contender for the Democratic presidential nomination, was shot and killed by an assassin. President Johnson then secured the nomination of Vice President Hubert H. Humphrey at the Democratic National Convention at Chicago, where violence again erupted as antiwar demonstrators were manhandled by local police. Humphrey lost the election to the Republican nominee, former vice president Richard Nixon, largely because fratricidal struggles had seriously weakened the Democratic Party. The extraordinary narrowness of Nixon's margin resulted from a third-party campaign by the former governor of Alabama, George Wallace, who attracted conservative votes that would otherwise have gone to Nixon. Democrats retained large majorities in both houses of Congress.

The Richard M. Nixon administration. *Foreign affairs.* Nixon and his national security adviser, Henry Kissinger, believed that American power relative to that of other nations had declined to the point where a fundamental reorientation was necessary. They sought improved relations with the Soviet Union to make possible reductions in military strength while at the same time enhancing American security. In 1969 the Nixon Doctrine called for allied nations, especially in Asia, to take more responsibility for their own defense. Nixon's policy of détente led to Strategic Arms Limitation Talks (SALT), which resulted in a treaty with the Soviet Union all but terminating antiballistic missile systems. In 1972 Nixon and Kissinger negotiated an Interim Agreement that limited the number of strategic offensive missiles each side could deploy in the future. Nixon also dramatically reversed Sino-American relations with a secret visit by Kissinger to Peking in July 1971. This led to a presidential visit the following year and to the establishment of strong ties between the two nations. Nixon then visited Moscow as well, showing that détente with the rival communist powers did not mean that he would play them off against one another.

Nixon's policy of détente

The limits of détente were tested by the Arab-Israeli Yom Kippur War of October 1973, in which the United States supported Israel and the Soviet Union the Arabs. Nixon managed the crisis well, preventing the confrontation with the Soviets from getting out of hand and negotiating a cease-fire that made possible later improvements in Israeli-Egyptian relations. Nixon and Kissinger dramatically altered U.S. foreign relations, modifying containment, re-

ducing the importance of alliances, and making the balance of power and the dual relationship with the Soviet Union and China keystones of national policy.

Meanwhile, inconclusive fighting continued in Vietnam, and unproductive peace talks continued in Paris. Although in 1969 Nixon announced his policy of "Vietnamization," according to which more and more of the fighting was to be assumed by South Vietnam itself, he began by expanding the fighting in Southeast Asia with a 1970 "incursion" into Cambodia that would have drastic effects upon that country. This incident aroused strong protest; student demonstrations at Kent State University in Ohio led on May 4 to a confrontation with troops of the Ohio National Guard, who fired on the students without orders, killing four and wounding several others.

Antiwar demonstrations

National revulsion at this act led to serious disorders at many universities and forced some of them to close for the remainder of the term. Many Americans were also disillusioned by revelations in 1971 about the conduct of the war, especially the treatment of Vietnamese civilians. Further antiwar demonstrations followed the 1971 U.S. invasion of Laos and Nixon's decision to resume intensive bombing of North Vietnam in 1972.

Despite continued U.S. involvement in the war, peace negotiations with North Vietnam slowly progressed, and the long-awaited cease-fire agreement was finally signed on Jan. 27, 1973. The agreement, which provided for exchange of prisoners of war and for U.S. withdrawal from South Vietnam without any similar commitment from the North Vietnamese, ended 12 years of U.S. military effort that had divided the American people profoundly, brought down a president, and taken some 58,000 American lives.

Domestic affairs. When Chief Justice Earl Warren, who had presided over the most liberal Supreme Court in history, retired in 1969, Nixon replaced him with the conservative Warren Burger. Three other retirements enabled Nixon to appoint a total of four moderate or conservative justices. The Burger court, though it was expected to, did not reverse the policies laid down by its predecessor.

Nixon's most original legislative proposal, a family assistance plan that would have replaced welfare with a minimum income, was rejected by Congress for partisan reasons. Congress did, however, enact the president's revenue-sharing program, which provided direct grants to state and local governments, but the program's effectiveness was diminished by substantial reductions in other forms of aid. Congress also expanded social security and federally subsidized housing. In 1972 the Congress, with the support of the president, adopted a proposed constitutional amendment guaranteeing equal rights for women.

The Equal Rights Amendment

Despite widespread support, the Equal Rights Amendment, or ERA, as it was called, failed to secure ratification in a sufficient number of states. (Subsequent legislation and court decisions, however, gave women in substance what the ERA had been designed to secure.)

Nixon gave high priority to the control of inflation and the reduction of military expenditures. But the cost of living continued to rise, until by June 1970 it was 30 percent above the level of 1960; industrial production declined, as did the stock market. By mid-1971 unemployment had reached a 10-year peak of 6 percent, and inflation continued. Wage and price controls were instituted, the dollar was devalued twice, and the limitation on the national debt was raised three times in 1972 alone. The U.S. trade deficit improved, but inflation continued unchecked.

The Watergate scandal. The scandal that was to haunt Nixon's second term, and lead to its premature end, first surfaced in June 1972, when five men were arrested for breaking into the Democratic national headquarters at the Watergate office–apartment building in Washington. When it was learned that the burglars had been hired by the Committee to Re-Elect the President (CRP), John Mitchell, a former U.S. attorney general, resigned as director of CRP, even though responsibility for the crime appeared to be limited to underlings. These events, however, had no effect on the election that fall. Even though the Democrats retained majorities in both the Senate and the House, Nixon won a landslide victory over Democratic nominee Senator George McGovern of South Dakota, capturing every state except Massachusetts and the District of Columbia.

In 1973, however, it was revealed that an attempt to suppress knowledge of the connection between the Watergate affair and CRP involved highly placed members of the White House staff. In response, a Senate select committee was formed and opened hearings in May, and Nixon appointed Archibald Cox as a special prosecutor to investigate the scandal. Amid conflicting testimony, almost daily disclosures of further scandals, and continuing resignations of administrative personnel, a battle developed between the legislative and executive branches of government. Nixon attempted to stop the investigation by firing Cox, leading Attorney General Elliot Richardson and Deputy Attorney General William D. Ruckelshaus to resign. This "Saturday night massacre" of Justice Department officials did not, however, stem the flow of damaging revelations, confessions, and indictments.

The Watergate affair itself was further complicated by the revelation of other irregularities. It became known that a security unit in the White House had engaged in illegal activities under the cloak of national security. Nixon's personal finances were questioned, and Vice President Spiro T. Agnew resigned after pleading no contest to charges of income tax evasion. On Dec. 6, 1973, Nixon's nominee, Congressman Gerald R. Ford of Michigan, was approved by Congress as the new vice president.

On May 9, 1974, the Judiciary Committee of the House of Representatives began hearing evidence relating to a possible impeachment proceeding. On July 27–30 it voted to recommend that Nixon be impeached on three charges. On August 5 Nixon obeyed a Supreme Court order to release transcripts of three tape-recorded conversations, and he admitted that, as evidenced in the recordings, he had taken steps to direct the Federal Bureau of Investigation away from the White House when its inquiries into the Watergate burglary were leading it toward his staff.

The admission created a storm. Nixon's support in Congress vanished, and it seemed probable that he would be impeached. On the evening of August 8, in a television address, Nixon announced his resignation, effective the next day. At noon on August 9, Vice President Ford was sworn in as his successor, the first president not elected either to the office or to the vice presidency.

Nixon's resignation

The Gerald R. Ford administration. Ford's was essentially a caretaker government. He had no mandate and no broad political base, his party was tainted by Watergate, and he angered many when he granted Nixon an unconditional pardon on Sept. 8, 1974. Henry Kissinger remained secretary of state and conducted foreign policy along the lines previously laid down by Nixon and himself. Ford's principal concern was the economy, which had begun to show signs of weakness. A brief Arab oil embargo during the Yom Kippur War had led to a quadrupling of oil prices, and the oil shock produced both galloping inflation and a recession. Prices rose more than 10 percent in 1974, and unemployment reached 9.2 percent in May 1974, a rate not seen since the Great Depression. Ford was no more able than Nixon to deal with the combination of inflation and recession, called "stagflation," and Congress had no remedies either. A limited exception was the Energy Policy and Conservation Act of 1975. It reduced the price of domestic oil, gave the president more control over energy, plus the right to impose gasoline rationing and conservation measures, and created a strategic petroleum reserve. None of this accomplished much, however, either then or later, because presidents were reluctant to employ such measures. Congress also set mandatory fuel economy standards for post-1977 automobiles in an attempt to reduce gasoline consumption. For the most part, however, Congress and the president were at odds, Ford vetoing no fewer than 50 bills during his short term in office.

In the election of 1976 Ford won the nomination of his party, fighting off a strong challenge by Ronald Reagan, the former governor of California. In a crowded field of contenders the little-known ex-governor of Georgia, Jimmy Carter, won the Democratic nomination by starting early and making a virtue of his inexperience. Ford, despite Watergate and stagflation, nearly won the election,

Carter receiving the smallest electoral margin since 1916.

The Jimmy Carter administration. *Foreign affairs.* More than any other president, Carter used diplomacy to promote human rights, especially with regard to the governments of South Korea, Iran, Argentina, South Africa, and Rhodesia (Zimbabwe). Efforts to continue the détente with Russia foundered as the Soviets supported revolutions in Africa, deployed medium-range nuclear weapons in Europe, and occupied Afghanistan. Relations with the People's Republic of China, on the other hand, improved, and full diplomatic recognition of the Communist government took effect on Jan. 1, 1979. In September 1977 the United States and Panama signed two treaties giving control of the Panama Canal to Panama in the year 2000 and providing for the neutrality of the waterway. The Senate confirmed both treaties by single-vote margins.

Carter's most noted achievement was to sponsor a great step toward peace in the Middle East. In September 1978 he met with Egyptian president Anwar el-Sādāt and Israeli prime minister Menachem Begin at a two-week negotiating session at Camp David, Md., and on September 17 Carter announced that two accords had been signed establishing the terms for a peace treaty between Egypt and Israel. Further torturous negotiations followed, including visits by Carter to Tel Aviv and Cairo, before the peace treaty was signed in Washington, D.C., on March 26, 1979.

The Iran hostage crisis

Carter's greatest defeat was administered by Iran. Following the overthrow of Mohammad Reza Shah Pahlavi, who had been supported by the United States, the Islāmic Republic of Iran was proclaimed in Iran on Feb. 1, 1979, under the leadership of the Ayatollah Ruhollah Khomeini. In November militants seized the U.S. embassy in Tehrān and held its occupants hostage. An attempt to rescue the hostages in April 1980 failed, and the hostages were not released until Carter left office in January 1981. Carter's inability to either resolve the hostage crisis or to manage American perceptions of it disabled him as a leader.

Domestic policy. Carter's effectiveness in domestic affairs was generally hampered by his failure to establish good relations with Congress, his frequent changes of course, the distractions caused by foreign problems, and his inability to inspire public confidence. His major domestic effort was directed against the energy crisis, though with indifferent results. Inflation continued to rise, and in the summer of 1979 Carter appointed Paul Volcker as chairman of the Federal Reserve Board. Volcker raised interest rates to unprecedented levels, which resulted in a severe recession but brought inflation under control.

Carter's failure to provide decisive leadership hurt him in the election of 1980. With some difficulty, he won renomination as the Democratic candidate. Ronald Reagan was the Republican nominee, while Republican John B. Anderson of Illinois headed a third ticket. Reagan easily defeated the discredited Carter, who would have lost the election even if he had received the 5,600,000 votes that went to Anderson. Furthermore, the Republicans gained control of the Senate for the first time since 1954.

The Ronald Reagan administration. Reagan took office pledged to reverse the trend toward big government and

Reagan's economic reforms

to rejuvenate the economy by implementing supply-side economics, based on the theory that cutting taxes would stimulate so much growth that tax revenues would actually rise. After surviving an attempted assassination in March 1981 Reagan began pushing his plan for economic "revolution" through Congress. In May Congress approved his program, which would reduce income taxes by 25 percent over a three-year period, cut federal spending on social programs, and greatly accelerate a military buildup that had begun under Carter. The recession that had resulted from Volcker's policy of ending inflation through high interest rates deepened in 1981, but by 1984 it was clearly waning, without a resurgence of inflation. The U.S. economy moved into one of the strongest postwar recoveries, although interest rates did not drop into single digits until 1985.

In foreign affairs Reagan often took bold action, but the results were usually disappointing. His efforts to unseat the leftist Sandinista regime in Nicaragua through aid to the Contra rebels was unpopular and unsuccessful. U.S.–Soviet relations, which were the chilliest they had been since the height of the Cold War, did not improve when Reagan publicly called the Soviet Union an "evil empire." Reagan's decision to send a battalion of U.S. marines to Lebanon in support of a cease-fire resulted in a terrorist attack in 1983, in which some 260 marines were killed. On Oct. 21, 1983, he launched an invasion of the Caribbean nation of Grenada, where Cuban influence was growing. U.S. forces prevailed, despite much bungling, at a cost of 42 American, 160 Grenadan, and 71 Cuban lives. Popular at home, the invasion was criticized almost everywhere else. Relations with China worsened at first but improved in 1984 with an exchange of state visits.

Despite Reagan's poor foreign record, he benefited in the election of 1984 from a high degree of personal popularity, from the defeat of inflation, which, aided by the development of an oil glut, fell to 3.8 percent in 1983, and from the beginnings of economic recovery. This combination proved fatal to the hopes of the Democratic nominee, former vice president Walter Mondale of Minnesota, and his running mate, Congresswoman Geraldine Ferraro of New York, the first female vice presidential candidate ever to be named by a major party. Reagan overwhelmed Mondale, winning 49 states, a new record.

Reagan's second term was more successful than his first in regard to foreign affairs. In 1987 he negotiated an intermediate-range nuclear forces (INF) treaty with the Soviet Union, eliminating two classes of weapon systems that each nation had deployed in Europe. This was the first arms limitation agreement ever to result in the actual destruction of existing weapons. Relations between the superpowers had improved radically by 1988, owing primarily to the new Soviet premier, Mikhail Gorbachev, whose reforms at home were matched by equally great changes in foreign policy. An exchange of unusually warm state visits in 1988 was followed by Soviet promises of substantial force reductions, especially in Europe. Though Gorbachev retained the initiative throughout, Reagan took advantage of his opportunities. When he left office, Soviet-American relations were at their highest point since 1943.

Reagan's domestic policies were unchanged. His popularity remained consistently high, dipping only briefly in 1987 after it was learned that his administration had secretly sold arms to Iran in exchange for American hostages and then had illegally used the profits to subsidize the so-called Contras, a rebel force seeking to overthrow the left-wing government of Nicaragua. In the short run his economic measures succeeded. Inflation remained low, as did unemployment, while economic growth continued. There were worrisome aspects, however. While spending for domestic programs fell, military spending continued to rise, and

Growth of the budget deficit

revenues did not increase as supply-siders had predicted. The result was a staggering growth in the budget deficit. The United States, which had been a creditor nation in 1980, was now the world's largest debtor nation, owing some half a trillion dollars to foreigners. Furthermore, although economic recovery had been strong, individual income in constant dollars was still lower than in the early 1970s, and family income remained constant only because many more married women had entered the labour force. Savings were at an all-time low, and productivity gains were averaging only about 1 percent a year. Reagan had solved the short-term problems of inflation and recession, but he did so with borrowed money and without touching the deeper sources of America's economic decline. Thus, while his successor took office at a time of peace and prosperity, he inherited difficult problems as well.

Reagan's hand-picked successor was Vice President George Bush of Texas. The Democrats nominated Michael Dukakis, the governor of Massachusetts. Both candidates were politically moderate. Bush won the election on the record of peace and prosperity he inherited from President Reagan, and by putting Dukakis on the defensive.

For later developments in the history of the United States, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

For coverage of related topics in the *Macropædia* and *Micropædia,* see the *Propædia,* sections 961, 964, 965, and 973, and the *Index.* (W.L.O'N.)

NEW ENGLAND

Connecticut

One of the original 13 states and one of the six New England states, Connecticut is located in the northeastern corner of the United States. In area it is the third smallest state in the nation, with 5,018 square miles (12,997 square kilometres), and it ranks among the most densely populated. It lies athwart the great urban-industrial complex along the Atlantic coast, with Massachusetts on the north, Rhode Island on the east, Long Island Sound (an arm of the Atlantic Ocean) on the south, and New York on the west. Hartford is the capital. Connecticut takes its name from an Algonquian word meaning "land on the long tidal river."

Connecticut, with its many beaches and harbours, its forest-clad hills, and its village greens that are often surrounded by houses dating from the 17th and 18th centuries, represents a special blend of modern urban life, rustic landscape, and historic sites. It is a highly industrial and service-oriented state, and its personal income per capita is among the highest in the nation. The strength of its economy lies in a skilled working force, much of it fabricating products that have been manufactured in Connecticut since the products were invented.

The population is heavily urban. The state has no single large city, however, and the intense crowding characteristic of many urban areas is not found in Connecticut. It continues its long tradition of being a prosperous state, with in-migration attracted by the good employment opportunities, excellent educational facilities, and pleasant living conditions for the majority of its people.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Essentially a rectangle in shape, 90 miles (145 kilometres) west to east and an average of 55 miles north to south, Connecticut covers the southern portion of the New England section of the Appalachian Mountain system. It contains three major regions: the Western Upland, the Central Lowland (Connecticut Valley), and the Eastern Upland. The northern part of the Western Upland, often called the Berkshire Hills, contains the highest elevation in the state, 2,380 feet (725 metres), on the southern slope of Mount Frissell in the northwest corner. It is drained by one major river, the Housatonic, and numerous tributaries.

Physiographic regions

The Central Lowland is different in character, being a downfaulted block of land approximately 20 miles wide at the Massachusetts border and narrowing as it progresses toward the sea, which it meets at New Haven. It is filled with sandstone and shale. Periodic volcanic activity pushed immense quantities of molten rock to the surface and produced the igneous deposits of the central valley. These layers of sandstones and traprock have been faulted, broken, and tipped so that there are numerous small ridges, some reaching as high as 1,000 feet above their valleys. Within the lowland, the Connecticut and other rivers have eroded the soft sandstones into broad valleys.

The Eastern Upland resembles the Western in being a hilly region drained by numerous rivers. Their valleys come together to form the Thames River, which reaches Long Island Sound at New London. Elevations in this area rarely reach above 1,300 feet. In both uplands the hilltops tend to be level and have been cleared for agriculture.

Climate. In Connecticut's moderate climate, the average January temperature is around 26° F (−3° C), and most of the state receives from 35 to 45 inches (889 to 1,143 millimetres) of snow each year. In the northwest, however, the average snowfall exceeds 75 inches. Snow may remain on the ground until March, but usually mild spells and rains melt it earlier in the year. Summers average between 70° and 75° F (21° to 24° C), with occasional heat waves driving the daytime temperatures above 90° F (32° C). Precipitation, averaging from three to four inches per month, is evenly distributed. The coastal portions have somewhat warmer winters and cooler summers than does the interior, while the northwestern uplands are high enough to have cooler and longer winters with heavier falls of snow. Occasionally, hurricanes have caused flooding and other damage, particularly along the coastline. Tornadoes and severe thunderstorms sometimes occur in the Connecticut River valley.

Perhaps the most marked characteristic of Connecticut's weather is its changeability. Cold waves and heat waves, storms and fine weather can alternate with each other weekly or even daily. The statement of Hartford resident Mark Twain, "If you don't like Connecticut weather, wait a minute," has become a widely appropriated and adapted proposition.

Plant and animal life. Originally, Connecticut was a forested region. The few Indian clearings, the swampy floodplains, and the tidal marshes accounted for about 5 percent of the total area. The southern two-thirds is largely an oak forest. The northern border belongs to the northern hardwood region of birch, beech, maple, and hemlock. A few higher elevations and some sandy sections support a coniferous forest. Virtually all of the primeval forest has been cut, and the current woodland that covers nearly two-thirds of the state is a mixed forest.

The animal life when the first European settlers arrived included deer, bear, wolves, foxes, and numerous smaller species, such as raccoon, muskrat, porcupines, weasels, and beaver. Deer are still found in the less densely settled regions, but in general the populations of larger animals have been severely reduced. Most birds are migratory, but chickadees, blue jays, and the immigrant English sparrows are year-round residents.

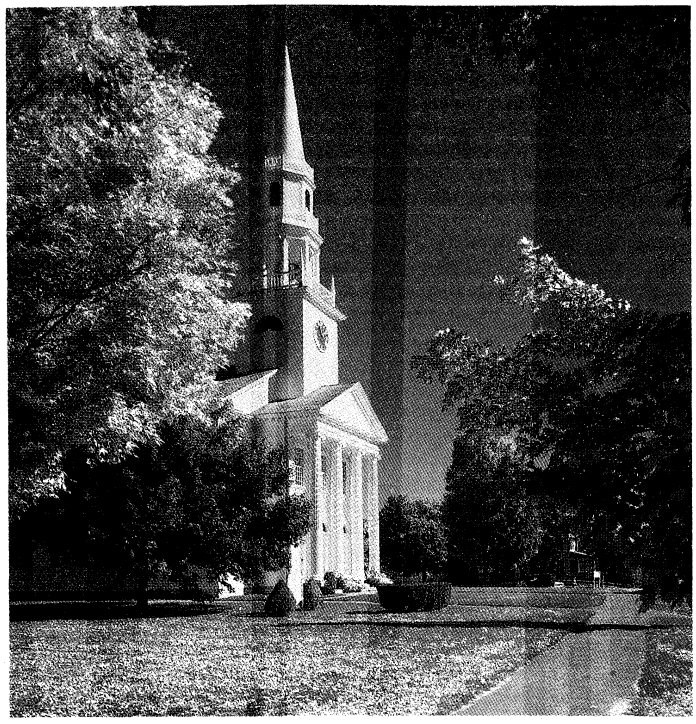
Settlement patterns. Most regions in Connecticut are not clearly defined, although Fairfield county in the southwest section is uniquely oriented toward New York City, serving as a suburb for many commuters. With two of the state's largest cities, Stamford and Bridgeport, Fairfield is one of the most populous counties in the state. The northwestern and northeastern quarters of Connecticut are less densely populated. They have some agriculture, but most residents there, as elsewhere in the state, work in the manufacturing cities and towns along the rivers.

Connecticut's small towns represent a territorial concept that is equivalent to a township in other parts of the country. Within each town, a town centre is surrounded by the town hall, schools, churches, usually a village green, a number of houses, and often a tiny business district with several stores. Elsewhere within the town, other hamlets may contain similar communal gatherings. If the hamlet is on a stream, the houses often cluster around a red brick factory that was erected in the 19th century to run its machinery from a waterwheel in the river. Such mill villages are to be found throughout the state, although many of the factories have been abandoned. Farmsteads and cultivated fields once lay between such small population nodes, but the roads connecting these villages have become sparsely lined with rural, nonfarm homes.

Rural and urban settlement

City status in Connecticut is determined not by population but by vote of the residents to change their governmental system from a town meeting to a city form. All of the larger towns and cities are manufacturing centres; some originated as mill towns and grew with their factories. The power source changed from water to steam and later to electricity, and often the products manufactured have changed to satisfy a new economic and social structure, but each city and town prides itself on the uniqueness that often is associated with its products.

The people. *Ethnic composition.* The Algonquian Indians, the original occupants of Connecticut, comprised about 16 separate tribes with some 5,000 to 7,000 members. The first European settlers were English, coming directly from England or by way of the Massachusetts Bay colony. During the 17th and 18th centuries population growth occurred primarily through an excess of births over deaths; immigrants, mainly from the British Isles, arrived at a rather slow rate. At the time of the first U.S. census, in



Church in Litchfield, Conn.
D. Forbert/Shostal Associates

1790, Connecticut had a homogeneous population, about 96 percent of which was of English ancestry; blacks accounted for only about 2 percent of the population.

Patterns of immigration

The immigration of the Irish, beginning in the 1840s, and of French Canadians after the Civil War, continued throughout the 19th century. Later in the 19th century the primary sources of foreign immigration shifted to southern and eastern Europe—Italy, Poland, the Austro-Hungarian Empire, and Russia. Each immigrant group tended to congregate in certain parts of the state. Thus New Haven and its suburbs are populated with large numbers of descendants of Italian immigrants; Poles are concentrated in the Naugatuck valley; and French Canadians live in the northeast. The immigration of blacks into Connecticut after World War II showed the same tendency. Most blacks live in the five largest cities. New Haven and Hartford are more than 30 percent black. Puerto Ricans have moved from their island and from New York City into Connecticut's urban centres. Hartford has a large concentration of Jamaicans. Other West Indies islands are also well represented in the major cities.

Demographic trends. From 1790 to 1840 the state's growth rate hovered between 4 and 8 percent per decade. Connecticut was—considering its small size and its limited agricultural resources—quite adequately filled. During the 19th century thousands of Connecticut residents, especially the young, migrated to better agricultural lands in the western part of the country; their places were taken by newcomers from Europe. The state's population growth passed the national rate in 1900 and, with one exception, did not fall below it until 1980.

For more than 300 years the distribution of Connecticut's people has reflected the region's changing economy and the resources of the land. Settlement began in the middle Connecticut Valley, where the soils were good, and on the coast, where maritime activities, trading, and fishing supplemented the living that the settlers were able to derive from the land. The upland areas were not fully occupied until the late 18th century, yet by 1790 the population was fairly evenly distributed across the state. Towns with better agricultural lands or with other resources—marine or mineral—had denser populations. During the 19th century the rise of water-powered manufacturing attracted young people from the agricultural upland towns to the growing mill towns, and virtually all of the upland towns lost population. Manufacturing towns grew rapidly.

The movement of people and industry into the cities dominated the population movements until 1950. Since then, Bridgeport, Hartford, and New Haven, the three largest cities, have had a general movement of population to the suburbs and to the former agricultural hill towns.

The economy. With limited natural resources, a well-educated and innovative citizenry has enabled Connecticut to reach high levels of productivity. Connecticut's creative genius has produced large shares of significant inventions and patents.

An economic overview

Connecticut, like other areas of the Northeast, lost much traditional industry to the Sunbelt in the 1940s and '50s, but the '70s and '80s were marked by economic rebound. In the 1980s Connecticut ranked first in personal income per capita and had one of the lowest unemployment rates in the nation. Labour unions have been strong and may be given partial credit for the high wages and good working conditions characteristic of most Connecticut factories. Business is also a powerful force in the state. The Connecticut Business and Industry Association is sophisticated and influential, and there are many active local chambers of commerce.

Agriculture and fishing. Since 1870, agriculture has declined in importance, and it is now a relatively minor element in the economy. The precipitous decline in the number of farms resulted in the enactment in 1978 of a farmland preservation program. Livestock and animal products are the major source of farm income. Connecticut's farms produce substantial quantities of milk, eggs, poultry, and vegetables for local consumption and one important export crop, shade-grown tobacco, used mainly for cigar wrappers.

Except for the oyster industry and the historically important whaling industry, commercial fishing has never been very important in the state. The oyster industry has attempted a comeback from the devastation caused by natural elements and pollution of the coastal waters.

Industry. The foundation of Connecticut's economy is manufacturing, which employs about one-fourth of the state's work force. High-technology companies, small development firms, and service industries have all shown growth since the 1970s. Defense contracts play a vital role in Connecticut's manufacturing. In addition to military products, thousands of items that are sold on a worldwide basis are produced in the state. Among the items that have been manufactured in Connecticut by long tradition are pins, clocks, silverware, sewing machines, Winchester rifles, and many brass products.

Historically, mining was important; but the last iron and copper mines closed long ago, and the state's high ranking in value added by manufacture is due mainly to the import of nearly all raw materials. Only sand, gravel, stone, feldspar, clay, and mica are still produced within the state. A large number of minerals were first discovered in Connecticut by Yale University mineralogists, and many quarries still yield interesting finds to collectors.

The insurance industry

Connecticut is often referred to as the nation's insurance centre and Hartford as The Insurance City. Marine insurance was the first concern of Connecticut companies, and eventually the coverages that they offered expanded to many forms of casualty insurance. Some of the largest insurance companies of the United States are based in Connecticut.

Transportation. Connecticut's railroad network is a basic link in the Boston–New York City transportation pattern. The first railroads were constructed to bring the produce of the agricultural interior to Connecticut ports. Each of the larger river valleys—the Housatonic, Naugatuck, Connecticut, Willimantic, and Quinebaug—supported its own railroad. The line along the coast was completed in 1852. Until 1930 the railroads flourished with the expanding Connecticut economy, but highway competition for passengers and products reduced railroad traffic severely. Most of the river lines have dropped passenger service; freight service continues on some, but on others it has been abandoned. Amtrak operates two passenger lines through Connecticut: the New York-to-Boston route and the route from New York to Springfield, Mass. Service on the New York–New Haven line has deteriorated despite

its heavy use as a commuter facility between southern Connecticut and New York City. Rehabilitation began in the early 1980s.

Interstate highways crisscross the state, but they are concentrated in the densely settled coastal and Connecticut Valley regions. Connecticut pioneered in the limited-access highway. The first section of the Merritt Parkway, from New York to near Milford, opened in 1938 and is often acclaimed as one of the most scenic and best designed. On June 28, 1983, several lives were lost with the collapse of the Mianus River bridge. The governor and legislature responded with the nation's first comprehensive infrastructure renewal program for roads and bridges.

Bradley International Airport, north of Hartford, is the major air terminal, but there are many other airports throughout the state that offer regional services. The port of New Haven is one of the largest in New England, and the U.S. Coast Guard Academy facilities are at New London.

Administration and social conditions. *Government.* Connecticut has enjoyed 350 years of constitutional government, from the Fundamental Orders of 1638 to the present constitution of 1965. Connecticut's state government is headed by a strong governor who is elected for a four-year term. The governor initiates legislation, prepares the state budget, appoints department heads, and can veto individual items of an appropriation bill. An Office of Policy and Management develops the governor's proposed budget and oversees executive agencies. Major oversight is also provided by an independent, bipartisan auditors' office.

Executive branch

Connecticut's legislative branch, the General Assembly, is composed of a 151-member House and a 36-member Senate. It met biennially until a constitutional amendment adopted in 1970 provided for annual legislative sessions. The 36 senatorial districts are approximately equal in population. The House of Representatives was originally based on towns, with each town, regardless of size, having at least one representative. The 1965 constitution reapportioned the lower branch so that it, like the Senate, is based on equivalent population. Connecticut's General Assembly is different from most state legislatures in that all business is conducted through joint House–Senate standing committees. Although the legislature is in session for only three to five months each year, there is interim committee activity. Major legislative issues include the budget, education, solid-waste disposal, drugs, local aid and property tax relief, crime, and housing.

The state's judiciary is headed by the Supreme Court. Superior courts were formed in 1978 by a merger of the courts of common pleas and the juvenile courts. The justices of the Supreme Court and of the superior courts are nominated by the governor and appointed by the General Assembly for eight-year terms. Probate judges are elected on partisan ballots for four-year terms.

Below the state government are 169 local units called towns. They are creations of the state, with their rights and responsibilities set out in state statutes. There is, nonetheless, a long-standing, intense tradition of local autonomy. These local governments maintain roads and provide elementary and secondary education and police and fire protection. Larger municipalities also provide water and sewage facilities and other services. Originally, government was based on the town meeting, at which the citizens elected selectmen to run the town between the annual meetings. As populations increased and problems of administration became more complex, other systems were substituted. Most larger communities have opted for a city form with an elected mayor and council. Some smaller communities have elected mayors; some have town or city managers. Many towns have retained the town meeting or have substituted the representative town meeting.

The state government provides funds to local governments for the many social programs that are operated. Although Connecticut has no broad-based income tax, it does tax interest, dividends, and capital gains above certain income levels. State government relies heavily on high sales and business taxes for revenue. Local governments are almost entirely reliant on property taxes, supplemented by state and federal aid.

Education. From the earliest days, every town has been required to maintain public elementary schools and, as the town grew in size, secondary schools as well. Connecticut is renowned for its many private schools and colleges. Yale University (1701), in New Haven, is regarded as one of the world's great universities; other institutions, such as Wesleyan University (1831) in Middletown, have national recognition. Public higher education has expanded considerably. The community college system, founded in 1965, has more than 10 colleges. Also under the aegis of the State Board of Governors for Higher Education are the University of Connecticut (1881) at Storrs, with several branches, four other state universities, and five technical colleges.

Major universities

Health and welfare. The community and the state have become increasingly involved in health and medical care. Most people live within 10 miles of hospital services, and doctors and other medical personnel are numerous. There are many community health clinics in addition to the advanced medical centres of the University of Connecticut at Farmington and of the Yale–New Haven Hospital. A Health Care Cost Commission has endeavoured to control rapidly spiraling cost increases. In relation to most states, Connecticut provides generous welfare benefits. Departments for the aged and for children and youth services have been established to meet the special needs of communities.

Despite inner-city blight and abandoned housing, progress has been made by urban redevelopment programs in Connecticut's larger cities. Urban renewal programs in New Haven during the 1950s and '60s became a prototype for the nation. Much work in rehabilitating urban areas remains to be done, however, especially in residential neighbourhoods. There is also a shortage of lower- and middle-income housing.

To correct abuses in the free enterprise system, Connecticut enacted numerous regulations. The first child-labour law was passed in 1842, but it was ineffectual: for 30 years more, hundreds of children continued to work long hours in the textile mills. A labour department was set up by the state in 1873, and since then hundreds of laws and regulations have been enacted to control working conditions. The length of the working day, minimum wage rates, equal pay for equal work, and similar protective regulations have been passed. State departments supervise banks, insurance companies, and the public utilities, and in 1959 the Department of Consumer Protection was organized to consolidate several existing agencies.

Labour legislation

Cultural life. Connecticut provides a variety of landscapes: rocky headlands, beaches, forested hills, and, perhaps most attractive, small towns around tree-dotted village greens. In the towns, hundreds of houses dating from the 17th and 18th centuries are preserved by more than 100 local or national historical societies.

Numerous sites important in Connecticut's past or associated with illustrious individuals are maintained by state or private organizations. These include the Putnam Wolf Den in Pomfret, Mount Riga Furnace in Salisbury, Fort Griswold State Park in Groton, Old New-Gate Prison and Copper Mine in East Granby, the Mark Twain Memorial home in Hartford, the Tapping Reeve House and Law School in Litchfield, and the (William) Gillette Castle State Park in East Haddam. Perhaps the best known is Mystic Seaport and Marine Museum in Mystic, where a small New England seaport has been recreated with all its ships and shops. The outdoorsman can tramp the many miles of trails and camp in one of the 30 state forests, covering more than 130,000 acres (52,500 hectares), or in one of 90 state parks, comprising some 30,000 acres (12,000 hectares). Colourful autumn foliage impresses many visitors to Connecticut and the rest of New England. Sport fishing, particularly for bluefish, is popular in Long Island Sound.

Art exhibitions are held annually in many cities, a number of which have art galleries and museums. The best known are the Yale University Art Gallery, the Wadsworth Atheneum in Hartford, and the New Britain Museum of American Art. Symphony concerts and concerts by smaller groups are presented regularly in the larger communi-

Theatre companies

ties. Several educational institutions have public concerts throughout the year. Repertory companies operating in or near resort areas in the summer include the Westport County Playhouse in Westport and the Oakdale Musical Theatre in Wallingford. The American Shakespeare Theatre in Stratford, the Long Wharf Theatre in New Haven, and the Goodspeed Opera House in East Haddam are well known. The Yale School of Drama was, at its founding in 1925, the first such school at an institution of higher learning. Southwestern Connecticut is also within easy reach of the vast artistic resources of New York City.

HISTORY

Colonization. In contrast to many of the other New England areas, relations between Indians and the early settlers in Connecticut were good. Trading posts were established along the Connecticut River by the Dutch from New Amsterdam and by the English from the Plymouth colony, but the first permanent European settlers in the state came from the Massachusetts Bay colony to the middle Connecticut Valley during 1633–35 and to the Saybrook–New Haven coastal strip during 1635–38. In 1665 the Connecticut River settlements and the New Haven colony were united, and the general outline of the state emerged, although its borders were not finally demarcated until 1881, more than 200 years later. The New Haven colony was unsuccessful in an attempt to settle Delaware Bay, and the united Connecticut colony, despite its charter provisions, lost its claim to a strip of land extending to the Pacific. Following the American Revolution, settlers from Connecticut, with claims in the Midwest, were among the first to move into an area that became known as the Western Reserve, now northeastern Ohio.

Political, economic, and social maturation. The political development of the colony began with the Fundamental Orders of Connecticut (1638), a civil covenant by the settlers establishing the system by which the river towns of Windsor, Hartford, and Wethersfield agreed to govern themselves. The orders created an annual assembly of legislators and provided for the election of a governor. This was superseded by the royal charter of 1662, a liberal document that provided for virtual self-government by the propertied men of orthodox faith in the colony. It served Connecticut well until it was replaced by the state constitution adopted in 1818, a document that after being amended many times was replaced by a new constitution in 1965, reflecting the more complex needs of contemporary government. The Congregational church was disestablished by the constitution of 1818.

Connecticut remained an agricultural region of farms with a few small urban areas—Hartford, New Haven, New London, and Middletown—until the early 19th century. The economy began to change, however, after 1800, when textile factories were established, and, by 1850, employees in manufacturing outnumbered those in agriculture. The shift to manufacturing had been aided by the inventive genius of a number of Connecticut residents. Eli Whitney, well known for his invention of the cotton gin in 1794, developed the idea of machine-made parts for guns. An order for muskets from the federal government enabled him to build a musket factory in Hamden. The principle of interchangeable parts, adapted to clock manufacturing by Eli Terry of Plymouth in 1802, rapidly became basic to all manufacturing.

Industrial development

The economic, social, and political innovations that emerged in the 19th and 20th centuries were often resisted at first, but eventually they were accepted. Slavery, first attacked by legislation in 1784, was not abolished completely until 1848. The constitution of 1818 granted suffrage to men with certain property qualifications, but women's suffrage came only through federal enactment in 1920.

Connecticut's concentration of military and small-arms manufacture contributed much to the nation's efforts in World Wars I and II and the Korean and Vietnam wars.

While Connecticut has a healthy two-party system, Democrats have dominated state politics since the Great Depression of the 1930s. Governors Wilbur Cross, Chester Bowles, Abraham Ribicoff, and John Dempsey all presided over periods of unprecedented economic development, school construction, civil rights activity, and increased health and social services. In 1974 Ella Grasso became the first woman in any state elected in her own right to the office of governor.

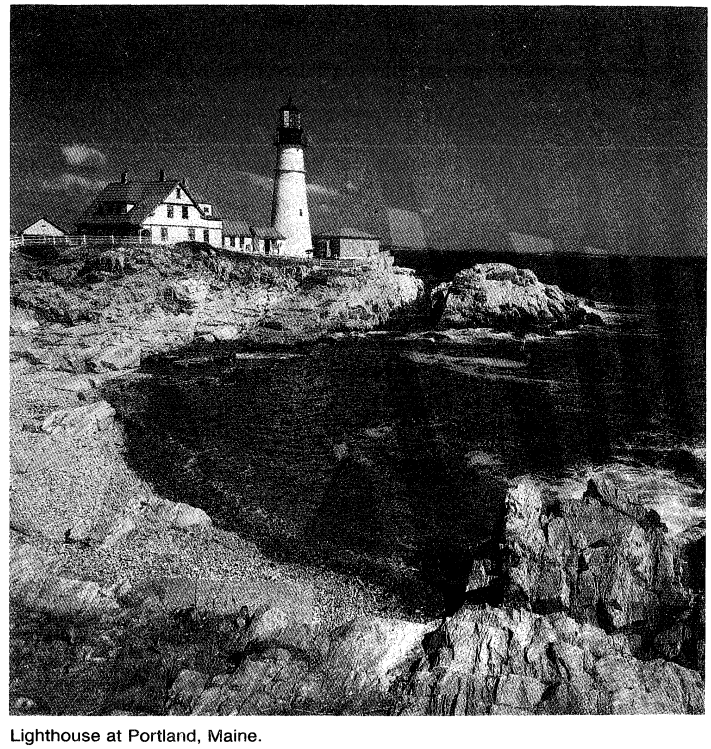
Connecticut displays sharp contrasts between areas of great wealth and deep poverty. The central cities of Hartford, New Haven, and Bridgeport are particularly affected. In this sense there are two Connecticuts. Closing the gap remains the state's major challenge. (J.B.Ho./I.J.S.)

Maine

Maine, the largest of the six New England states in area, lies at the northeastern tip of the United States. Its 33,265 square miles (86,156 square kilometres), including 2,270 square miles of inland water area, represent nearly one-half of the total area of New England. Maine is bounded on the northwest and northeast by the Canadian provinces of Quebec and New Brunswick, respectively, and on the west by New Hampshire. The famed rocky coastline of the state is angled from southwest to northeast along the Atlantic Ocean. Maine's capital has been Augusta since its admission on March 15, 1820, as the 23rd state of the Union. There are two theories of the derivation of the state's name: that the state was named for the former French province of Maine and that it was so named for being the "mainland," as opposed to the coastal islands.

Maine is the most sparsely populated state east of the Mississippi River. Nearly 90 percent of its total land area is under forest cover. It is also, by most statistical measures, an economically depressed state, but the rugged beauty and challenge of its climate and landscape and the character of its people have given Maine an importance beyond its economic and political power. Limited economic growth, in fact, has contributed to the preservation of much of its natural appearance. Since 1970 the state's southwestern coastal counties have recorded accelerating growth rates, increased residential and commercial construction, and increased tax revenues. Maine's economy remains dependent on manufacturing related to the timber harvest (mostly paper and paper products). The extractive industries of fishing, mining, and agriculture have been surpassed by trade and the recreational and service industries, these latter accounting for an increasing percentage of the state's income. The state epitomizes the increasingly

Vernon Sigl/Shostal Associates



Lighthouse at Portland, Maine.

difficult national choices between preservation of environmental quality and potential economic expansion.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The Appalachian Mountain chain extends into Maine from New Hampshire, terminating in Mount Katahdin, at 5,268 feet (1,606 metres) the state's tallest peak. The western and northwestern borders adjoining New Hampshire and Quebec have the most rugged terrain, with numerous glacier-scoured peaks, lakes, and narrow valleys. South and east of the mountain areas lie rolling hills and smaller mountains and the broad valleys of the Saco, Androscoggin, Kennebec, and Penobscot rivers.

Mountainous areas

From Kittery, at the southern tip of the state, to Cape Elizabeth, just southwest of the state's largest city, Portland, long sand beaches are interrupted intermittently by rocky promontories. North and east of Cape Elizabeth the coastline of Maine is a series of peninsulas, narrow estuaries, bays, fjords, and coves, once glacier-covered mountains and valleys now partially submerged in the post-Ice Age rise in sea level. The Camden Hills and Mount Desert Island are the largest of the coastal mountains. The tides along this famous rockbound coast are among the strongest in the world, running between 12 and 24 feet (3.7 and 7.3 metres). Off the coast of the state lie about 1,200 islands, some no more than rocky ledges, others topped with trees and sheltering the homes of fishermen, lobstermen, and summer residents. All told, the coast of Maine—including the bays, islands, and inlets washed by the tides—totals some 3,500 miles (5,630 kilometres).

Drainage and soils. Most of Maine's river systems flow from north to south. The St. John River and its principal tributary, the Allagash, are the major exceptions, flowing north and then east along the northern border of Maine and turning south through New Brunswick, Can., to the sea. The state is dotted with 2,500 lakes and ponds, the largest of which is Moosehead Lake (120 square miles). Soils in southwestern Maine were formed primarily from granite; coastal, central, and eastern soils are composed of shale, sand, and limestone; while the soils of Aroostook county, in the northeast, which are among the most productive in the state, are largely composed of caribou loam.

Lakes and ponds

Climate. Maine has three relatively well-defined climatic areas: southern interior, coastal, and northern. The southern and coastal regions are influenced by air masses from the south and west. North of the land dividing the St. John and Penobscot river basins, air masses moving down the St. Lawrence River basin tend to prevail. Mean annual temperatures range from 37° to 39° F (3° to 4° C) in the north and from 43° to 45° F (6° to 7° C) in the southern interior and coastal regions. Mean temperatures are about 62° F (17° C) throughout the state during the summer and 20° F (−7° C) during the winter. Clear days range from about 100 per year in the south to only 70 in the north, and annual precipitation averages 36 to 48 inches (910 to 1,220 millimetres). Snowfall averages more than 100 inches in the north and at higher elevations.

Plant and animal life. Flora and fauna represent a combination of subarctic and Appalachian species. Forests include heavy stands of pine, spruce, and fir among the softwoods. Sugar maple, yellow birch, aspen, and paper birch dominate the extensive stands of hardwoods. Among the fauna are deer, moose, black bear, fox, lynx, hare, raccoon, porcupine, skunk, and woodchuck. Songbirds, lake birds, seabirds, and many game species abound throughout the state. Among the many aquatic species are the seal, whale, porpoise, lobster, shrimp, clam, haddock, cod, mackerel, and Atlantic and landlocked salmon, as well as many freshwater game fishes.

Settlement patterns. Maine's general population distribution reflects the early patterns of settlement along the coast and the river valleys, with vast sections of the interior covered with forest and virtually uninhabited except for occasional lumber encampments. About one-half of the population is concentrated in four southwestern counties: Androscoggin, Cumberland, Kennebec, and York. Almost one-half of Maine's residents live in what are classified as urban areas, but there are only a few cities of 25,000 or more inhabitants.

Coastal Maine is best known, from anecdotes and dialect stories, as the traditional home of the "Down East Yankee." Many communities in the region, relatively isolated from the principal avenues of highway traffic, were once bustling centres of ocean commerce and river trade. Population movements in the state have blurred some of the regional differences, but within the coastal region there are three distinct areas. The southwestern coast, predominantly a resort area, extends from Kittery to the Portland metropolitan region on Casco Bay. The midcoast region, marked by a combination of fishing and maritime activities, vacation and retirement homes, and resort centres, runs from Bath (long a shipbuilding centre) and the mouth of the Kennebec River to Belfast, on the western shore of Penobscot Bay. The eastern coastal region of Maine begins on the eastern shore of Penobscot Bay and ends at Calais, on the St. Croix River, at the New Brunswick border.

The coastal region

Central and southern Maine form a contiguous region covering the southern half of the state, from the New Hampshire border to the Penobscot River. It contains the bulk of the population and most of the industrial and commercial activities. The western border areas of York, Cumberland, Oxford, and Franklin counties form a subregion within the state, but increased mobility is tending to erode the distinctive community and speech patterns of those sections.

Aroostook county, a region by itself, is often referred to as "the" county. Central and southern Aroostook areas were settled by English and Irish immigrants whose speech patterns continue to resemble those of their neighbours across the border in New Brunswick more than they do the broad *a*'s and dropped *r*'s of the rest of Maine speech. The St. John valley along the northern border of Aroostook county was settled by Acadians of French descent from Nova Scotia and New Brunswick. The communities of the valley retain their French character and speech.

Maine's rocky terrain limits the size of farms in most areas of the state. With the exception of Aroostook county and a few broad valleys in the central region, the fields are small, and in many cases they are marked by old stone walls or separated by wooded lots.

Interior rural communities in Maine vary according to the terrain and their economic history. Some consist only of a crossroads settlement with a store, gas station, post office, and three or four homes; others have a church, school, a few stores, and small establishments clustered around a millsite; still others have the traditional village green, often with the typical white frame, single-spired New England church, as well as such social centres as a grange hall. Communities in the state that prospered during the height of the lumber trade are marked, where the terrain permits, by broad avenues and imposing wooden homes. Coastal communities are similar, with commercial areas on the waterfront and social, cultural, and residential centres on higher ground.

Rural life

Maine's largest urban communities are Portland, Lewiston–Auburn, Bangor, Augusta, Biddeford, and Waterville. Portland is the centre of a metropolitan area spreading inland and around the harbour city, which lies on Casco Bay. It is the commercial and transportation hub of the state, whose economy has a growing and diversified industrial base, including paper manufacturing, steel fabrication, light manufacturing, and assembly. Bangor, an old lumbering town at the head of navigation on the Penobscot River, is the commercial centre for eastern and northern Maine. Augusta, the state capital, lies at the head of navigation on the Kennebec River. State government is the principal source of employment for the city, but it is also the site of textile, shoe, and paper industries.

The twin cities of Lewiston and Auburn form the second largest urban centre in the state. Long dependent on textile and shoe manufacturing, the two communities have pursued aggressive industrial development programs and have diversified into electronics and light manufacturing. They also serve as a commercial and trade centre for the Androscoggin valley and eastern Oxford county. Biddeford, south of Portland at the base of Saco Falls, is a lumbering and textile centre. Waterville, north of Augusta on the Kennebec, with its neighbouring communities of

Winslow and Fairfield, is a pulp-and-paper and textile centre and a commercial and trade centre for the central and northern Kennebec valley.

The people. The original Down East Yankees were English and Scotch-Irish Protestant immigrants who made the most substantial and persistent early European settlements in Maine. They set the style of dour and taciturn industry and dry wit that is characteristic of Maine legends and stories. Their descendants dominated the political and economic life of the state during most of its development, and they constitute its largest population group, particularly in the smaller communities and rural areas.

Contrary to popular impressions, however, Yankees are not the sole inhabitants of Maine. Two groups of French descent make up the second largest ethnic bloc in the state. The Acadians, originally from Brittany and Normandy, were driven out of Nova Scotia in 1763 by the British; many of them settled in the St. John valley, which now forms the northern border of Maine, while others made the long trip to Louisiana. The later French-Canadian migration from Quebec province began with the growth of the lumber and textile industries following the American Civil War. French is the primary language in much of the St. John valley, and it is the second language in Maine's industrial cities. Irish immigration to the state began in the 18th century, and the Irish and the French make up the bulk of Maine's Roman Catholic population. French Huguenot and German settlements were established early near the coast. During the 1870s the state encouraged the building of a Swedish settlement in Aroostook county as part of a program for agricultural development and population growth.

French language in Maine

Most of the remaining few thousand members of the original American Indian population of the area live on state reservations. The nonwhite population of Maine is less than 2 percent.

The economy. Maine's forest and waterpower resources invited exploitation during the early years of the Industrial Revolution; for a long time, skilled, low-cost labour provided an advantage to the textile and shoe industries until those industries moved their operations to factories in low-wage areas of the South and overseas.

Aroostook county, where the potato is the main crop, is one of the few areas with rich agricultural soils. Terrain and soil conditions throughout most of the state are inadequate for large-scale farming. With the exception of lobster production, fishing is a marginal industry. As a result of these factors, Maine is a relatively poor state, with the lowest income per capita in New England.

Since 1955 the state government has promoted an active economic development program through the Department of Economic and Community Development. Public and private agencies have combined to form the Finance Authority of Maine and the Maine Municipal Bond Bank to encourage investment and provide loan guarantees. The state also has used the services and financial assistance of the federal Economic Development Administration and of the Small Business Administration.

Economic development programs

Economic resources and components. Maine's primary natural resources are timber, sand, gravel, limestone and building stone, fish, and shellfish. There are limited deposits of copper, zinc, feldspar, and semiprecious stones. Peat is mined for horticultural use. Soils and climate have contributed to the production of high-quality potatoes and of apples, blueberries, and other fruits. Dairying is also an important activity.

The services sector represents the largest component in the market value of Maine goods and services; the manufacturing sector is second, and the trade sector is third. Pulp and paper constitute the largest item in manufactured products, potatoes and poultry in farm income, and lobsters in the fishery industry. Tourists—attracted by Maine's picturesque lakes, streams, and coast and the opportunities for swimming, boating, fishing, hunting, hiking, and winter sports—account for a large portion of retail sales and service income.

Electrical power. More than three-fourths of the electrical energy generated within the state is produced in steam plants; a large percentage of the steam is supplied by nuclear power, hydroelectric stations and diesel and gas-turbine units providing the remainder. Many of the state's hydroelectric sites have been restored, and biomass generators have been constructed in eastern Maine. Agreements to purchase surplus power from Canadian hydroelectric generating stations have been negotiated. Since 1972 a consortium of New England utility companies has operated the Maine Yankee atomic power plant at Wiscasset.

Transportation. Maine depends heavily on its roads for ground transportation. Railroads carry freight but no longer carry passengers. Buses provide interstate, intrastate, and some suburban and urban passenger transportation. Portland and Searsport are the major seaports. State and private passenger and freight ferry services operate to many of the coastal islands, and Portland and Bar Harbor have ferry connections with Yarmouth, N.S., Can. Several airlines operate from Presque Isle, Bangor, and Portland to points outside Maine; commuter airlines provide intrastate and interstate service to other Maine communities. Some international nonscheduled air passenger and freight traffic is routed through Bangor International Airport.

Administration and social conditions. *Government.* The constitution of the state was adopted in 1819. It is based on the constitution of the Commonwealth of Massachusetts and reflects colonial traditions of checks and balances. The governor, Maine's chief executive officer, is checked by the members of the Maine House of Representatives and Senate and their joint standing committees. This legislature elects several constitutional officers, including the attorney general, the secretary of state, the auditor, and the state treasurer. Department heads, appointed by the governor, are approved or rejected by the Senate. In 1957 the state legislature approved a constitutional amendment extending the governor's term to four years, with a two-term limit.

Executive officers

The legislature, which has a 151-member House and a 35-member Senate, is elected every two years. The president of the Senate is the constitutional successor to the governor. Maine has a three-tiered judicial branch, including district judges, a superior court, and a supreme court. Probate courts serve at the county level.

Maine's 16 counties traditionally have provided an administrative framework for the superior court system, law enforcement, land records, and probate practice and for some road maintenance and construction functions. Town government, with the annual town meeting and a board of selectmen, prevails in most communities. More than 20 communities operate under city charters. Professional managers are used in most cities and in many towns.

Local communities depend for revenues on property taxes, on automobile excise taxes, on fees for hunting, driving, and other licenses, on state aid for education, roads, and welfare, and on federal grants-in-aid. State revenues are obtained from a corporate and personal income tax, inheritance tax, sales and use taxes, motor fuel taxes, tobacco and alcoholic beverage taxes, licenses and miscellaneous taxes, federal grants-in-aid, and a state lottery.

With the election of 1954, traditional Republican dominance in Maine's state offices and national representation ended. Thereafter, Democrats have competed successfully with Republicans for the governorship and for federal and state legislative seats. Party officials are elected in local caucuses and state conventions. Nominations for county and state offices are obtained through primary elections, but Maine has no presidential-preference primary.

Education. Local governments are responsible for public elementary and secondary education, under the general supervision of a state board of education. Most rural areas are served by multicommunity school administrative districts. The state operates technical institutes for post-secondary vocational training. The University of Maine, established in 1865 in Orono as a college of agriculture and mechanic arts, has been reorganized into a seven-branch system and offers a broad range of undergraduate and graduate curricula. Private liberal arts colleges include Bowdoin College (Brunswick; 1794), Bates College (Lewiston; 1864), and Colby College (Waterville; 1813).

Welfare. Maine's chronic economic problems are reflected in the high incidence of poverty. The largest pro-

portion of poverty is found in the rural counties of the state, particularly in the eastern coastal counties and in Aroostook county. Public awareness of Maine's poverty and of the particular difficulties faced by the state's small Indian and black populations has led to vigorous efforts by community action groups, civil rights organizations, and health and housing associations to improve economic opportunity, as well as to deal with the related problems of housing, education, and health. Most poverty assistance is administered by the state Department of Health and Welfare.

Cultural life. In its culture, as in its social and economic development, Maine reveals the attributes of both a struggling frontier community and an eclectic society immersed in commerce with other cultures. Folktales, songs, local humour, and the short stories and poems of native authors are direct, earthy, and filled with a sense of the awkward absurdities of man's attempt to subdue nature.

Cultural heritage

The tools of the woodsman, farmer, and fisherman are clean and simple, as are the lines of country homes, meetinghouses, and working boats. The great mansions of the old seaports, among some of the finest memorials to an earlier America, are filled with chairs, tables, chests, books, prints, hangings, screens, pottery, and bric-a-brac gathered on the many voyages of Maine's seamen to Europe and Asia, as well as with examples of the shipbuilders' and sailors' arts of wood carving and scrimshaw. Maine has, in addition, the unique contributions of such groups as the Shakers and its own local versions of the Federal, Greek Revival, Gothic, and Victorian periods of American architecture.

Maine has had a revival in crafts production, including pottery, metalworking, block and silk-screen printing, weaving, furniture making, and carving. State agencies, historical societies, museums, and local associations are engaged in preserving historic sites and in the collection, preservation, and presentation of materials on Maine's heritage. The Marine Maritime Museum is located in Bath.

Maine has been, and continues to be, the birthplace or the permanent or seasonal home of well-known figures in the American arts. They have included such writers as Henry Wadsworth Longfellow, Harriet Beecher Stowe, Sarah Orne Jewett, Edward Arlington Robinson, and Edna St. Vincent Millay; the painters Winslow Homer, John Marin, Edward Hopper, and Andrew Wyeth; and composer Walter Piston. Among the state's largest and finest museums are the Portland Museum of Art; the Colby (College) Museum of Art, which has a large collection of works by Winslow Homer; and the Museum of Art on the Bowdoin College campus. Active cultural programs are sponsored by the state's colleges and universities, museums, community symphonies, workshops and camps, and the numerous summer theatres.

Acadia National Park, which extends over most of Mount Desert Island and Isle au Haut, was the first national park east of the Mississippi River. Other recreational attractions include Baxter State Park, a wilderness area of 200,000 acres (80,940 hectares) surrounding Mount Katahdin; the 92-mile Allagash Wilderness Waterway; and more than 100 state parks and historic sites.

HISTORY

Algonquian Indians were the earliest known settlers in Maine. They lived along the river valleys and the coasts, hunting, fishing, and planting crops. Few of them survived the arrival of the European settlers. But the earlier tribes are remembered in numerous place-names; in the sites of their camps and burial grounds; in ancient trails and water routes; in the use of the canoe, the snowshoe, and the toboggan; in crops such as corn (maize), beans, and squash; and in the revived concern for the natural environment.

The Indians

Explorations and disputes. The first European explorations of Maine are shrouded in mystery. Evidence that Norsemen landed on the coast is scant and disputed, and serious questions exist about some of the early British claims based on John Cabot's voyages in the late 1490s. Portuguese, Spanish, French, and English explorers did probe the islands, the bays, and the rivers of the "maine" throughout the 16th century; by the first decade of the 17th century, summer fisheries had been established on some of the coastal islands and fur trade had begun with the Indians.

An area claimed by both the French and English crowns was an intermittent battleground between the English, the Indians, and the French from 1615 until 1675 and a constant battleground from that date until 1763, when the British conquered the French in eastern Canada.

Maine was given separate provincial status in New England under royal patents granted by Charles I. The Puritans of Massachusetts took over the territory when the proprietor, Sir Ferdinando Gorges, backed the losing side in the English Civil Wars. Frontier settlers in Maine chafed under Massachusetts rule, but the merchants of the coastal towns resisted the separation movement until the War of 1812, when popular resentment against the failure of the Massachusetts Commonwealth to protect the District of Maine against British raids tipped the scales in favour of separation. Maine entered the Union as a free state under the Missouri Compromise of 1820.

The northeast boundary of the state was a matter of serious controversy between the United States and Britain. The Treaty of Paris (1783) identified the boundary in part as extending along the middle of the St. Croix River to its source and from there north to highlands running northwest to the "head of Connecticut river." Identifying these highlands proved to be difficult. Efforts at arbitration failed in 1831, and the disputed area was the scene of the so-called Aroostook War of 1838–39. In March 1839 General Winfield Scott arranged a truce calling for joint occupancy of the disputed territory. This remained in effect until 1842, when a settlement was reached that divided the territory virtually in half.

Statehood. Maine intrigued entrepreneurs, who hoped to make their fortune in furs, fisheries, timber, and land development. The first three proved to be lucrative for a few, but the climate, border troubles, and the availability of more fertile land in the newer territories to the west curtailed settlement of the area before and after statehood. The period of greatest economic growth came between 1830 and 1860, when lumber, ice, granite, lime (extracted from limestone), and fishing and shipbuilding dominated the state's economy. Coastal communities flourished and railroads developed as Maine merchants traded around the world.

Economic growth

The American Civil War and the Industrial Revolution diverted workers and capital from Maine during the last decades of the 19th century. Textiles and paper products became the primary sources of manufacturing employment, while fisheries and agriculture continued as important but uncertain sources of income. The details of economic activity changed during the first half of the 20th century, but the overall picture remained one of precarious prosperity and extreme susceptibility to swings in the national economy.

Political development. Maine's social and political history has been dominated by struggles against the adversity of frontier life and economic limitations, coupled with strong drives within the state for social reform, including world peace, abolition of slavery, prohibition, and women's suffrage. Jeffersonian and Jacksonian Democrats held sway from statehood until the rise of the Whigs and the emergence of the Republican Party. The Abolitionist movement gave the Republican Party its start in Maine in 1854, and the Grand Old Party dominated the state for almost a century. Democrats scored temporary gains in the elections of 1910 and 1912 and, during the Great Depression, in the elections of 1932 and 1934, but it was not until 1954 that sustained competition began to develop between the major parties.

Since the 1950s Maine has had Republican, Democratic, and Independent governors, each presiding over an economy that has recorded slow but steady gains. These gains have been the result of an increase in the service economy of southwestern coastal Maine, an economy energized by the in-migration of new residents, many from Northeastern corridor states. The resulting pressures on native coastal residents have produced local and state regulations designed to plan for and control economic growth rates

that have exceeded 10 percent per year in Maine's southern counties. (E.S.M./J.N.C.)

Massachusetts

Like others of the 13 British colonies along America's Atlantic seaboard in the 17th and 18th centuries, Massachusetts was founded by people seeking in a wilderness for a new way of life involving such then-untried notions as freedom of religion and self-government. These and other ideals were severely tested during more than 150 years of colonial life, but they came to provide much of the ideological underpinning of the American Revolution, from which Massachusetts emerged as one of the founding and leading members of the new United States.

One of the six New England states lying in the northeastern corner of the nation, the Commonwealth of Massachusetts, as it is known officially, is bounded on the north by Vermont and New Hampshire, on the east and southeast by the Atlantic Ocean, on the south by Rhode Island and Connecticut, and on the west by New York. It covers 8,284 square miles (21,456 square kilometres) and ranks 45th among the states in area. The residents represent an amalgamation of the prototypical Yankee spirit of an earlier America and the energies of the later immigrants who flocked to its cities in the 19th century.

Massachusetts has been, nearly from its founding, a leading force in American education. During the 19th century Boston, its capital, became synonymous with the highest attainments in America's cultural and artistic life, and the state as a whole provided industrial and financial leadership for the nation. Though these latter positions have long since been yielded to larger and faster-growing states and regions, the history and people of Massachusetts have left an indelible mark on the development of the American consciousness.

PHYSICAL AND HUMAN GEOGRAPHY

The land. The Massachusetts coastline is about 1,500 miles (2,400 kilometres) in length, yet the cross-country distances are only 190 miles from east to west and 110 miles from north to south. The jagged coast winds from Rhode Island around Cape Cod, in and out of scenic harbours along the shore south of Boston, through Boston harbour and up the North Shore, swinging around the painters' paradise of Cape Ann, to New Hampshire.

Relief. The indented coast of Massachusetts was formed by the great glaciers that in places covered the land with several thousand feet of ice. When the last ice disappeared some 11,000 years ago, massive chunks of rocks were exposed along the shore. Hard, flat land stretches out beyond, becoming stony upland pastures near the central part of the state and a gently hilly country in the west. Except toward the west, the land is rocky, often sandy, and not fertile.

Cape Cod and the offshore islands

In the southeast Cape Cod juts out into the ocean. This 65-mile-long appendage is rectangular in shape except at its easternmost point, where it hooks northward. Its offshore waters are among the most treacherous in the country. Henry David Thoreau wrote that to the people of Provincetown, at the tip of the cape, the sea is their garden and the dog that growls at their door is the Atlantic Ocean. Tufts of grass spring up along the sand dunes, and gnarled jack pines and scrub oaks, some only head high, grow in bunches. Off the southeastern coast lie the islands of Nantucket and Martha's Vineyard, lashed by the gray Atlantic in winter but in summer alive with thousands of tourists and longtime seasonal residents.

Central Massachusetts comprises rolling plains fed by innumerable streams. Beyond lies the broad and fertile Connecticut Valley and the Berkshire Hills. The now-paved Mohawk Trail crosses the Berkshires, the Hoosac Range on the east and the Taconic Range on the west. The state's highest point, 3,491 feet (1,064 metres), is Mount Greylock on the Taconic side near Adams. In North Adams a natural bridge of white marble has been formed by the wind and water, and at nearby Sutton is a half-mile-long gorge that knifes through the rock, exposing 600,000,000 years of geologic history.

Farmland along the Connecticut River near Sunderland, Mass.
Gene Ahrens/Shostal Associates

Rivers and lakes

Drainage. The land is veined with rivers—19 main systems, the best known of which are the Connecticut, Charles, and Merrimack. More than 1,100 ponds or lakes lie among the hollows of the hills, one in almost every one of the more than 350 communities. Many bear long Indian names, most notably Lake Chargoggagoggmanchauggagoggchaubunagungamaugg (which, translated, means "You Fish on Your Side; I Fish on My Side; Nobody Fishes in the Middle"). The best-known small body of water, however, is Walden Pond, immortalized by Thoreau.

Although clean-up efforts have lessened the pollution, nearly all the rivers and many lakes are unsuitable for swimming. This has long been true of the meandering Charles, which separates Boston and Cambridge and which is favoured by college rowing crews, canoeists, and sailboat enthusiasts.

Climate. The state has a temperate climate. The climate is colder but drier in western Massachusetts, although its winter snowfalls may be more severe than those nearer the coast. July is the hottest month, averaging about 71° F (22° C), in contrast to 26° F (−3° C) in January. Annual precipitation averages 42 inches (1,067 millimetres) in Boston and 44 to 45 inches in Worcester and Pittsfield.

Plant and animal life. Despite its industrialization, Massachusetts has preserved many of its forests, and there are now nearly 150 state forests, reservations, and parks. Public hunting grounds amount to some 23,000 acres (9,300 hectares). Three national wildlife refuges and the Cape Cod National Seashore allow further contact with nature. Not far from downtown Boston is the Arnold Arboretum, which has one of the largest collections of trees and shrubs in the United States.

Few large animals remain in the wild, but an occasional bear or moose is sighted. Other animals seen in the woods include deer, beaver, muskrat, mink, otter, snowshoe hare, red fox, woodchuck, raccoon, and chipmunk. Along the shores the sandpiper, blue heron, American egret, sanderling, and turnstone can be seen. Water birds include the gull, scoter, cormorant, and loon, while those most often seen on land are the kingfisher, warbler, bobwhite, brown thrasher, sparrow hawk, yellow-shafted flicker, and whippoorwill. Game birds include ruffed grouse, wild turkey, and pheasant.

Settlement patterns. The earliest settlements were along the seacoast, with the population most heavily concentrated in those towns that lay at the mouths of rivers. The

settlers fanned inland along these streams, drawing on them at first for farm use and later for the power to run mills. Early towns were also settled along the Connecticut River in the western part of the state.

Today, the lure of the sea results in nearly equal popularity for all of the towns along the coast, where sunbathing, swimming, yachting, and fishing are a way of life. Among these coastal towns are Plymouth, with its long harbour; Duxbury, Marshfield, Scituate, and Cohasset, where the first suburbs sprang up in colonial days; the boating bays from Hingham to Boston; the beaches at Revere and Lynn; Gloucester and Cape Ann, famous for fishing; and Marblehead, the yachting capital of the world.

Coastal towns

Boston is surrounded by communities many of whose residents work in the city by day, sleep in the suburbs by night, and clog the highways commuting back and forth. Other urban cores include Springfield, Worcester, Fall River–New Bedford, Lowell–Lawrence, Pittsfield, and Fitchburg. These cities, which grew large during the Industrial Revolution, have since declined in population.

The people. *Ethnic composition.* Boston and San Francisco are often referred to as sister cities because of such similarities as busy oceanside ports, good restaurants, emphasis on culture, prominence of religion in civic and social life, fine architecture, and echoes of a colourful history. Their ethnic mix is comparable as well. The blend of peoples in Boston has spread across the state. Although Boston is heavily Irish, so too are the urban areas in western Massachusetts, and the native brogue of Irish immigrants is more likely to be heard around Springfield, Westfield, and Holyoke than it is in Boston. In the first half of the 20th century, large numbers of Italians followed the Irish immigrations of the 1800s. By the 1960s and '70s the newcomers to the state were mostly Spanish-speaking, from Cuba and Puerto Rico.

The English stock that still forms the backbone of the population is intermingled with Slovaks, Poles, Canadians, Russian Jews, Greeks, Scandinavians, Syrians, Germans, French, and Chinese. Fall River and New Bedford are the homes of many Portuguese and Cape Verdians. Blacks are concentrated mainly in the Roxbury and Dorchester sections of Boston and in Cambridge and Springfield.

Few Indians remain in Massachusetts, despite the prevalence of Indian place-names. The state was named by Captain John Smith for the Massachuset tribe, whose name meant "Near the Great Hill," believed to refer to Blue Hill that rises south of Boston in an otherwise flat area.

Religion. Massachusetts is now largely Roman Catholic, though its religious foundation was solidly Protestant. The Pilgrims, who established the Plymouth colony in 1620, and the Puritan settlers came to Massachusetts mainly for religious reasons. The *Mayflower* colonists were a small groups of Separatists who had fled to Holland from England to practice their religion without official interference. Economic hardship and a desire to establish an identity free of Dutch influence prompted them to seek out America. The Puritans, persecuted in England because of their religious beliefs, sought ecclesiastical reform, but only within the church structure. They were not advocates of religious tolerance, as other Protestant groups and radical thinkers discovered. Many with differing religious views were banished, including Roger Williams from Salem and Anne Hutchinson from Boston. Unrepentent Quakers and Anabaptists were banished, and a few were executed.

Religious Separatists and other dissenters

The Puritan Massachusetts Bay Company essentially established a theocracy, with close ties between the government and the clergy. The leaders felt comfortable not only in establishing patterns of government by interpreting the colony charter but also in interpreting the will of God for the people. The arrangement fell short of its purpose. When in 1634 Governor John Winthrop refused to call a meeting of the legislature, or General Court, the freemen demanded to see the charter. He acceded, divulging his infringement on the rights of the legislature, and a bill was quickly passed vesting governmental power in the freemen.

During the subsequent development of Massachusetts, religion remained important. The Puritan Congregational church was not formally disestablished until an amendment to the state constitution was passed in 1833. Following colonial patterns, churches often are found in the most prominent places of the towns and villages, symbolizing their traditional role in social life.

Social hierarchies. All the people of Massachusetts may be created equal, but some gained an edge earlier than others. The variety of peoples in the state fails to alter the fact that the major concentration of wealth and power continues to be controlled by the 800 or so families who trace their pedigrees to the *Mayflower* and those who in the following centuries so successfully trod the avenues of commerce, finance, and culture that they came to be considered among the ranks of that still-relevant cadre, the Proper Bostonians. Many descendants of later immigrants also have found their way to the top of the financial—and often political—ladder. The proud tradition of family participation in the building of the state extends also to those of less-exalted position: a large proportion of the residents of small-town Massachusetts, especially in the west, can claim many generations of Yankee background.

The economy. The economy of Massachusetts today is based largely on technological research and development, service industries, and tourism. This represents a major shift from the state's preindustrial agricultural basis in the 17th and 18th centuries and the heavy manufacturing that characterized the 19th and the first half of the 20th centuries.

Fishing and agriculture. Foreign trade, fishing, and agriculture long buoyed the economy. Salem sailors brought exotic goods from China, the West Indies, and other faraway lands. Fishing was lucrative, adventuresome, and dangerous—more than 10,000 Gloucester fishermen have lost their lives over the centuries. Fishing and shipbuilding went hand in hand. Between 1789 and 1810 the Massachusetts fleet grew 10-fold, some of it to aid in defense against British and French aggressions on the high seas. Yankee sailors also found much "black gold" in the slave trade between West Africa and Southern ports.

The golden age of the Yankee seafarers

At the height of the whaling boom in the 19th century, 329 whaling vessels sailed from New Bedford, in addition to others from Nantucket and other ports, bringing in $10,000,000 worth of cargo each year in their holds. This great industry was not to last, however: by the turn of the century its contribution to the state's economy had dwindled to only a fraction of its former importance. Fishing later suffered substantial reverses as well. A $42,000,000 annual business in the early 1960s, fishing began to wane late in the decade because of foreign competition in the traditional Atlantic fishing grounds and the depletion from overfishing of such species as haddock and lobster. By the late 1970s, however, the industry had made a comeback, and Massachusetts usually ranks among the top three or four U.S. states in value of fish landings.

The generally rocky soils support only truck gardening, although the purple sandy bogs of southeastern Massachusetts and Cape Cod produce about 50 percent of the U.S. cranberry supply. Greenhouse and nursery products are the main source of farm income, followed by dairy products.

Industry. Massachusetts has been a manufacturing state since the early 1640s, when John Winthrop, Jr. (son of Governor Winthrop), opened a saltworks in Beverly and ironworks in Saugus and Quincy. Francis Cabot Lowell was largely responsible, however, for raising the state to its manufacturing eminence. Lowell went to England to study methods of textile operations and built a power loom in Waltham in 1814. He died in 1817, but his associates developed Lowell, the city built of bricks, with its mills driven by the Merrimack River.

Yankee ingenuity fostered much early handicraft-based industry, though the influx of unskilled, low-paid labourers from Europe during the 19th century was the necessary ingredient for the mass production that developed in the state's shoe and textile factories. One of the first and largest shoe plants in America was the United Shoe Machinery Corporation in Beverly, while the building of the Springfield armory in 1777 boosted industry in western Massachusetts at the same time that it aided the Revolutionary cause. Other well-known goods from

Massachusetts factories included watches from Waltham, Salem, and Boston; rockers from Gardner; cutlery and hand tools from Greenfield; guns and motorcycles from Springfield; leather goods from Peabody; shovels (which were used by the forty-niners during the California gold rush) from North Eaton; envelopes from Worcester; paper from Holyoke; silverware from Newburyport; and razor blades from Boston.

Electronics and communications

Today, the electronics and communications industries draw heavily upon the many educational institutions in and around Boston. The suburbs of Boston have become known for their research-and-development facilities, which have contributed significantly to computer technology. Copper and iron were once mined in Massachusetts, but mineral production is currently limited to sand and gravel, stone, and clay.

Transportation. "Never, in these United States, has the brain of man conceived, or the hand of man fashioned, so perfect a thing as the clipper ship," wrote historian Samuel Eliot Morison in *Maritime History of Massachusetts.* All clipper ships were built between 1850 and 1855, and from then on a kind of World Series of ship racing began. The champion was Donald McKay's *Flying Cloud,* which sailed to San Francisco in 89 days, went 374 miles in one day, and averaged 13.5 knots over four days. Records were not the only motivating factor: the clippers carried 1,700 tons of cargo.

Symbolic of Massachusetts' close relation to the sea, the first lighthouse in the United States, Boston Light, was built off that busy port in 1716.

Water formed the Bay State's highway system for 200 years. Rivers such as the Connecticut and Merrimack and man-made canals such as the Middlesex served early needs well. The Boston Post Road and the Mohawk Trail were the most heavily traveled of the early roadways. Opened to Boston–New York mail in 1673, the Post Road consisted of three routes. The Mohawk, an Indian footpath that was converted to an ox road by the settlers, became the first interstate toll-free road, called Shunpike, in 1786.

In 1826 the nation's first railroad brought granite from the quarries of Quincy and Charlestown for the building of the Bunker Hill Monument in Charlestown. The cars were horse drawn. A steam railroad connected Springfield and Worcester in 1839, and 15 systems were shuttling freight among western Massachusetts cities by 1855. Among the most impressive feats of early railroad building was the 4.5-mile Hoosac Tunnel, drilled under the Hoosac Range between 1851 and 1875. The first electric street railway was built in Brockton, and Boston had the nation's first passenger subway, as well as an elevated system. Boston's Logan International Airport, stretching parallel to the harbour, is one of the few large air terminals in close proximity to a major city.

Administration and social conditions. *Government.* From the Mayflower Compact, drawn up by the Pilgrims in 1620 when the concept of "the divine right of kings" dominated Europe and the idea of self-government was little more than an exotic notion, a form of government evolved of which the people could feel themselves a part. In 1630 the Puritans settled in Massachusetts under the authority of a charter granted them by King Charles I of England. The charter was similar to those of many other trading companies, except that it allowed the officers of the company to meet in Massachusetts rather than in England. The relative isolation of the colony and the lack of interference from England allowed the development of a virtually autonomous government. The state's legislature, known officially as the Great and General Court, traces its origins to John Winthrop and his 18 assistants. In a dispute over a stray pig, the court became bicameral in 1644. The assistants became the upper chamber, while two deputies elected from each town constituted the House.

The state constitution

After independence was declared, the General Court drew up a constitution for the State of Massachusetts Bay. It was rejected by the people, in part because it lacked a declaration of rights but also because it was not written by an elected constitutional convention. In 1779 a new convention was elected and convened in Cambridge. John Adams was the principal author of the new constitution of the Commonwealth of Massachusetts, which was ratified in 1780. Many of its features were directly incorporated into the federal Constitution. One of its provisions permits the governor and his council or the legislature to seek advisory opinions on questions pertaining to the scope of gubernatorial or legislative power from justices of the Supreme Judicial Court. Today, Massachusetts is the only one of the 13 original states still governed under its first constitution, which is the oldest governing constitution in the world. It has, however, been amended many times.

The first meeting of the General Court as the legislative body of the new state took place in October 1780, exactly 150 years after the first meeting of the Puritans' Great and General Court. It today comprises 40 senators and 160 representatives; both houses serve two-year terms. The state's judiciary mainly divides into the district courts for handling minor matters, superior courts for trial by jury, and the Supreme Judicial Court. Justices are appointed by the governor with the advice and consent of the Executive Council. The legislature establishes details concerning the arrangement and operation of the judicial system.

The legislative process in Massachusetts is characterized by several distinctive features: legislation may be introduced by citizen petitions; bills are referred to the appropriate joint standing committee prior to legislative debate; public hearings are held by the committees for most legislative proposals; and the committees are required to submit a report to the General Court for each bill, with their recommendation.

Another political phenomenon that grew up shortly after the settlers arrived was the town meeting, which started as a forum for settling local quarrels and grew to what is in many smaller towns the community event of the year. (As the poet James Russell Lowell observed, "Puritanism, believing itself quick with the seed of religious liberty, laid, without knowing it, the egg of democracy.") The first recorded meeting was in Dorchester in 1633, when citizens were summoned by the roll of a drum. A year later Charlestown organized the first Board of Selectmen, the emergence of such local government balancing the power of the colony's executive. The relationship between the General Court and municipal governments, which was not specified in the constitution, was finally addressed with the passage of the Home Rule Amendment in 1966. This provides that municipalities can make changes in their government without securing the permission of the legislature. A county system was patterned after the English model, in which the greater powers reside in townships and cities rather than in the counties, which serve chiefly judicial purposes.

Massachusetts politics is now dominated by the Democratic Party. Fueled by the massive Irish immigration of the 1840s and '50s, the Democrats slowly broke the Republican monopoly on state political offices. During the late 19th and early 20th centuries, politics became a means to a better life—to a place alongside the "Boston Brahmins" of *Mayflower* heritage—for the Irish and other immigrant groups who experienced great discrimination and hostility. In 1881 Lawrence became the first major city to elect an Irish Catholic mayor; Boston followed suit in 1884. The Boston Irish politician has become legendary, mostly because of Mayor (and Governor) James Michael Curley, a skillful orator from a lowly background who was jailed twice, once while in office.

The state has played an important role in national politics. It has contributed five presidents—John Adams, John Quincy Adams, Calvin Coolidge, John F. Kennedy, and George Bush—as well as a great number of Cabinet officers, career officers, diplomats, and congressional leaders.

Education. Education lies close to the heart of Massachusetts' social and cultural life. Harvard College (now Harvard University) was founded in 1636 in New Towne (now Cambridge). Although it was designed originally to provide the wilderness colony with a continuing supply of trained clergy rather than an educated lay population, its graduates became community leaders and schooling soon was provided colonywide. In 1647, towns with 50 householders were required to support an elementary school; those with 100, a secondary school.

Founding of Harvard University

Massachusetts became a pioneer as well in kindergarten and secondary education and developed a uniform state public school system in 1840. The state has numerous private preparatory schools of national ranking. Roxbury Latin School, founded in 1645, is among the nation's oldest.

Many of the nation's oldest and most prestigious institutions of higher learning, in addition to Harvard, are located in Massachusetts. The largest, both in Boston, are Boston University (1839) and Northeastern University (1898). Nearby are the Massachusetts Institute of Technology (Cambridge; 1861) and Tufts (Medford; 1852) and Brandeis (Waltham; 1948) universities. Amherst (Amherst; 1821) and Williams (Williamstown; 1793) colleges have perpetuated traditions of academic excellence at small schools, while Mount Holyoke (South Hadley; 1837), Wellesley (Wellesley; 1870), Smith (Northampton; 1871), and Radcliffe (Cambridge; 1879) colleges have been pioneers in women's education. Boston College (Chestnut Hill; 1863) and College of the Holy Cross (Worcester; 1843) are major Roman Catholic institutions. The University of Massachusetts (Amherst and Boston; 1863) is the principal state university.

World medical centre

Health and welfare. Massachusetts is one of the chief medical centres of the world, particularly in the area of specialists and specialty hospitals. It has also been a leader in research, notably at Boston's Children's Cancer Research Foundation. An urgent contemporary challenge is the delivery of medical services to the poor, who have been affected by urbanization and the gradual disappearance of the family physician.

The state was an early leader in the fight to improve social conditions. Regulatory laws were passed, beginning in the early 19th century, to protect residents. State boards, under the supervision of the governor, later grew out of the need to improve conditions in health, education, welfare, labour, banking, insurance, and prisons. The state recognized its responsibility as early as 1818, when it opened an asylum for the insane. The first state almshouse opened in 1854, while the nation's first public health hospital for tubercular patients began operations in 1898. Despite these early gains, care for the mentally ill, alcoholics, addicts, the homeless, and juvenile delinquents remains a problem.

Following colonial tradition, welfare remained the province of the municipalities until it was taken over by the state in 1970. Although the new program was fraught with difficulty, it was an improvement over the system that existed in the mid-19th century, when citizens were still imprisoned for debt. State programs now stress training and educating welfare recipients to allow them to become self-sufficient. Massachusetts' welfare payments are well above the national average.

Cultural life. The blending of an Old World heritage and a New World spirit produced a bountiful cultural environment in Massachusetts.

Place in American arts and letters

Literature was virtually lost in the leaden language of the early writings and sermons, though poets such as Anne Bradstreet and Edward Taylor rose well above the level of dogmatizing, and Jonathan Edwards combined taut language and a brilliant theological mind. During what has been called the American renaissance, however, beginning around the time of the Revolution and lasting through much of the 19th century, the state nourished many writers who might be said to have founded the bases of American literature—and who brought it recognition outside the new nation.

The group of writers who brought fame to Concord are an indication of the inspiration of this period. A deep sense of both community responsibility and individualism may be traced through the writings of Ralph Waldo Emerson, Henry David Thoreau, Nathaniel Hawthorne, and Louisa May Alcott, all of whom were neighbours. The eloquence of Emerson, preacher, philosopher, and poet, carried his concepts of individual spiritual freedom to faraway lands, while Hawthorne found tranquillity in the small town after growing up in the shadows of Salem witches near the House of Seven Gables.

The mountains of Pittsfield also provided a congenial working environment for Hawthorne, as well as for Herman Melville, Oliver Wendell Holmes, and Henry Wadsworth Longfellow (the latter two combining, respectively, medicine and scholarship with their writings). Among other famous writers of the era were John Greenleaf Whittier and James Russell Lowell, as well as Emily Dickinson, today generally acclaimed as one of the finest American poets of the 19th century.

Performing arts

The universities have become central to many of the performing arts (theatre, dance, and music) in Massachusetts, and the Boston Symphony Orchestra generally is regarded as among the finest musical ensembles in the world. Its Tanglewood concerts at Lenox in the Berkshires (begun in 1938) are, with the Jacob's Pillow Dance Festival at nearby Becket, among the major attractions of the New England summer.

Boston's museums appeal to a variety of interests, ranging from the renowned collections of the Museum of Fine Arts, to the Computer Museum, to the John F. Kennedy Library and Museum. Visitors with children often gravitate toward the Museum of Science, the New England Aquarium, and the Children's Museum, which pioneered the use of participatory exhibits. Important collections of historical records are held by the Massachusetts Historical Society, the Massachusetts Archives, the Boston Athenaeum, the Boston Public Library, and the New England Historic Genealogical Society.

Historical sites in Boston draw many tourists. The Freedom Trail provides a trip that includes Boston Common, the old and new (1795) state houses, Park Street Church, the Old Granary Burying Ground, Old Corner Bookstore, Faneuil Hall, Paul Revere House, the Old North Church, and the USS *Constitution,* better known as *Old Ironsides.*

Outside the capital the past seems still alive in three villages: Plimoth Plantation, Old Sturbridge Village, and Shaker Village in Hancock, where the sect established its communal-church concept in the 1780s. Harvard Square in Cambridge is a favourite tourist stop for its potpourri of people, its proximity to Harvard and the Massachusetts Institute of Technology, and the history imbedded in the cobblestone atmosphere along its narrow side streets.

Salem prefers to forget its witch-hunting period of the late 17th century, but visitors to the House of Seven Gables and other "haunted houses" keep the memories alive. In the elm-shrouded Chestnut Street area are Federal-style homes that reflect the days of prosperity for merchants, shipowners, diplomats, congressmen, and writers.

Along the South Shore are Quincy, where the humble homes of the eminent Adams family are located next door to one another, and Hingham, where the Old Ship Church is the oldest surviving church of the 13 colonies. The Bourne Whaling Museum in New Bedford includes a half-size reproduction of a whaling vessel and some 600 logbooks, and the Seamen's Bethel (chapel) there was immortalized by Melville in *Moby Dick.*

Historic western Massachusetts

West of Boston lies Concord and its Old Manse, home of the Emersons and, for four years, of the Hawthornes. Past the Old Mill and Longfellow's Wayside Inn in Sudbury are Worcester and then Springfield, where the armoury and arsenal are reminders of the city's famous rifle. In nearby Pelham the Town Hall complex has the oldest continuously used meetinghouse in the country and a monument to Captain Daniel Shays, who led a rebellion of poor farmers in 1786. Chesterwood in Stockbridge was the site of the studio of Daniel Chester French, sculptor of the great seated Lincoln statue in Washington's Lincoln Memorial. Some of the doors of houses in Old Deerfield bear the marks of Indian tomahawks wielded during the raids of the early 18th century.

Private clubs, both social and athletic, long have been Massachusetts institutions, especially for golf, tennis, and yachting. Among the most exclusive are the Brookline Country Club and the Longwood Cricket Club in Brookline, the Myopia Hunt Club in Hamilton, and various yacht clubs along the North Shore above Boston, particularly in Marblehead.

Athletics have come to form a subculture among all social classes. The professional teams—Boston's Red Sox in baseball, Bruins in hockey, and Celtics in basketball and

the Foxboro-based New England Patriots in football—attract the most attention, but the state also gives considerable emphasis to high school and college athletics.

HISTORY

Although the landing of the Pilgrims on Nov. 21, 1620, was important, Indians had found this corner of the country some 3,500 years earlier, and Leif Eriksson and his Norsemen may have landed somewhere in the Cape Cod region in 1003. European seafarers tapped the fertile fishing areas throughout the 1500s, the French explorer Samuel de Champlain mapped the area in 1605, and in 1614 Captain John Smith of the Virginia colony drafted a detailed map of the New England coast from Penobscot Bay in Maine to Cape Cod.

European settlement. Prior to 1685 there were two separate colonies within the boundaries of present-day Massachusetts. The area around Plymouth and Cape Cod, settled by the Pilgrims, was known as Plymouth colony, or the Old Colony. By the mid-1640s its population numbered about 3,000 people. The Pilgrims were never granted a royal charter; their government was based on the Mayflower Compact, a document signed by 41 male passengers on the *Mayflower* five weeks before their arrival in the New World. The compact was hardly democratic since it called for rule by the elite, but it established an elective system and a basis for limited consent of the governed as the source of authority. The Old Colony was rapidly overshadowed by its Puritan neighbour to the north, the Massachusetts Bay Company. Fueled by the Puritan migration of the 1630s, and an aggressive sense of authority, the Massachusetts Bay colony expanded rapidly. By the mid-1640s it numbered more than 20,000 people, and it began absorbing settlements in Maine and New Hampshire. The government of the colony was based on a providential interpretation of the royal charter granted by King Charles I, which was transferred to the new settlement by John Winthrop. The exhortation by Winthrop, "For wee must Consider that wee shall be as a Citty uppon a Hill, the eies of all people are uppon us . . . ," underlines the strength of conviction of the Puritan mission.

The two colonies

The Puritan government often operated as an independent state, to the point of minting its own money and even conducting its own foreign affairs. King Charles II finally abrogated the colony's charter in 1684 for repeatedly overstepping its authority. In 1691 a new charter was granted to the Province of Massachusetts Bay after the Glorious Revolution brought William and Mary to the throne in England. The new province formally united Plymouth, Maine, and the islands of Nantucket and Martha's Vineyard with Massachusetts—a configuration that remained until 1820, when Maine was established as a separate state.

Settlers had feared Massachusetts for its hostile Indians, but until 1675 relative peace prevailed because of a pact with Massasoit, chief of the Wampanoag. This accord was ended by King Philip (Metacom), Massasoit's son. His open warfare, King Philip's War (1675–76), ended with his own death, but only after hundreds of settlers had been killed and some 50 towns raided in southeastern and central Massachusetts. Repeated expeditions against the Indians were common in the 18th century, as Massachusetts men joined with British troops to fight the French and their Indian allies.

Commercial and industrial expansion marked 18th-century Massachusetts and resulted in the rapid settlement of new communities, many spurred by speculation. Between 1692 and 1765, 111 new towns and districts were incorporated, while the population increased to 222,563.

Revolutionary period and statehood. "The shot heard round the world" initiated a new order in Massachusetts and her sister provinces. The struggle had actually begun several years earlier, when a new spirit grew out of years of physical struggle and radical ideas involving such concepts as equality, freedom, and unity. Events in Boston—the fight against the writs of assistance, the Boston Massacre, the Boston Tea Party and resulting closure of the port of Boston, the Battles of Lexington and Concord, Bunker Hill, and the evacuation of the British troops from Boston—inspired song and verse that came to typify the spirit of the Revolutionary era. Agrarian unrest in 1786–87 resulted in the only military threat to the new commonwealth. Governor James Bowdoin was forced to call out a special state army of 4,400 men to suppress Shays's Rebellion. The unrest and fear generated by the armed insurrection probably helped advance support for the ratification of the new U.S. Constitution. A year later, in 1788, Massachusetts became the sixth state to ratify the Constitution.

Revolutionary War Battles

Massachusetts was in the forefront of the Industrial Revolution, and the resulting changes guaranteed that by the mid-19th century the state would be vastly different from its colonial antecedents. A decline in agricultural lands fostered both a migration away from Massachusetts and the development of large-scale manufacturing enterprises producing textiles, shoes, and machinery. The rural outlook of the state was lost with the rise of a number of urban areas, connected by turnpikes, canals, and, later, railroads. The shattering of ethnic and religious homogeneity through immigrant migration, especially the arrival of the Irish, accentuated these changes. Property requirements were removed for voters, the Congregational church was disestablished, black Massachusetts regiments fought in the American Civil War, and Irish politicians began to be elected to public office. The population of Massachusetts continued to expand, although at a slower rate than the rest of the country, until by 1860 it had become the second most densely populated state in the nation.

The 20th century. The consequences of the Industrial Revolution—increasing urbanization, an economy based on manufacturing, and a large immigrant population of low-paid workers—had a major impact on Massachusetts in the 20th century. Most noticeable was the shift of the textile and shoe industries out of Massachusetts to Southern and Midwestern states. Labour unrest, economic stagnation, and urban decay followed. The two world wars brought only brief respites from this decline. The advent of the electronics and communications industries after World War II finally brought this cycle to a halt. Aided by federal money for research and development, numerous small corporations began to draw on the expertise of academics from Boston and Cambridge. High-technology industry began as a suburban phenomenon, but it has also revitalized many of the larger cities, with their large mill complexes now home to numerous research and development firms. This renewal has allowed Massachusetts to maintain its financial, educational, and cultural prominence. (J.S.D./M.L.Cl.)

Economic changes

New Hampshire

New Hampshire, one of the 13 original U.S. states, is located in New England at the extreme northeastern corner of the nation. With an area of 9,279 square miles (24,032 square kilometres), it is bounded on the north by the Canadian province of Quebec, on the east by Maine and a 16-mile (25-kilometre) stretch of the Atlantic Ocean, on the south by Massachusetts, and on the west by Vermont. The capital is Concord, located in the south central part of the state.

The Granite State, as New Hampshire is popularly known, is a study in contrasts. Since the late 19th century it has been among the half-dozen most-industrialized states in the Union, yet it is frequently portrayed as agricultural and pastoral. Vermont and New Hampshire supposedly constitute a "Yankee Kingdom" dominated by WASPs (white Anglo-Saxon Protestants), yet the state has a large population of residents with French-Canadian, German, Italian, Polish, and other non-English ancestors. Its political reputation is probusiness and conservative, yet the single largest internal source of state funds is a business profits tax. New Hampshire's regional subdivisions are so distinct that numerous people have suggested it be divided in thirds, with roughly equal parts being added to Maine, Vermont, and Massachusetts.

Despite these contrasts, the state has developed a distinct identity. Central to that identity is the image of governmental frugality: New Hampshire has no general sales or individual income taxes but still manages regularly to have

a budget surplus. Frugality at the state level has accentuated the dispersal of responsibility to towns. Although town governments exist in all the New England states, in no state do they carry as much authority as in New Hampshire. Still another component of that identity is a craggy adherence to tradition, symbolized powerfully by the rock profile in Franconia Notch known as the Old Man of the Mountain. The combination of frugality, decentralization, traditionalism, industrialization, ethnicity, and geographic diversity makes New Hampshire very attractive to many Americans. East of the Mississippi, only Florida and Delaware have outpaced it in rate of population growth since 1950.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The basic physical features of New Hampshire are the result of the last glacial age (approximately 70,000 to 10,000 years ago), during which the Wisconsin ice sheet moved like a huge bulldozer across New England from the northwest to the southeast. Loose sand, silt, clay, and gravel were deposited as masses of glacial till that, near the town of Greenland, are 395 feet (120 metres) in depth. The mountain notches of New Hampshire—Crawford, Dixville, Franconia, and Pinkham—are the result of the glacial action, as are the potholes and cirques found in the state. The great glacier left many deltas and hillocks of stratified deposits. New Hampshire's lakes are also the results of glacial action.

Glacier-formed topography

The mountains are the most striking feature of New Hampshire's landscape. There are about 1,500 classified elevations, including eight mountains rising more than a mile in altitude, 61 rising more than three-quarters of a mile in height, and 157 with an elevation of a half mile or more. The best-known is Mount Washington, at 6,288 feet (1,917 metres) the third highest peak in the nation east of the Mississippi River. The mean elevation of the state is about 1,000 feet above sea level.

Drainage and soils. New Hampshire has five main river-drainage basins. The largest is that of the Merrimack River in the central part of the state. Second in size is the Connecticut River drainage basin along the western border. The remaining waters flow into the Saco, Piscataqua, and Androscoggin rivers, known collectively as the coastal rivers, as well as into several smaller streams. There are some rich deposits of deep soil along these rivers, but in general the soils within the state are rocky, thin, and difficult to farm.

Climate. New Hampshire's climate is highly varied. In winter temperatures may drop below 0° F (−18° C) for days at a time. Summers are relatively cool, and the mean annual temperature is about 44° F (6.7° C). Annual precipitation is approximately 42 inches (1,067 millimetres) and is rather evenly distributed over the four seasons.

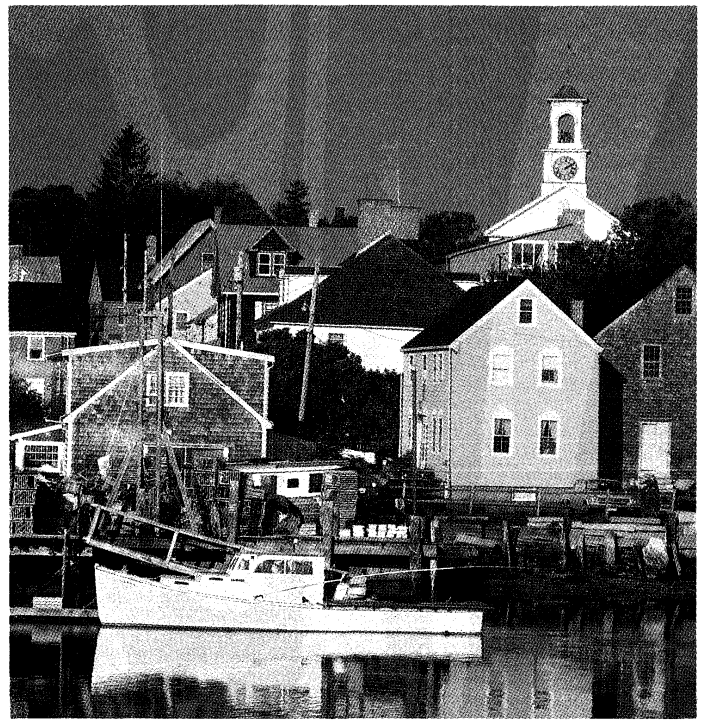
Plant and animal life. More than 80 percent of New Hampshire remains under forest cover. The majority of the trees found in the eastern United States are indigenous to the state. The most valuable single species of tree has always been the white pine.

The wooded areas support a flourishing wildlife. Whitetail deer are numerous everywhere; the moose herd is limited largely to Coos county. There is a hunting season for deer each year, but the killing of moose is prohibited. Beaver, once almost exterminated, have been protected and are making a comeback. Black bear are relatively common, while smaller mammals like rabbit, squirrel, raccoon, fox, and mink are plentiful. There is an abundance of birdlife, including such species as grouse, woodcock, pheasant, and duck. State rearing stations keep the interior lakes and rivers well stocked for fishing. There has been much concern about the effects of pollution on aquatic life, and strenuous efforts, both public and private, are under way to prevent further contamination of lakes, streams, and coastal waters.

Settlement patterns. New Hampshire has six distinct regions, each deeply rooted in the state's history. The heavily forested White Mountain area in the north is popular with mountaineers and tourists in summer and winter alike. The lakes region around Lake Winnipesaukee is a favoured locality for summer camps and resorts and for water sports. The seacoast region, which includes Portsmouth, Dover, Exeter, and Hampton, has many maritime activities. The south central, or Merrimack, region surrounds Manchester and Nashua and is the most heavily industrialized section of the state. The Dartmouth–Sunapee Lake region in the west central portion of New Hampshire is dotted with educational institutions and summer homes. The area around Mount Monadnock, in the southwestern corner of the state, is noted for many small industries and such attractions as the MacDowell Colony, a residential retreat in Peterborough for artists, and for the Cathedral of the Pines in Rindge, an outdoor shrine dedicated to the nation's war dead. Each region is officially organized and finances its own promotional activities. The 11 percent of the state's area that constitutes the White Mountain National Forest is almost uninhabited.

Major regions

Craig Blouin—New England Stock Photo



Fishing boat at the harbour at Portsmouth, N.H.

The people. In the first U.S. census, in 1790, New Hampshire had a population of 141,885. Since then each decennial count has recorded a growth, except for that of 1870, when there was an extensive post-Civil War exodus to the Midwest. The urban population is concentrated to a large extent in the southern and southeastern regions, and the larger urban centres are all located south of the White Mountains. In the area nearest Boston, New Hampshire has become a bedroom suburb for thousands of commuters to that metropolis.

In the colonial period the majority of the people were of English origin, but a significant influx of Scotch-Irish, who were largely Presbyterian in faith, began in 1719. They settled in the south central and southwestern portions of New Hampshire and named their principal towns Derry, Londonderry, Antrim, and Dublin.

New Hampshire had a system of town churches in which any officially recognized denomination could be designated at the annual town meeting to receive public tax support. Prior to the American Revolution five denominations were officially recognized: Congregational, Baptist, Presbyterian, Quaker, and Church of England. The system was discarded in 1819 by the Toleration Act passed by the legislature. Since then all churches have been privately supported, and any denomination may function freely.

During the 19th and 20th centuries waves of immigrants came into the state from central and eastern Europe. The first Roman Catholic congregation was established in 1823, the first Roman Catholic school in 1859, and a statewide diocese in 1884. The first Jewish congregation was orga-

nized in 1892 and the first Greek Orthodox church in 1905. The New Hampshire Council of Churches, organized in 1945, has developed broad ecumenical policies to include many faiths.

Blacks comprise less than 1 percent of the population, and those of Asian origin are even fewer. The largest group not directly descended from origins in the British Isles are the French Canadians, or Canado-Americaines, who first began to arrive in the years immediately after the Civil War, chiefly from Quebec. They were attracted mainly to such industrial cities as Manchester, Nashua, Laconia, and Berlin. By the end of World War I, New Hampshire had the largest percentage of French Canadians among the states.

French-Canadian population

The economy. The historic shoemaking, woodworking, apparel, and textile industries have declined in productivity and employment, while space-age industries have grown rapidly. Among nonmanufacturing industries that also have grown significantly are public utilities, insurance and banking, medical and health service establishments, miscellaneous business services, and wholesale and retail trades. Tourism is also of major importance.

Agriculture, forestry, and fisheries. Agricultural acreage has decreased by two-thirds in the 20th century. Dairy products are the chief source of agricultural income, followed by ornamental and greenhouse plants and a variety of fruits and vegetables. More than half of the timber harvested is used for sawlogs, and nearly 40 percent for pulpwood to supply paper and newsprint industries. The state ranks low in commercial fishing yields but has marketable catches of lobsters and deep-sea fish.

Industry. Machinery and electrical- and electronic-goods manufacturing has largely supplanted the textile and garment industry, much of which either closed or moved to the South after 1945. Paper and wood products are the next largest industries. The only large industrial centre north of the White Mountains is Berlin, with large paper and pulp mills.

There are many rock quarries, chiefly of granite, throughout New Hampshire. Sand and gravel are the major minerals produced in terms of both tonnage and value. Other native minerals that have commercial value include garnet, zircon, beryl, and bog iron.

In 1955 the state legislature established an Industrial Park Authority, now the Industrial Development Authority, to attract more business into New Hampshire. Supplemented by the endeavours of individual cities and towns and by various private organizations, the authority has been successful in modernizing the industrial and business life of the state. An organized effort began in 1925 to publicize New Hampshire as a tourist area, and this effort has steadily expanded. With the increase in popularity of winter sports following World War II, New Hampshire became as attractive to visitors in the cold months as it had been in the summer.

Winter sports

Transportation. The New Hampshire railroad network is now barely one-fourth its former length. Hundreds of miles of right-of-way have been abandoned, and there is no passenger service except on a six-mile cog railway on Mount Washington. The line opened in 1869 as the first construction of its kind in the world, and it is now used to carry tourists behind steam locomotives during the summer. As the railroad disappeared, the bus and highway systems of the state greatly increased. Several branches of the interstate highway system have given New Hampshire residents easy access to the major metropolitan areas of the Northeast. Regularly scheduled commercial air carriers serve airports in Keene, Lebanon, and Manchester; New Hampshire also has dozens of smaller landing strips for private planes. The transportation system as a whole makes it possible for many people to live in New Hampshire while working out of state and for the important tourist industry to function smoothly.

Administration and social conditions. *Government.* The constitution of New Hampshire, the second oldest among the 50 states, was adopted in 1784. Every 10 years the residents may vote on the question of holding a convention to consider modifications of the constitution. Proposals that pass these conventions must be approved by two-thirds of the voters at a popular referendum. Several conventions have been held in the state's history, and a large number of modifications have been adopted, but a considerable portion of the original constitution remains intact.

The governor is elected for a two-year term and is assisted by a five-member executive council, a survival from the colonial era. Members, elected every two years from five geographic districts in the state, must approve most appointments to state offices and all appointments to judicial posts.

The state legislature, the General Court, comprises more than 400 members elected every two years. The lower body, the House of Representatives, has between 375 and 400 delegates, depending on changes in population. It is thus one of the largest legislative bodies in the English-speaking world. The state Senate has 24 members.

The state's judicial functions take place in three levels of courts: municipal and district courts, county superior and probate courts, and the state Supreme Court. All justices are appointed by the governor and executive council and serve during good behaviour or until they are 70 years old. The state police, in addition to their usual duties, are at the disposal of the governor for emergency assignments.

Counties, cities, and towns are the units of organization for government below the state level. The 10 counties have few nonjudicial functions. Each of the 13 cities has its own charter providing for a mayor–council, a council–manager, or a commission system of government. Virtually all the 221 incorporated towns rely on the annual town meeting to guide policy. Towns have a variety of elected and appointed committees, but those matters of general concern recommended by these advisory groups appear as articles in a warrant distributed before the town meeting. With an elected moderator presiding, the town votes on every article. It also elects a wide range of local officials. Elected selectmen, from three to five in number, manage daily affairs. In larger towns they are authorized to have a manager to assist them.

Local government

New Hampshire has an unusual system for raising revenue. The state has clung tenaciously to its distinction of having neither a general sales tax nor a comprehensive state individual income tax. Because it raises some money from dog and horse racing, from state-operated liquor stores, and from a lottery, New Hampshire has a reputation for relying on "sin" taxes. The state actually derives only a small portion of its revenue from these sources, however. Much more important are a business profits tax, license fees, and taxes on meals, lodging, and motor fuel. The state provides the standard health and social welfare services; the provisions, however, are not as extensive as in adjoining states. Administration of the welfare system rests primarily with the towns, as does the responsibility for financing public elementary and secondary education. Local property taxes provide the bulk of this financing. The entire tax structure became a matter of intense partisan dispute in the 1960s and '70s. State budgetary surpluses in the 1980s and the advantages of local financial control have silenced many critics of the current system.

New Hampshire has always been a two-party state. Prior to the Civil War, New Hampshire was overwhelmingly Democratic, but from then to the depression of the 1930s it was dominated by the Republicans. The voters, however, have opted on several occasions for Democratic presidential candidates—Woodrow Wilson, Franklin D. Roosevelt, and Lyndon B. Johnson—and have sent a number of Democrats to Congress.

Education. New Hampshire has had a public school system since 1647, when, as a part of Massachusetts, it was required to provide different kinds of schools depending on community size. The statewide system is administered by a board of education headed by a commissioner. Each town is constituted a school district, with its main support coming from local property taxes. Towns frequently join to form consolidated public high-school districts. The land-grant college that became the University of New Hampshire was founded in 1866 in Hanover and moved to Durham in 1893. Affiliated colleges are located at Plymouth and Keene. There are several state vocational colleges and a technical institute; an educational television

system at the state university has been operated since 1959.

Colleges

New Hampshire also has many private educational institutions. Phillips Exeter Academy and St. Paul's School are the best known among the state's preparatory schools. Colleges include Dartmouth, founded at Hanover in 1769, St. Anselm (Manchester; 1889), and Colby-Sawyer (New London; 1837).

Cultural life. New Hampshire has several outstanding cultural institutions. The MacDowell Colony, a retreat for musicians and writers, was founded in 1907 and is a memorial to the composer and Peterborough resident Edward MacDowell (1860–1908). The 86-acre (35-hectare) former home of the noted sculptor Augustus Saint-Gaudens (1848–1907) at Cornish was designated a National Historic Site in 1964. The Currier Gallery of Art and an institute of arts and sciences are located in Manchester; Dartmouth College is a distinguished fine arts centre, and there is an arts and science centre in Nashua. Several towns have smaller galleries, art centres, and museums. Summer theatres flourish in a dozen or more resort areas.

Recreation and tourism are important parts of the cultural pattern of the state. New Hampshire has some 30 ski areas, many of which operate their lifts in the summer for sightseers. There are more than 200 youth camps and more than 70 golf courses. Historical sites throughout the state include restored colonial homes and Fort Constitution at New Castle. There are more than 30 state parks and more than 800,000 acres (324,000 hectares) of publicly owned forest preserves.

HISTORY

American Indian population. Before contact with the English about 3,000 American Indians inhabited what eventually became New Hampshire. They were organized into clans, semiautonomous bands, and larger tribal entities; the Pennacook, with their central village near the town that bears their name, were by far the most powerful of these tribes. The entire Indian population was part of the linguistically unified Algonquian culture that dominated northeastern North America. Disease, war, and migration quickly reduced the population after contact with English settlers. By 1700 few Indians resided within colonial boundaries. The primary contemporary reminder of American Indian inhabitation is in place-names like Lake Winnipesaukee, Kancamagus Highway, and Mount Passaconaway.

The English colony. The New Hampshire region was included in a series of grants made by the English crown to John Mason and others during the 1620s. A fishing and trading settlement was established in 1623, and in 1629 the name New Hampshire, after the English county of Hampshire, was applied to a grant for a region between the Merrimack and Piscataqua rivers. The towns of Dover, Portsmouth, Exeter, and Hampton were the main settlements.

Early colonial disputes

From 1641 to 1679 the region was administered by the colonial government of Massachusetts. Following territorial and religious disputes between Massachusetts and Mason's heirs, New Hampshire became a separate royal province in 1679. Bitter boundary feuds with Massachusetts and New York over that part of the New Hampshire grant that became Vermont continued almost until the American Revolution. Benning Wentworth held the post of colonial governor from 1741 to 1767, the longest tenure of any royal governor in any of the colonies.

In 1767 the colony took its first census and reported about 52,700 people. By 1772 the state was divided into five counties, to which five others have been added since 1800. New Hampshire soldiers played an active part in the colonial wars between Great Britain and France from 1689 to 1763. By the end of the colonial period the seat of government was at Portsmouth, and there were 147 chartered towns in the province.

Revolution and statehood. In December 1774 there was armed resistance to the British at New Castle, where Fort William and Mary (now Fort Constitution) was seized by colonists. The citizens of New Hampshire were overwhelmingly in sympathy with the aims of the revolutionary leaders. The state furnished two brigadier generals to the Continental Army, three regiments of regular troops, and hundreds of short-term militiamen. New Hampshire officially became a state in 1776 and issued its own Declaration of Independence several weeks before the national Declaration of July 4, 1776. New Hampshire's vote was the ninth and decisive vote in ratifying the Constitution of the United States in 1788.

Role in forming the nation

Following the establishment of the nation, the state grew rapidly. Agriculture, notably sheep raising, flourished, and manufacturing developed along the fast-flowing rivers, particularly in Manchester. When the railroads came to the Northeast, an extensive rail network was constructed in New Hampshire. Portsmouth and its surrounding towns were shipbuilding centres. In 1846 Manchester became the first incorporated city in the state. New Hampshire was also the birthplace of such noted statesmen as Daniel Webster and President Franklin Pierce.

Contributions to war and politics. New Hampshire played an active part in the American Civil War, and, following that conflict, the state voted regularly for Republican candidates until the New Deal of the 1930s. The Progressive movement had an impact upon the state in the early 20th century, and the economic depression of the 1930s left scars that were only slowly effaced. In World Wars I and II and in the Korean and Vietnamese conflicts, soldiers from New Hampshire played important roles, and the shipbuilding and naval facilities at Portsmouth were of special value.

Because it holds the earliest presidential primary in the nation, New Hampshire has furnished the first testing ground for many candidacies. In 1905 Portsmouth was the site of the peace conference that ended the Russo-Japanese War, and in 1944 the state was host to the Bretton Woods Conference that founded the International Bank for Reconstruction and Development and the International Monetary Fund. The U.S. Air Force constructed a huge Strategic Air Command base at Newington in 1957, and the Portsmouth Naval Shipyard, across the Piscatagua River from Portsmouth in nearby Kittery (Maine), has continued to specialize in building and repairing submarines. (J.D.S./J.R.D.)

Rhode Island

Rhode Island, one of the original 13 states and one of the six New England states, is often remembered for being the smallest state in the Union, only about 48 miles (77 kilometres) long and 37 miles (60 kilometres) wide. More remarkable, however, are the state's contributions to the forming of the new nation, of its concept of liberty of conscience—the legacy of Roger Williams, its founder—and of its influence on the industrial development of the United States.

Rhode Island is one of the most densely populated states in the nation. Providence, the founding settlement, is the capital and the cultural, political, and financial centre of the state. The extreme compactness of area, large population, and economic activity have tied Rhode Island closely to its neighbours, Connecticut on the west and Massachusetts on the north and east. The Rhode Island Sound on the south is the basis of the state's fishing industry.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The western two-thirds of the state is part of the New England Upland, with hills rising as high as 800 feet (240 metres) above sea level. The Narragansett, or Seaboard, Lowland comprises coastal lowlands and islands that are less than 200 feet (61 metres) in elevation. The highest point in the state is Jerimoth Hill, 812 feet (247 metres), near North Foster.

Drainage. Several river systems drain Rhode Island. The most important are the Blackstone, the Pawtuxet, and the Pawcatuck. The Blackstone River and its tributaries drain the northern part of the state. Originating in Massachusetts, the Blackstone once provided waterpower for the textile mills built at Woonsocket, Pawtucket, and a dozen villages in between. The Pawtuxet River drains the central part of the state. Its north branch was flooded

when Providence built a dam at the village of Kent. The resulting Scituate Reservoir is now the state's largest body of fresh water, supplying Providence and its neighbouring communities. The Pawcatuck River flows west across the southern part of the state into Block Island Sound south of Westerly.

Narragansett Bay

Mount Hope Bay feeds water from Massachusetts into Narragansett Bay, which has always been Rhode Island's greatest asset, providing a convenient waterway running two-thirds the length of the state. The commercial trade on which the wealth of Newport, Bristol, and Providence was founded provided some of the capital for the industrial development of the state. There is now a large ocean-borne bulk commerce centred on the bay.

Soils. The state's soils are primarily glacially derived. In the western area many rocks and boulders lie on rolling land with outcrops of granite and gneiss. Better soils with fewer rocks are found in the lowlands and adjacent bay islands, where the soil base is sedimentary. Along the west side of Narragansett Bay and in river valleys, glacial outwashes of sand and gravel are rapidly drained.

Climate. The state has a humid continental climate, with winds predominantly from the west. Marine influences are discernible in differences between coastal and inland locations. The average monthly temperature in January is 29° F (−2° C), in July 71° F (22° C), and for the year 50° F (10° C). Annual precipitation is about 42 inches (1,067 millimetres). The major weather characteristic is variability, with extreme weather conditions including tropical storms, ice storms, and heavy snow. By 1980, 29 storms of hurricane force had struck Rhode Island.

Plant and animal life. More than 60 percent of the state is forested with secondary tree growth. White pine grows in scattered locations. Several varieties of oak are abundant and, with other hardwoods, form the bulk of the timber harvest. Ash, hickory, and maple are widely dispersed, with some birch and hemlock in mixed woodlands. Swamp maple grows in wet places, while cedar, juniper, and poplar fill in abandoned fields and pastures. Favoured species of spruce, fir, and pine are grown on tree farms for commercial use.

Small animals such as rabbits, woodchucks, raccoons, gray squirrels, minks, and beavers are distributed widely outside urbanized areas. Foxes are less widely distributed, while white-tailed deer are found on Prudence and Block islands and in the western woodlands.

Settlement patterns. Rhode Island comprises five counties, eight cities, and 31 towns (townships). The major population centre is Providence and the nearby communities of Pawtucket, North Providence, Johnston, Cranston, Warwick, and East Providence. Warwick and Cranston are the state's second and third largest cities. Major clusters of population in Woonsocket, Newport, and Westerly are close to the state's borders.

Providence and other cities

The people. Before the arrival of the British colonists, about 7,000 Narraganset Indians lived along the west side of the bay and islands, where they fished and farmed in summer and moved into the forest in winter. They spoke a dialect of the Algonquian language family, as did their neighbours, the Wampanoag to the east and the Niantic and Nipmuc. The Narraganset were detribalized in 1880 when they sold their communal land to the state. The Narraganset continued their traditions, however, and maintained a group registry, and in 1934 they incorporated and reorganized their internal government to include an elected chief and a council. A suit to reclaim their land, based on a 1790 statute that prohibited sales of Indian land without federal approval, was settled out of court in 1978 for about 1,800 acres. A petition for tribal status was recognized in 1983, but the Narraganset's endeavour to gain reservation status was refused in 1985. Tribal programs and financial compensation continued, however.

During the colonial era, most settlers were English Protestants, although some Irish, Huguenots, Jews, and blacks also arrived then. Irish Catholics began to come in large numbers in the 1820s, and they settled in Providence, Pawtucket, and Newport. The largest groups of French-Canadians came during the American Civil War; they were followed by small numbers of Scandinavians, Germans, Portuguese, and Cape Verdians. New arrivals settled first in the river valley mill villages; later immigrants settled in the cities. At the turn of the century, immigrants included eastern Europeans—Poles, Russian Jews, Ukrainians, and Lithuanians—and Greeks, Armenians, and Syrian-Lebanese. This was also a period of significant Italian immigration.

After World War II, blacks came from the South or from New York City, and Hispanics came from Puerto Rico, the Dominican Republic, Colombia, and Mexico. They were followed during the 1970s and '80s by Southeast Asians.

The economy. Since the mid-20th century, high taxes, high energy costs, and low wages have combined to keep the Rhode Island economy from expanding. In 1985 the state passed legislation to reduce the tax burden and overhaul unemployment and worker's compensation programs. In the same year, the legislature created the Business Action Center, a problem-solving agency designed to help businesses with their concerns.

Resources. The major resources of Rhode Island are human: muscles, skills, intellectual creativity, and an increasing mastery of technology. In addition, the forests have been used for centuries for fuel and lumber. The soils, where cleared of rocks, offer good potential for producing food. Reservoirs and groundwater provide drinking water and are important industrial resources. Narragansett Bay and the Atlantic Ocean provide food, transportation, and recreational opportunities.

U.S. Navy installations

Narragansett Bay has also attracted the U.S. Navy. The Naval War College has operated at Newport since 1884; the naval presence grew through World War II until the closing of the Quonset Point Naval Air Station and the reassignment of ships in 1974. The navy, however, has continued as the second largest employer in Rhode Island, after state government.

Agriculture. Agriculture has decreased steadily in Rhode Island. The leading areas of agriculture are nurseries, lumbering, and harvesting turf. Dairy products and poultry are also important. Potatoes are the major field crop and apples the major fruit. In 1985 a bond issue was passed to buy up development rights and permit farmers to continue to farm; this is an endeavour to preserve rural areas and save a vanishing heritage.

Fishing. The most valuable shellfish landings are lobsters and quahogs (hard-shell clams). Finfishing takes place in Narragansett Bay, but more fish are caught offshore from Galilee, where a cooperative processing plant is located. Flounder, butterfish, whiting, scup, and cod are the main species. Despite larger trawlers and more technical equipment, yields have declined. Laws restrict fishing methods and the size and number of fish to be taken.

Industry. Relying in its early years almost entirely on subsistence agriculture and seaborne trade, Rhode Island became a pioneer manufacturing state, principally in textiles, after the American Revolution. Manufacturing concerns producing jewelry, silverware, electrical equipment, textiles, transportation equipment, and fabricated materials have given way in importance to the service industry (including tourism), wholesale and retail trade, government, finance, insurance, and real estate.

Transportation. Amtrak, Conrail, and the Providence and Worcester Railroad serve the state's rail needs. The major arteries for road travel are Interstate Highways 95 and 295 (the Providence bypass), which offer connections to cities in Connecticut and Massachusetts as well as between communities in Rhode Island. A limited-access connector road links I-95 to the state's main air terminal, T.F. Green State Airport, in Warwick. General aviation is also served by state airports at Smithfield, Newport, Westerly, Block Island, and Quonset.

The Rhode Island Port Authority and Economic Development Corporation operates port facilities for ocean-borne commerce at Providence and Davisville/Quonset Point. The port of Providence, at the head of Narragansett Bay, handles petroleum, automobiles, scrap iron, lumber, and steel. The Davisville/Quonset Point facility (formerly the Quonset Point Naval Air Station) is part of a large industrial park with rail and air connections.

Administration and social conditions. Atop the state capitol in Providence is the statue of "The Independent Man," a symbol of so much that is characteristic of Rhode Island. Roger Williams wrote that he had founded Providence as a place of refuge for "those distressed for cause of conscience," and the principle of absolute religious freedom has been an abiding article of Rhode Island's political philosophy.

Principle of religious freedom

Government. Rhode Islanders count their colonial charter (1663) as the first constitution. A new constitution, approved in 1986, replaced the 1842 constitution, which had been amended more than 50 times.

Ever since its founding, the state has shown a reluctance to permit elected officials to exercise extensive powers. In the early years legislatures and elected officials served for only six months. Prior to 1854 the state had five capitals—Providence, Newport, East Greenwich, Bristol, and South Kingstown—and the General Assembly traveled from one to another. In 1900 Providence was chosen as the sole capital.

Elected officials are the governor, lieutenant governor, secretary of state, attorney general, and treasurer, each serving unlimited two-year terms. The governor has veto power and the power to name certain department heads. The 150 members of the General Assembly (50 senators, 100 representatives) are paid by the day for 60 legislative days. Attempts to amend the state constitution to increase legislators' pay have been defeated by the voters.

The state has a district court system; superior courts; a Supreme Court, which also can give advisory opinions if requested; and municipal and probate courts, the latter often identical with the town council in the smaller communities. There is also a Family Court, which handles both juvenile and domestic cases. Supreme court judges are elected by both houses of the General Assembly sitting as a committee of the whole, and since the 1930s they have had tenure.

Most of the cities in Rhode Island operate with a mayor and city council form of government, but East Providence and Newport have city managers, with mayors chosen from among the councilmen to act as the ceremonial heads of the local government. Most of the towns are governed by a town council, but in some cases operations are conducted by a town manager. Some call an annual financial town meeting.

Rhode Island has a sales tax, which exempts food and prescription drugs. In 1971, after many defeats in the General Assembly, the state adopted a personal income tax. Lotteries were used widely in colonial times and after the Revolution to raise capital for civic improvements; they were prohibited later but were reestablished in 1974. A share of the state's income derives from its tax on pari-mutuel betting at the Lincoln Greyhound Track and Newport Jai Alai.

There are jails in cities and towns throughout the state, but most persons awaiting trial or sentenced to prison are sent to the state Adult Correctional Institutions (ACI) at Cranston. In 1975 the ACI adopted a work-release program to rehabilitate prisoners. Rhode Island has a strong state police force that operates throughout the state.

Education. In 1969 a Board of Regents was created and charged with responsibility over all public education, from elementary schools through the state-operated colleges and the university. In 1981 the board was divided into the Board of Regents for Elementary and Secondary Education and the Board of Governors for Higher Education. A number of private preparatory schools, both sectarian and nonsectarian, send graduates to many of the major colleges and universities, especially in the East.

Higher education

Rhode Island is strong in its institutions of higher education. Brown University in Providence, founded in 1764 as Rhode Island College, is one of the major Eastern universities that make up the so-called Ivy League. It is noted for its library facilities, especially the John Carter Brown Library of early Americana. The Rhode Island School of Design, in Providence (founded 1877), is widely known, primarily for its training in the visual and graphic arts. The University of Rhode Island, in Kingston, is a land-grant institution dating from 1888. Roman Catholic colleges include Providence College (1917) and Salve Regina-The Newport College (1934) in Newport. Rhode Island College, a four-year public liberal arts college, dates from 1854. Other institutions include Bryant College (1863), Johnson and Wales University (1914), the New England Institute of Technology (1940), Roger Williams College (1948), and the Community College of Rhode Island (1960).

Health and welfare. Local communities are handicapped by limitations on property taxes as a major source of revenue. The state's capacity to raise taxes, however, has permitted it to play a significant role in establishing acceptable standards of support for municipal services. State funds subsidize teachers' pay, provide grants for school construction, protect public health, support the needy, aid dependents, and offer incentives for achieving specified goals in these areas.

The state maintains elaborate facilities to care for the sick and indigent of all ages. Rhode Island has a national reputation for deinstitutionalization of clients into group homes. A system of outpatient care and of referral for mental health patients is available.

Biomedical research has been important since the establishment in 1975 of the Brown University Program in Medicine, which operates a full biological science program. More than 350 laboratories support university research at hospital centres. The University of Rhode Island has a large biological sciences faculty and conducts research in pharmacology.

Cultural life. Library facilities are plentiful throughout the state. The Redwood Library, in Newport, and the Providence Athenaeum, both proprietary libraries housed in architecturally important buildings, have roots going back to the mid-18th century. The public libraries of Providence and Westerly have important holdings, the former having special collections on whaling, printing, slavery, and Irish literature. The library of the Rhode Island Historical Society, in Providence, has more than 1,000,000 manuscripts and is especially strong in its holdings of the state's genealogical records. The society operates John Brown House, a merchant's mansion in Providence; built in 1786, the house is furnished with masterpieces of the Newport school of cabinetmakers and with other 18th-century antiques. The Museum of Rhode Island History, Aldrich House, also run by the society, is a converted mansion that features changing exhibits.

The Museum of Art of the Rhode Island School of Design has impressive collections, including early Rhode Island furniture and silver. Roger Williams National Memorial (Providence), Slater Mill Historic Site (Pawtucket), and South County Museum (Narragansett) offer specialized insights into other aspects of the state's past.

Preservation of early homes

Preservation societies in both Providence (1956) and Newport (1945) restore and preserve surviving colonial homes, while the state Historical Preservation Commission is a major centre for research and publications. There are more than 11,500 Rhode Island properties listed on the National Historic Register. The Preservation Society of Newport County operates as museums several mansions that were formerly the summer homes of millionaires. The Newport Historical Society museum, with its fine collections; Touro Synagogue, a magnificent example of colonial architecture; Old Colony House; Redwood Library; Hunter House; the restored colonial homes of the Point section; and the International Lawn Tennis Hall of Fame and Tennis Museum in the Newport Casino building give Newport an extraordinary and varied cultural heritage.

The Rhode Island Philharmonic Orchestra, the Rhode Island Civic Chorale and Orchestra, the Community Chorus of Westerly, and numerous smaller groups and organizations are among the state's musical resources. Newport has an annual chamber music series, as well as several series featuring renowned soloists. Many of the restored houses in Newport are the setting for these performances, and jazz, modern, and folk festivals are held at Fort Adams. The State Ballet of Rhode Island performs throughout the state. The Trinity Square Repertory Company (1964), with its own home in Providence, is renowned for producing works by new playwrights, as well as for staging novel productions of classic works.

Many recreational activities are centred on the state's water. For many years the waters off Newport have been the site of the yacht races for the America's Cup. An annual tuna tournament is held in Rhode Island and Block Island sounds. Newport Casino, one of the early centres of tournament tennis, has an annual grass-court tournament of national importance.

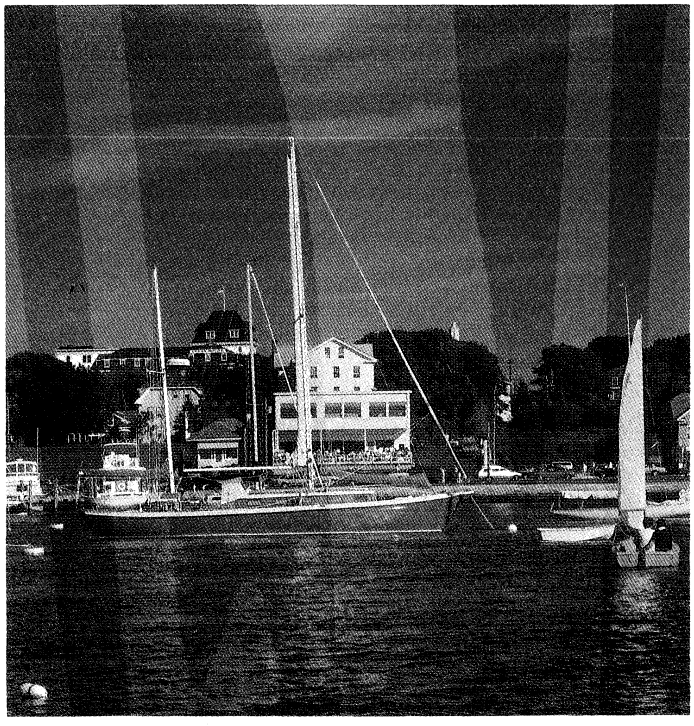
HISTORY

Colonial period. The name Rhode Island owes its origin to the 16th-century Italian explorer Giovanni da Verrazzano, who compared the size of Block Island to the Mediterranean island of Rhodes. The earliest settlers thought that Verrazzano had referred to the island which the Indians called Aquidneck and thus began calling it Rhode Island.

In the state's official name—The State of Rhode Island and Providence Plantations—lies a clue to its founding. The first settlement was made by the minister Roger Williams and a few followers at Providence, near the head of Narragansett Bay, in 1636. They were either under edict of banishment from Massachusetts Bay colony—Williams for advocating freedom of conscience in religion—or were in trouble with the authorities there. In 1638 a group of Bostonians, in similar difficulties, purchased the island of Aquidneck, now Rhode Island, from Indians and settled Portsmouth. Factional strife split this settlement, and William Coddington and his adherents moved to the southern end of the island, where they founded Newport, leaving Anne Hutchinson and her followers in Portsmouth. In 1643 Samuel Gorton and a dissident group settled Warwick.

Williams went to England in 1643 and returned the following year with a royal patent for the colony, but the four towns could not agree on a form of government until 1647, when a loose confederacy was established. The colony was never accepted into membership in the United Colonies of New England—comprising Plymouth, Massachusetts, Connecticut, and New Haven—and it was constantly threatened with a takeover by these governments. Coddington, having himself made ruler for life of the island towns, split the colony between the mainland and the island towns. Williams and John Clarke, the latter representing island elements unhappy about Coddington's commission, sailed for England in 1651, succeeded in getting the commission rescinded, and in 1654 set up a reunited government. Clarke remained in England and, in 1663, won a royal charter that was to be the basis of colonial and state government for 180 years.

Ulrike Welsch—Photo Researchers



Sailboats at Watch Hill, R.I.

Although the colony never officially joined the New England colonies in King Philip's War (1675–76), it suffered greatly. All mainland settlements were burned, including, in 1676, many houses in Providence. Most of the mainland settlers took refuge on Rhode Island, which was not attacked. The Great Swamp Fight, which broke the power of the Narraganset Indians, took place in December 1675 west of the present village of Kingston.

Early commerce with the West Indies

Rhode Island had commerce with the West Indies, selling horses, barrel staves, and salt fish. Eventually, some merchants plied the triangular trade: taking rum to the African coast, where it was traded for slaves; carrying the slaves to Charleston, S.C., or to the West Indies, where they were traded for molasses; and carrying the molasses to Rhode Island, where it was distilled into rum. The passage of the Sugar Act by Parliament in 1764 seriously restricted this trade, and the colony, never enthusiastic about obeying unpopular laws, began to indulge in smuggling of sugar and molasses. In 1772 the British customs vessel *Gaspee,* patrolling Narragansett Bay, ran aground off Namquit (now Gaspee) Point while pursuing a suspected smuggler; that night it was burned by a group of townsmen from Providence. This has been widely regarded as the first act of outright violence against the British crown in the period leading to the American Revolution.

Revolution and independence. During the war Newport was occupied by the British. In 1778 a land force under General John Sullivan and the French fleet commanded by the Count d'Estaing cooperated in an operation designed to dislodge them. Before the French troops could be landed, however, a British fleet appeared in the bay; d'Estaing halted the landing and set out in pursuit. Two days later, before the ships had actually engaged, they were dispersed by a storm. The American ground forces, lacking French assistance, were forced to retreat from the island. At Butts Hill they fought a strong rearguard action that became known as the Battle of Rhode Island and in which a battalion of freed slaves distinguished itself. A Rhode Islander, General Nathanael Greene, distinguished himself as Washington's second-in-command.

After the war Rhode Island was reluctant to ratify the Constitution until the Bill of Rights was proposed in the form of 10 amendments. The state's largely agricultural population was opposed to joining the Union, while the merchants of Providence and Newport worked hard for ratification. When threats of commercial isolation from the other states were raised, Rhode Island accepted the document in May 1790 by a margin of two votes.

Newport, preeminent before the war, lost much of its economic power during the British occupation, and Providence, led by such merchants as the four Brown brothers, John, Joseph, Nicholas, and Moses, assumed the leadership.

The Dorr Rebellion

In 1842 a movement for widening the franchise, limited under the 1663 charter to freeholders and their eldest sons, resulted in a conflict known as the Dorr Rebellion. Led by Thomas Wilson Dorr, the son of an aristocratic family, the faction favouring universal suffrage held a convention in 1841 and adopted a constitution embodying this principle. At an election held under this constitution, Dorr was elected governor in 1842, but the election was not accepted as legal by the legislature or the state Supreme Court. When his forces were repulsed in an attempt to seize the arsenal in Providence, Dorr fled the state. Upon his return, he was tried on a charge of high treason, convicted, and sentenced to life imprisonment; he served only one year, however, and was released in 1845. By that time the state had adopted a revised constitution considerably broadening the basis of the franchise, but it was not until the mid-20th century that full rights to vote in all elections were extended to all citizens at the age of 21 (later 18).

In the years after the American Civil War the Republican Party, led by such political bosses as General Charles R. Brayton, dominated the state, mainly because the cities, which were increasingly Democratic, were not proportionately represented in either chamber of the General As-

sembly. Theodore Francis Green, a Democrat, was elected governor in 1932 and reelected in 1934. With the Democratic Party in power, a different prolabour domination followed. Of the 14 governors elected between 1933 and 1985, 10 were Democrats and four were Republicans. On the national level, Republicans Dwight D. Eisenhower, Richard M. Nixon, and Ronald Reagan carried the state.

A number of Rhode Island's concerns have become bipartisan in nature: improvements in the quality of drinking water and of the state's rivers and the bay, improved education, and the development of a stronger economy.

A new constitution was approved in 1986 that permits constitutional amendments after a majority vote in both houses of the General Assembly and a majority of those voting in a general election. Further, it provides a mechanism for voters to convene a constitutional convention on a regular basis at least once every 12 years.

(B.F.S./M.I.Wr.)

Vermont

Admitted as the 14th state on March 4, 1791, Vermont is one of the six New England states in the northeastern corner of the nation. Its 9,614 square miles (24,900 square kilometres) have relatively few inhabitants, and its capital, Montpelier, is one of the least-populous U.S. state capitals. On the south, Vermont borders Massachusetts; on the west, New York; and on the north, Quebec, Can. From the Canadian to the Massachusetts border, the Connecticut River separates Vermont from New Hampshire on the east. The river, from the mean low-water line on the western bank, is entirely within New Hampshire.

In many ways Vermont is a vigorous survivor of an earlier, simpler time in the United States. Millions of people visit the state each year, and many thousands of out-of-state residents maintain second homes in Vermont. These people primarily seek the beauty and tranquillity of Vermont's mountains and narrow valleys and the sense of the nation's past that pervades the entire state. The steeples of white wooden churches rising above small, mountain-bound towns with trim village greens; the herds of dairy cattle on sloping mountain pastures; and the red-gold leaves of tree-lined autumnal lanes are aspects of scenic Vermont that, in painting and photography, have become symbols of the rural United States.

George A. Robinson—f/STOP Pictures

Cross-country skier near Cambridge, Vt., with Mount Mansfield in the background.

Many people left their birthplaces in Vermont to join the movement westward and to America's cities. In turn, many creative personalities have sought the spiritual refuge offered by the state. Vermont has never stood in the mainstream of the nation's history, but its people and land have poured into their country a strength and a sense of continuity that joins the achievements of the nation's past with the purposes of its present.

PHYSICAL AND HUMAN GEOGRAPHY

The land. The land of Vermont does not have great variety, but in place of this it substitutes an intensity and pervasiveness of those features it does possess.

The Green Mountains

Relief. The Green Mountains that cover most of the state are a northeastward extension of the Appalachian Mountains that run from Canada into northern Alabama. They provide Vermont with a north–south backbone that ranges from 20 to 36 miles (32 to 58 kilometres) in width. Thirty-one mountains in the state rise to more than 3,500 feet (1,100 metres), and most of this tilted landscape is rocky with thin topsoil. Only 15 percent of the state's terrain, mostly in the Champlain Valley, is level land with fertile soil and high productive capability. Vermont's average altitude is about 1,000 feet above sea level. Mount Mansfield, at 4,393 feet (1,339 metres), is its highest point; and Lake Champlain, at 95 feet (29 metres), is its lowest. On the Vermont–Massachusetts border, the northern end of the Hoosac Range enters the state, and the Taconic Range rises along the southwestern side. North of the Taconic Range are the Red Sandrock Hills, which extend along Lake Champlain to St. Albans.

Drainage. These ranges are broken by only a few river valleys, such as the Winooski, Lamoille, and Missisquoi, all flowing westward into Lake Champlain. Part of the Missisquoi turns north through Canada before returning to Vermont. Lake Champlain's waters empty northward into the Richelieu River and flow 80 miles into the St. Lawrence. The longest river entirely within the state is the Lamoille (84 miles), followed by Otter Creek (75 miles), which rises in southwestern Vermont and flows northward into Lake Champlain. Several small streams, the largest of which is the White River, flow from the central highlands into the Connecticut River. The western portion of Lake Champlain is in New York; and 75 percent of Lake Memphremagog, the second largest lake associated with Vermont, lies in Canada. The largest of the 400 natural lakes entirely in Vermont is Lake Bomoseen (2,395 acres).

Rainfall and temperature

Climate. Snowfall in Vermont usually averages between 70 and 80 inches (1,800 and 2,000 millimetres) in the valleys and up to 110 inches in the mountains. Total annual precipitation varies from 34 inches in the eastern and western sections to more than 40 inches in the mountains. Winter temperatures can drop to −34° F (−37° C) and lower, but in the summer they rarely rise above 90° F (32° C). Pleasant summer days often turn cool after nightfall. The annual growing season is only about 120 days—somewhat longer in the low-lying Champlain Valley—because frost usually comes in September and may strike as late as the beginning of June. The short growing season and rocky soil make dairying the dominant form of commercial farming. About 80 percent of farm income comes from dairy products. The state has more cows per capita than any other state except Wisconsin.

Plant and animal life. A century ago many of Vermont's hilltops had been cleared as pastures and open fields. As farmers abandoned the hillsides, the open spaces quickly filled with trees. Pine, spruce, fir, and hemlock are common; maple and birch are among the deciduous trees. The state tree is the sugar maple, which reflects Vermont's prominence in maple sugar and syrup production. The wooded areas, with their small brooks and springs, produce an amazing variety of ferns and wildflowers; in the spring and summer they are filled with the many species of birds common to the Northeast.

Vermont has a huge deer population, and deer hunting is an autumn ritual. Bears are often seen, but wild members of the cat family are rare. Small animals abound in Vermont, and fishing in the lakes and streams, including ice fishing in winter, is popular.

Cities and towns

Settlement patterns. Most Vermonters live in valley cities and towns. Burlington, in the Champlain Valley, is Vermont's largest community, followed by Rutland, in the Otter valley; Bennington, in the Walloomsac valley; and Essex, near Burlington. The Green Mountains were long a barrier between eastern and western Vermont, and judgeships and political candidates often were chosen to balance an eastern and western sectionalism. Although regional division is now a minor factor, some observers detect it emerging between southern and northern Vermont, presumably a reflection of the south's influx of newcomers and resort developments. Others sense a dichotomy involving small towns and large towns, which revolves around such public issues as state constitutional reforms, welfare aid, and educational innovations. Three isolated northeastern counties have been known since 1949 as the "Northeast Kingdom." Pragmatically, however, the major sense of regionalism is derived from large towns, which form a centre for surrounding rural areas.

The people. Because so many Vermonters are descended from early Americans of Protestant background and English heritage, the people of the state are almost prototypical Yankees. There is scarcely a town in Vermont that does not have a white frame church on its village green or main street. Virtually every Protestant sect is represented in Vermont, with a heavy concentration of Presbyterians in the Caledonia county area of northeastern Vermont. The name Caledonia, the Roman designation for northern Britain, was brought by the Scottish immigrants who first settled the region in the 1770s.

In 1848, when railroads were first built in Vermont, a large number of Irish immigrants were hired as labourers. Many of their descendants live today in Rutland, Burlington, St. Albans, and other large towns. During the 1900s French Canadians from Quebec province settled in the state, many of them in the woolen-mill town of Winooski and others on farms along the northern border. Almost 10 percent of all Vermont residents speak French as their first language. Immigrants from northern Italy carried with them centuries of quarrying and stone-carving tradition to Barre and other granite-producing areas. They have given Barre a character quite different from what the visitor expects to find in a Vermont city. Other quarry workers from northern Spain settled in the Barre–Montpelier area. Many Welshmen worked in the slate mines of western Vermont because they were familiar with this type of mining in their native land. Immigrants from Poland sought work in Brattleboro, Springfield, and other manufacturing towns. The slight need for industrial labour, and the rural character of the state, have attracted few blacks from the nation's South, and they number less than one-half of 1 percent of the population. The Roman Catholic diocese of Burlington includes all of Vermont, and Roman Catholics make up about one-third of Vermont's total population.

Famous emigrants

Vermonters continue their long-standing pattern of leaving the state in order to find employment elsewhere. This tiny state has produced many leaders of American business and political life. John Deere, for example, designed the plow that cultivated the fields of the Midwest; Frederick Billings built the Northern Pacific Railway; Joseph Smith and Brigham Young led the Mormon church; and Gurdon Saltonstall Hubbard was the first meat-packer in Chicago. Presidents Chester A. Arthur and Calvin Coolidge were born, respectively, in Fairfield and Plymouth, and both parents of President Rutherford B. Hayes were Vermonters who migrated to Ohio. The Taft family, famous in U.S. statesmanship, traces its Ohio origins to Vermonter Alphonso Taft, who left Townshend and settled in Cincinnati.

The economy. Farming has declined as family farms have been combined into larger units, and an increasing number of farmers, unable to modernize their equipment and expand their herds, have sold their lands. Labour is not generally well organized in the state.

Industry. Vermont's approximately 1,000 manufacturing plants reflect national industrial trends. Machine-tool plants in Springfield tend to expand and retract along with the national economy. The St. Albans area in Franklin county especially has suffered from the decline of the railroad industry, but the state as a whole has one of the highest employment percentages in the nation. Textile mills were once major employers in many cities, such as Winooski, but many of these have closed or moved to the South. Some computer industries have moved into Vermont. Other Vermont firms have become subsidiaries of national firms. The state's Agency of Development and Community Affairs searches constantly for new industry, while an industrial park authority offers site guidelines.

Many Vermont industries are small companies that provide specialized products. Wood and paper products are natural for a state so heavily timbered, and about one-third of Vermont's manufacturing plants make bowls, hockey sticks, furniture, and paper of different kinds. Printing is among the major industries.

Mining. Extractive industries contribute significantly to Vermont's economy. Quarries in Barre are among the largest granite pits in the world; and marble from Proctor is used for constructing commercial and public buildings—examples include the U.S. Supreme Court building in Washington, D.C., and the United Nations headquarters in New York City. Slate quarries operate along the Vermont–New York border.

Attractions for visitors

Tourism. Vacation resorts, motels and hotels, and related services employ thousands of Vermonters to serve the many tourists that visit the state each year. Skiing facilities at Stowe, Dover, Sherburne, and the Mad River valley are among the many Vermont winter resorts that attract people from throughout the Northeast. During the summer, visitors hike mountain trails, search through antique shops, study exhibits in the many museums, attend musical and dramatic performances, or bicycle or drive through the state and take photographs. The travel information office of the Agency of Development and Community Affairs is active in promoting tourism, and its *Vermont Life* magazine depicts the state's scenic resources.

Transportation. Passenger service ceased on Vermont's railroads in 1966 but was resumed on a limited basis by Amtrak in 1971. Airline service is limited in southern and central Vermont, but the airport at Burlington is one of the busiest in New England. The major transportation arteries are north–south routes in the Connecticut River valley and the lowland valley south of Lake Champlain. Interstate highways link Vermont with Massachusetts, New Hampshire, and Quebec province. Elsewhere, roads are often winding, narrow, and hilly, following the contours of the land. None of these factors is conducive to industrial expansion, and transportation remains a major problem.

Administration and social conditions. *Government.* On July 8, 1777, Vermont adopted a constitution, the first in the United States to prohibit slavery. It was revised in 1786, and in 1793 the present constitution was adopted. The amending process is awkward, and amendments may be considered only once every four years. Voters are required to take a "freeman's oath," a holdover from colonial America, and to be "of a quiet and peaceable behavior."

The governor of Vermont is elected for a two-year term and may run for reelection. Members of the General Assembly also serve two-year terms; the House of Representatives has 150 members, and the Senate 30. Vermont has only one member in the U.S. House of Representatives.

The judicial system in Vermont is headed by the Supreme Court, which has five members. Below it are 10 Superior Court judges and 15 district judges. These judges are appointed by the governor and confirmed by the state Senate to serve six-year terms. All other judges, including two assistant judges in each county, are elected by the people.

Every year on the first Tuesday in March, voters throughout the state meet in their town halls or community buildings to debate the town budget, road maintenance, and other topics and to elect local officials for the coming year. Because most communities are small and the state is compact, Vermonters have fairly direct access to elected officials, as well as ample opportunities to voice their opinions publicly on dominant issues.

Republican tradition

Vermont has traditionally been one of the most Republican states in the Union, but since the 1960s a balance between the two major U.S. parties has developed in state politics.

Education. In view of its limited tax base, Vermont supports its schools commendably. Locally elected boards, encouraged by the state, govern all primary and secondary education. The state provides for much of the funding, for teacher training and certification, and for various special programs in such areas as vocational training, arts and crafts, and rehabilitation.

The state system of higher education includes the University of Vermont (chartered in 1791) in Burlington and a number of liberal arts, technical, and other colleges. In addition, there are several outstanding private colleges and a law school in South Royalton. The school of languages and the writers' workshops at Middlebury College (1800) in Middlebury and Bread Loaf in Ripton are internationally known. Bennington College (1925), primarily a women's college, is known for its fine arts programs and for the major figures in the arts who regularly serve on its faculty. Marlboro College (1946) is noted for its work in the performing arts and for its summer music festivals. Norwich University (1819) in Northfield is the nation's oldest private military college.

Health and welfare. There is a state mental hospital at Waterbury; the state also maintains facilities for the mentally retarded. Vermont's correctional system includes five regional detention centres and facilities for juvenile offenders. In 1969 the state assumed responsibility for administering welfare programs, which formerly had been supervised by an "overseer of the poor" in each town.

Cultural life. Vermont enjoys a vigorous cultural life despite its rural and small-town character. Many artists and scholars have followed such famous literary figures as Sinclair Lewis, Pearl Buck, Robert Frost, and Robert Penn Warren in maintaining vacation homes in the state. Painters find inspiration in the landscape, and sculptors adapt old materials and forms from barns and antique shops to contemporary uses. Among the galleries in Vermont are the Robert Hull Fleming Museum at the University of Vermont, the Southern Vermont Art Center in Manchester, the Springfield Art and Historical Society in Springfield, the Chaffee Art Gallery in Rutland, the Wood Art Gallery in Montpelier, and the Johnson Art Gallery at Middlebury College.

Folk arts

Practitioners of various folk arts are numerous, and the state operates an arts-and-crafts service augmented by the Vermont Museum and Art Gallery Alliance. The Vermont Council on the Arts and the Vermont Historical Society often sponsor demonstrations by artists and craftsmen. Vermont has a rich heritage of folk music. In warm weather Vermont is vibrant with music and drama, including the annual Marlboro Music Festival and the Mozart Festival in Burlington. The Craftsbury Chamber Players attract large audiences to Hardwick. Summer theatre is popular in Stowe, Winooski Park, Weston, Dorset, and elsewhere in the state. The Vermont State Symphony Orchestra was the first in the nation to receive a legislative appropriation. The University of Vermont sponsors cultural events in Burlington and other communities, and lectures, films, and concerts are offered frequently by Vermont's other colleges and universities.

Vermont is proud of the way it preserves its heritage. The Shelburne Museum is called "The Museum of the American Spirit" because its historic buildings on 45 acres (18 hectares) contain a wealth of early artifacts. The Bennington Museum contains the oldest preserved Stars and Stripes carried in battle, a collection of the primitive-style paintings of Grandma Moses, and specimens produced by the large Bennington pottery industry. In Montpelier the Vermont Historical Society has created a museum inside a reconstructed Victorian landmark on the statehouse green. The Fairbanks Museum and Planetarium in St. Johnsbury is famous for its natural history displays. The Billings Farm Museum in Woodstock and the Ethan Allen Homestead in Burlington are popular attractions. Vermont has nearly 150 local historical societies and statewide groups, such as the Vermont Archaeological Society and the Lake Champlain Maritime Society.

Vermont maintains more than 100 covered bridges, most of which were constructed before 1912 and are protected by state law. A state division for historic preservation maintains almost 70 historic sites, including the Bennington Battle Monument, the Old Constitution House at Windsor, and the birthplace and family homestead of Calvin Coolidge at Plymouth. The State House in Montpelier is considered the state's most important historic site.

HISTORY

Exploration and settlement. The first inhabitants of Vermont were Indians of the Abnaki tribe who hunted wild animals and traveled on the waterways and footpaths of the region. In 1609 the French explorer Samuel de Champlain discovered a lake there to which he gave his name. The first permanent European settlement was established by the French in 1666 on Isle La Motte, an island in northern Lake Champlain. The name Vermont, derived from the French *vert* and *mont* ("green mountains"), was applied to the region because of the thick coniferous growth that kept its mountains green year-round. During the late 17th and early 18th centuries, Vermont served as a route for French and Indian incursions from Canada into Massachusetts.

French exploration and settlement

In 1724 the Dutch established a community in Pownal, and the first English-speaking settlers erected Fort Dummer on the Connecticut River near present-day Brattleboro. When the British won the French and Indian War (1754–63), the land was opened to settlement. At the time the American Revolution began in 1775, about 20,000 people were living in Vermont. Many Vermont towns bear the names of the Connecticut and Massachusetts towns from which the early settlers came.

Revolution and statehood. Although the region was explored long before the landing of the Pilgrims and was settled before the American Revolution, it began its early development not as a chartered royal colony but as a territory, the possession of which was disputed by New Hampshire and New York. In the decades before the Revolution, disputes—frequently armed conflicts—arose when land grants by New Hampshire conflicted with similar grants issued by New York. Between 1770 and 1775, many early settlers joined units of the Green Mountain Boys, led by Ethan Allen, and repulsed the Yorkers who tried to control Vermont. Later, when the American Revolution began, the same Green Mountain Boys also asserted their independence from England. Their successful assault on Fort Ticonderoga, on the New York side of Lake Champlain in May 1775, has been called the first offensive action of the Revolution. In 1777 the Vermonters created an independent republic—first as New Connecticut, then as Vermont—and they remained independent until they joined the Union on March 4, 1791.

Growth in the early 19th century

Vermont grew from about 85,500 inhabitants in 1790 to about 218,000 in 1810; but, by the 1830s, Vermonters were departing in large numbers for the expanding cities and the more fertile lands to the south and west. The opening of the Erie Canal from Albany, N.Y., to Lake Erie in 1825 and later improvements in transportation hastened this emigration. More than 35,000 Vermonters fought in the American Civil War. Vermont became the site of the only Civil War action north of Pennsylvania when a band of Confederates raided St. Albans in October 1864. Many veterans left Vermont after the war because the state provided inadequate natural resources to power industrialization and urbanization and because greater opportunities for individual advancement were available elsewhere. The rural character of the state was thus assured, and today Vermont is the most rural state in the United States, with two-thirds of its people living in towns of 2,500 residents or fewer.

Modern period. After the Civil War, dairying began to replace sheep farming, and the state's hill farms were gradually abandoned. The election of 1856 gave control of Vermont to the new antislavery Republican Party, a dominance that lasted until the 1960s. In the 1930s the first ski runs were built, which led, by the 1960s, to the development of a winter tourist industry. During this time many small manufacturing plants set up operations in the state, and much of the hill country was opened to vacation home development. Despite these changes, Vermont has retained much of its earlier character. (C.T.Mo./D.G.Sa.)

MIDDLE ATLANTIC REGION

Delaware

The first of the original 13 U.S. states to ratify the federal Constitution, Delaware occupies a small niche in the Boston–Washington, D.C., urban corridor along the Middle Atlantic seaboard. With 2,045 square miles (5,295 square kilometres), it is the second smallest state in the nation and one of the most densely populated. Its population, like its industry, is concentrated in the north, around Wilmington, where the major coastal highways and railways pass through from Pennsylvania and New Jersey on the north and east into Maryland on the south and west. The rest of the state comprises the northeastern corner of the Delmarva Peninsula, which it shares with Maryland and Virginia. Dover is the capital.

Historically, geographically, and economically, Delaware has its closest ties with Pennsylvania, particularly Philadelphia, where the Delaware River and other transportation arteries direct its commerce. The state's three counties—New Castle, Kent, and Sussex—had been established by 1680. Stability and conservatism have become characteristics of Delaware, especially in the southern areas, which until 1964 maintained a grip on political life vastly out of proportion to their population. As a result, old institutions have been tenaciously preserved.

The manufacturing complex in the north makes Delaware one of the most industrialized states, especially notable for chemical research. The state is often depicted as being dominated by corporations, especially by the vast du Pont industrial empire, but until the mid-1960s the industrial wealth of the Wilmington area was balanced by the political overrepresentation of the agricultural downstate region. All factions have united to perpetuate liberal incorporation laws that encourage many American businesses to make Delaware their nominal home state.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* With the exception of Florida, Delaware, located mainly within the Atlantic Coastal Plain, is the lowest-lying state in the nation. A long sand beach forms the state's oceanfront, stretching from the Maryland line, at Fenwick Island, to Cape Henlopen, at the mouth of Delaware Bay. Only one major break, Indian River Inlet, occurs along the 23-mile (37-kilometre) length of the beach. Much of the beach is a low bar between the ocean and a series of lagoons or shallow bays, but at Bethany Beach, near the southern boundary, and again at Rehoboth, near the northern end of the beach, the mainland reaches directly to the ocean.

Much of the shoreline of Delaware Bay is marshy. The mouths of tributary streams like the Murderkill, the Mispillion, and the St. Jones are so shallow that only fishing boats find safe harbours north of Lewes. Farther north, on the banks of the Delaware River, spots of high, dry land appear, as at Port Penn, New Castle, and Edge Moor; the state's main port, Wilmington, lies on the Christina River, a tributary of the Delaware.

Most of Delaware is drained by streams that run eastward to the Delaware River, Delaware Bay, and the Atlantic Ocean, but the Nanticoke River and its tributaries in southwestern Delaware flow into Chesapeake Bay. So does the Pocomoke River, which drains the Cypress Swamp, or so-called Burnt Swamp, in the extreme south of Delaware, athwart the Maryland line.

Most of the Coastal Plain is fertile and level, seldom more than 60 feet (18 metres) above sea level, but it becomes increasingly sandy to the south. Abundant woodlands, streams, and freshwater ponds interrupt the monotony of the landscape. Occasionally, as at Odessa, villages appear suddenly at the side of the road, with no more warning than the sight of a church steeple.

Near its northern edge the plain is intersected by the Chesapeake and Delaware Canal, which has been deepened and straightened for ocean shipping. It shortens the water route between Philadelphia and Baltimore by several hundred miles and also brings Baltimore closer to the ocean than via Chesapeake Bay. The canal is popularly considered to be the boundary between agricultural downstate Delaware and the northern industrial region. Though the land on either side of it is similar, many Delawareans are convinced that even the weather changes at the canal.

Chesapeake and Delaware Canal

Several high bridges over the canal, the giant twin bridges crossing the Delaware River north of New Castle, and the refinery stacks at Delaware City are the major landmarks on the horizon below the northwestern corner of the state, where the rolling hills of the Piedmont extend south from Pennsylvania. Within this area—less than 1/15 of the state's total area—and at its edge dwells most of Delaware's population. The highest point in the state, Ebright Road in New Castle county, is only 442 feet (135 metres) above sea level. Peculiar features are Iron and Chestnut hills, which protrude into the plain southwest of Newark and are scarred by open pits where iron ore once was mined.

The centre of Wilmington lies on hills sloping downward toward the confluence of the Christina and its major tributary, the Brandywine. There, navigable water brought shipping close to falls that provided power for manufacturing. The railroads and highways, which followed this fall line along the east coast, have kept Wilmington on major transportation routes between Philadelphia and Baltimore and have promoted the tendency for the urbanization of open land between Wilmington and other cities.

Climate. The climate of Delaware is humid and temperate. The average daily temperature at the New Castle County Airport in northern Delaware is 54° F (12° C), varying from an average high of 86° F (30° C) in July to an average low of 23° F (−5° C) in January. Temperatures in southern Delaware usually run about two degrees higher than these figures. August, which has the second hottest temperatures after July, is also the rainiest month, with an average precipitation of about 5.5 inches (140 millimetres), whereas February has the least precipitation, an average of almost three inches. The annual average precipitation is nearly 45 inches (1,140 millimetres).

Plant and animal life. Delaware is a transition zone between plants typical of Pennsylvania and New York and

Kevin Fleming

Brandywine Creek flowing through the Piedmont in northern Delaware.

those common to coastal Maryland and Virginia. Hardwoods are characteristic in the north, but pines become mixed with hardwoods in the south. Deer, foxes, raccoons, opossums, and muskrats are common. Beaches and marshes are a winter refuge for many wildfowl, as well as a stop on the migratory paths of such birds as sandpipers.

Settlement patterns. After 1945 the suburbs of Wilmington, largely unincorporated, received not only the people fleeing Wilmington but also most of the white newcomers to Delaware. More than half of the state's population lives outside the city of Wilmington but within commuting distance of it. This suburban band includes Delaware's second largest city, Newark. Thus, suburbia has become the seat of Delaware's population, political power, and wealth. Suburban areas are populated generally by whites, although there are a few black enclaves.

Though federal laws reduced the flow of immigrants after World War I, Delaware experienced its largest population growth in the middle of the 20th century. From 1950 to 1960 its population grew by about 40 percent, but the rate declined thereafter. Many of the newcomers were highly skilled scientists or technicians. Wilmington also received a large influx of blacks, many of them unskilled.

The people. As is characteristic of the Middle Atlantic states, the colonial population of Delaware was quite varied. Swedes (and the Finns who came with them), Netherlanders, and African slaves settled in Delaware before the English, mainly in present-day New Castle county in the north. The English settlers came not only from overseas but also from Pennsylvania and Maryland. Some of the settlers from Pennsylvania were Quaker artisans and merchants; the settlers from Maryland were often planters who brought slaves with them. With the English came some Welsh settlers and, after 1715, large numbers of Irish, particularly the Presbyterians of Scottish descent known as Scotch-Irish. Downstate Delaware was settled mainly by the English and by slaves.

Early immigration

After the Revolution a small group of French came to New Castle and Wilmington from the West Indies, and a few, including the progenitors of the du Pont family, came from France. In the mid-19th century there was a large immigration of Germans and Roman Catholic Irish, and at the end of the century Italians, Poles, and Jews came in large numbers, accompanied by smaller groups of Ukrainians, Russians, Scandinavians, and Greeks.

Ethnic groups in rural Delaware include Polish potato growers in Kent, who came from Long Island; Italian mushroom growers at Hockessin; a colony of Finns that originated at Iron Hill after World War I; an Amish settlement near Dover; and the historic groups of mixed-bloods, called Moors and Nanticokes, at Cheswold, in Kent county, and beside Indian River Bay, in Sussex county. In the mid-20th century there was an influx of Latin Americans, mainly Puerto Ricans, who settled in Wilmington, as had the first generation of the earlier immigrants.

The economy. Delaware's prosperity depends upon its favourable location: four of the 10 largest cities in the United States lie within 150 miles (240 kilometres) of Delaware.

Agriculture. Though the number of farmers continues to decline, agriculture remains important. Most of the farmers' cash income comes from poultry raising, centred in Sussex county. The soybean crop continues to be important, and other major agricultural products include corn, milk, and vegetables. The coastal and inland waters produce fish, clams, and crabs. The only mining is of gravel and sand.

Industry. The major economic enterprise in Delaware is manufacturing, especially the chemical industry. Wilmington boasts of being the chemical capital of the world because it is the centre of administration and research of several chemical companies; du Pont, Hercules, and ICI Americas (formerly Atlas) are the largest. Chief chemical products are pigments, nylon, and petrochemicals. Delaware also has automobile-assembly plants, a petroleum refinery, a synthetic rubber plant, packaging plants, textile mills, and various food-processing plants.

Chemical industry

Delaware has made a business of incorporating companies, many of which operate primarily in other states, since early in the 20th century. It offers them favourable laws (frequently revised to reflect changing business conditions), a convenient location, moderate taxation, stable institutions, and a judicial system with experience in corporate litigation.The corporation franchise tax is an important source of state revenues.

Transportation. The chief flow of highway traffic in Delaware is between Wilmington and its suburbs and the interstate traffic crossing northern Delaware between New York or Philadelphia and Baltimore or Washington. Slightly less important is the traffic up and down the state on the du Pont Highway. The state maintains all roads and bridges as well as through streets in municipalities. A joint Delaware–New Jersey agency operates both the twin bridges across the Delaware River near New Castle and a ferry between Lewes and Cape May.

Delaware lies on the railroad passenger line between Philadelphia and Baltimore. Freight service is also available to the southern state line and in northern Delaware. Local bus transportation in the Wilmington area is provided by a public authority. Wilmington is a major port along the Atlantic seaboard and is the site of the state's largest airport.

Administration and social conditions. *Government.* The constitution of Delaware, its fourth, was adopted in 1897 but has been amended many times. Amendments require a two-thirds vote in two successive legislatures, with an election intervening. The governor, who has no veto on amendments, serves a four-year term and may be re-elected only once. Traditionally, the legislature has been strong and the governor relatively weak, but adoption of the cabinet form of government in 1970 centralized and strengthened executive authority. The 62-member bicameral legislature is known as the General Assembly.

An unusual feature of Delaware's judicial system is the retention of the Court of Chancery, which handles equity cases involving civil rights and litigation concerning Delaware corporations. Most other states have merged their chancery into their law courts. The highest court is the Supreme Court, which hears appeals from the Chancery Court and the Superior Court. At the lowest level in the state judiciary are the magistrate courts, presided over by justices of the peace, who seldom are lawyers. All Delaware judges are appointed by the governor.

Weak county governments have been the rule in Delaware. Formerly, each was headed by an elected levy court that set the tax rate and appropriated funds. The levy courts of New Castle and Sussex, however, have been replaced by stronger elected councils, and New Castle also elects a county executive who appoints the chief administrative officers. Delaware is notable for having used the county subdivision known as a hundred, an ancient English governmental unit. It no longer has a governmental function and is retained purely as a geographic name.

Because of Delaware's small size, many things are done by the state that elsewhere would be left to local government. Consequently, state taxes and indebtedness are relatively high, whereas local equivalents are low. The largest source of state income is the tax on personal and corporate incomes. There is no general sales tax and no state property tax. Real estate taxes are the chief support of county and municipal governments. Schools are supported chiefly by the state, but school districts must raise part of the money for new buildings and school operations, including salaries, through property and other taxes, which must be given approval in a referendum.

Role of state government

Democrats and Republicans have been fairly evenly matched in Delaware, although the Democrats have the larger number of registered voters. Many voters decline to list party preference, and numerous swing votes may go to either side. Primaries had little significance until 1978, when they were first used for all offices.

After the American Civil War, Delaware Democrats used their control of such offices as assessor and tax collector to discourage blacks from qualifying as voters, but Republicans sought the black vote and, with its aid, won control of the state early in the 20th century. In 1932 the Democrats abandoned their all-white tradition. At first

they won black votes only for the national ticket, but gradually, during the next two decades, Delaware's blacks, like those in other Northern states, realigned themselves with the Democratic Party. Thereafter, only an exceptionally popular Republican won black support. Bipartisan support was largely responsible for passage of a fair-housing law.

Education. The University of Delaware, established in 1833 as Newark College, has grown from 1,500 students in 1952 to about 20,000 students. A small population makes a medical school too expensive, but the state has arrangements with Jefferson Medical College of Philadelphia to save places in each class for Delaware students. Similar arrangements are made for students in fields such as veterinary science and dentistry, in which no training is offered in Delaware's public institutions. The state government assumes the major responsibility for public education.

Health and welfare. The state has been called upon to provide an increasing number of services for its citizens. The demands for expenditures, especially in education and welfare, have been brought on partly by population growth and immigration of young families with children and partly by recognition of long-ignored needs. Kindergartens, schools for the handicapped, and mental health clinics have been established.

Major museums

Cultural life. Two major museums are located in the outskirts of Wilmington. The Henry Francis du Pont Winterthur Museum and Gardens is noted for its collection of American decorative arts, which are displayed in authentic period rooms. The Hagley Museum portrays the development of American manufacturing through preservation of the early mills and other structures of the du Pont company, as well as by indoor exhibits. Other interesting museums include the Delaware Museum of Natural History, Greenville; the Delaware State Museum, Dover; and the Old Town Hall, Wilmington.

A number of historic houses in the state are permanently open to the public, including the John Dickinson Mansion, near Dover; the Parson Thorne Mansion, in Milford; and several houses in Odessa and New Castle. Several blocks in New Castle surrounding the colonial capitol, known as the Old Court House, remind visitors of the restorations of colonial Williamsburg in Virginia—except that in New Castle very few buildings had to be restored. Immanuel Episcopal Church, on the Green, was begun in 1703; its graveyard contains numerous interesting stones. The Presbyterian Church nearby dates from 1707. No buildings survive from the Dutch period. Old Swedes Church in Wilmington was built in 1698 for a Swedish Lutheran congregation, but it is now Episcopalian. The Swedes brought a tradition of log construction to the New World, but none of their work remains except perhaps portions of a few small log structures.

The state's foremost research library is that of the University of Delaware. Among the specialized libraries, the Hagley Library, featuring business and industrial history, and the library division of the Winterthur Museum, specializing in the decorative arts and crafts, are internationally known. The Wilmington Free Library is the largest unit in the consolidated New Castle county library system. The Delaware State Library Commission serves the lower counties; most towns also support a library of their own.

Wilmington long has been known as a centre associated with a distinguished group of illustrators, many of them pupils, either directly or indirectly, of Howard Pyle, whose work is displayed at the Delaware Art Museum. N.C. Wyeth, a pupil of Pyle, made his home just across the Pennsylvania line at Chadds Ford, which members of his family have made famous as the home of the Brandywine school, a group of mainly genre and narrative painters.

Performing arts

Wilmington has a legitimate theatre, the Playhouse, as well as the Grand Opera House, restored as a state centre for the performing arts. The professional Delaware Theatre Company has its own building in Wilmington. The small village of Arden is remarkable for its theatrical traditions, both amateur and professional, which include annual productions of Gilbert and Sullivan operettas.

Delaware's ocean beaches are popular not only with Delawareans but also with people from neighbouring areas, especially Washington, D.C. Rehoboth and Indian River bays are boating, fishing, and clamming centres. State parks, such as Lum's Pond, are also used for recreation. The week-long Delaware State Fair is held annually in Harrington. Pari-mutuel betting lures crowds to racetracks in Stanton, Dover, and Brandywine Hundred.

HISTORY

The Indians. When the first Europeans arrived, the Delaware (or Lenni Lenape) Indians lived in northern and central Delaware and also along the river shore in Pennsylvania and New Jersey. Their language was a version of the Algonquian tongue. Politically decentralized (each village ran its own affairs), they were a peaceful people, supporting themselves by farming, hunting, and fishing. The more warlike Minqua, or Susquehannock, living to the west, frequently attacked the Lenape. Several Algonquian-language tribes, such as the Nanticoke, Assateague, and Choptank, lived in southern Delaware.

The colony. The Dutch who established the first European settlement in Delaware at Lewes in 1631 were killed by Indians, and it was not until 1638 that a permanent settlement was planted—by Swedes at Fort Christina, now Wilmington; they reputedly erected America's first log cabins in this colony of New Sweden. The Dutch from New Amsterdam (New York) conquered the Swedes in 1655, and the English seized the colony from the Dutch in 1664. Thereafter, except for a brief Dutch reconquest in 1673, Delaware was administered as part of New York until 1682, when the Duke of York ceded it to William Penn, who wanted it so that his colony of Pennsylvania could have access to the ocean. Though Penn tried to unite the Delaware counties with Pennsylvania, both sides resented union. In 1704 he allowed Delaware an assembly of its own. Pennsylvania and Delaware shared an appointed governor until the Revolution. Only in 1776 did the name Delaware—deriving from Sir Thomas West, 12th Baron De La Warr, a governor of Virginia—become official, though it had been applied to the bay in 1610 and gradually thereafter to the adjoining land.

Revolution and statehood. During the Revolution, Delaware was invaded by a British army en route to Philadelphia and was constantly menaced by British ships. The event best remembered, however, is the spectacular ride (July 1–2, 1776) of Caesar Rodney from his home to Philadelphia to break a tie in the Delaware delegation and cast Delaware's vote for independence. The proudest boast of Delaware is that its speedy ratification of the Constitution, on Dec. 7, 1787, gave Delaware its right to be called "the first state."

19th-century political sentiments

As national political parties arose, Delaware became a Federalist state, adhering to the party of Alexander Hamilton and John Adams well into the 1820s. In the next period Delaware became as fervently Whig as it had been Federalist.

Civil War and aftermath. The advent of the Civil War did not seriously tempt Delaware to secession. Delaware had been slave territory since its days as a Dutch colony, but the number of slaves had declined drastically, mainly through voluntary manumissions, from 8,900 in 1790 to 1,800 in 1860. More important was Delaware's economic bond with Pennsylvania and the North, strengthened by the river trade and the new railroad network. Though Unionist in sentiment, Delaware never voted for Lincoln, and the Reconstruction that followed the Civil War drove many voters to the Democratic Party in sympathy with the occupied South. By the end of the 19th century, however, economic realities had regained importance, and Delaware became firmly Republican and remained so until well into the Great Depression of the 1930s.

Economic growth. Delaware's principal city, Wilmington, became a manufacturing city, and by 1920 it contained not only half of the state's population but also at least a similar proportion of its wealth and economic energy. Diversity characterized the products of Wilmington factories, but in the 20th century the city became renowned as an administrative centre for the nation's chemical industry. In the 1980s, after Delaware lifted the limit on interest charges, many major banks from New York City and elsewhere set up subsidiaries in the state. (J.A.Mu.)

District of Columbia

For coverage of the District of Columbia, see the *Macropædia* article WASHINGTON, D.C.

Maryland

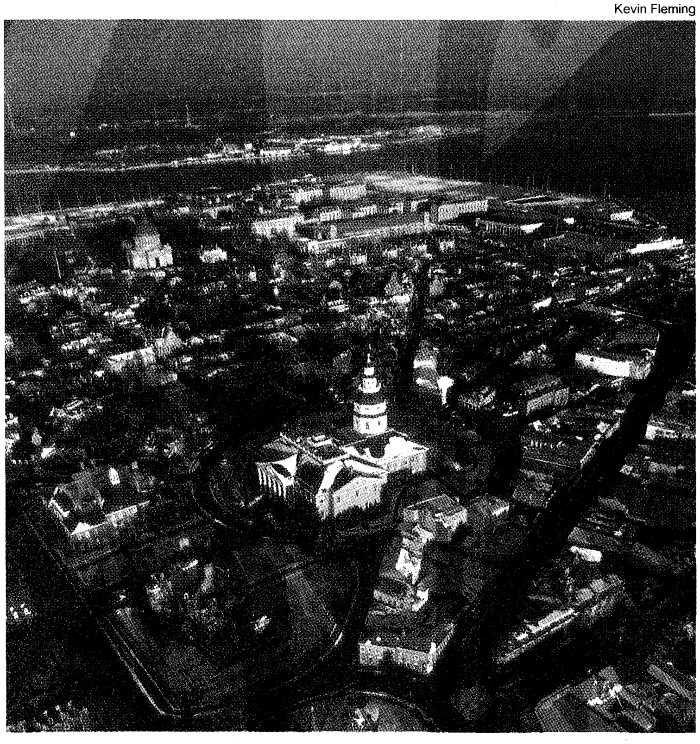
One of the original 13 states of the United States, Maryland lies at the centre of the Eastern Seaboard, astride the great industrial–population complex that stretches from Maine to Virginia. Its small size—10,460 square miles (27,092 square kilometres), about 6 percent of which is water, excluding Chesapeake Bay—belies the great diversity of its landscapes and of the ways of life that they foster: from the low-lying and water-oriented Eastern Shore and Chesapeake Bay area, through the metropolitan hurly-burly of Baltimore, its largest city, to the forested Appalachian foothills and mountains of its western reaches.

Maryland is thought to have been named in honour of the wife of King Charles I by a grateful Cecil Calvert, Lord Baltimore, who in 1632 was granted a charter for the land as a haven in which his fellow Roman Catholics might escape the restrictions placed on them in England. A deep sense of history is notable in the quiet charm of Annapolis, its capital since 1694, and in the white-domed, pillared statehouse, built in 1772, the nation's oldest statehouse in continuous use. Annapolis' 40-block Colonial Historic District contains more pre-Revolutionary structures than any other U.S. historic district. The narrow, crooked streets of Annapolis, the houses abutting directly on the brickwork sidewalks, the graceful tree-covered green about the statehouse, and the myriad masts of boats at dock or anchor in the harbour reflect more an earlier America than a state geared to the latest in technology.

Pivot between North and South

Geography, too, has provided Maryland a role in U.S. history, as a pivot between the North and the South. Its northern border with Pennsylvania is the famous Mason and Dixon Line, drawn in the 1760s to settle disputes between the Penn and Calvert families and traditionally regarded as the boundary between the North and the South. To the south much of the boundary with Virginia is formed by the Potomac River, a symbolic barrier during the Civil War. On the north bank of the Potomac lies the District of Columbia, coterminous with the city of Washington, a small enclave ceded by Maryland in 1791 for the site of the national capital. East of the Chesapeake, the Eastern Shore shares the Delmarva Peninsula with Delaware on the north and Virginia on the south. In the mountainous west Maryland's panhandle, which is joined to the rest of the state by a narrow waist, interlocks with the eastern panhandle of West Virginia.

Kevin Fleming



Maryland State House in Annapolis with the U.S. Naval Academy, the Severn River, and Chesapeake Bay in the background.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The Atlantic Coastal Plain covers about one-half of Maryland's land area, yielding to the Piedmont Plateau at a fall line running from the northern tip of the District of Columbia through Baltimore and to near the northeastern corner of the state. The Catoctin ridgeline in the west forms the gateway to the Appalachians.

To the south the Coastal Plain is sandy; to the north it is loamy and fertile. Its water edges, called salt marshes, or wetlands, exasperate mapmakers as erosion periodically fills in a swamp or deletes an entire island: Blakistone, for example, is about one-tenth its 1634 size. The Chesapeake's 23 estuarial tributaries provide the state with some 3,200 miles (5,150 kilometres) of shoreline—subject to frequent change. The most important of nature-made revisions was an irruption of the ocean, during a storm in 1933, through Fenwick Island into Sinepuxent Bay, just below Ocean City. The lower portion of this barrier sand reef is now a national seashore and state park known as Assateague Island, and the inlet has become a boon to Ocean City's resort fishing fleet.

The Piedmont Plateau has good farming soil except for belts of clay that are mined for brick kilns; from the beginning, the exteriors of Maryland buildings have glowed with salmon-coloured brick made from the state's clay. To the west and parallel to the fall line, the low Parr's Ridge forms a barrier between the Potomac and Chesapeake Bay.

Maryland's share of the Appalachian Mountains comprises a series of forested barriers, with many of the intervening valleys still uncleared. Backbone Mountain, hugging the West Virginia line, is the highest point in Maryland, at 3,360 feet (1,024 metres).

Chesapeake Bay

The most salient feature of Maryland's topography is Chesapeake Bay, which serves the port of Baltimore and divides the Eastern Shore from what was once called Maryland Main. On a summer weekend as many as 100,000 sailboats and powerboats may be seen on the water. But the bay has drawbacks. Swimmers shun its brackish, murky water after the late-summer onset of billions of small, stinging jellyfish, and the cross-bay bridges are often filled to capacity with the crush of summer weekenders going to and coming from the ocean beaches.

Dredging is necessary to maintain the 50-foot ship channel to Baltimore and to the Chesapeake and Delaware Canal. The bay must also be protected against pollution by the municipalities, industries, and farms in its drainage area. The bay was once lined with oysters, but silt, pollutants, and hostile microorganisms have pushed the beds up into tributary rivers and diminished the yield even there. The largest catch is the blue crab, which arrives on a dinner table in such delectable forms as crab soup, crab cakes, steamed hard-shell crabs, soft-shell crabs, and crab imperial. The bay, which was called by the Baltimore sage H.L. Mencken a "great big outdoor protein factory," still affords a precarious living to hundreds of watermen.

Climate. Maryland has two climates. It is continental in the west, with temperature records from −40° F (−40° C) to more than 100° F (38° C). In the east a humid subtropical climate is dependent on the Azores High (a pressure area that moderates the weather but does not prevent ice formation almost every winter on Chesapeake Bay's northern tributaries), with summer calms as high as 107° F (42° C), and nearly 100 percent relative humidity. Ordinarily, rains are enough to make reservoirs overflow and to enable Baltimore and Washington, D.C., to draw all the soft water needed for drinking. Storms sweep in from the west and south, except in late summer, when the fringes of passing hurricanes often drench Maryland from the east.

Settlement patterns. Sectionalism within Maryland is dictated by terrain. The Eastern Shore farmers concentrate

on chickens, corn (maize), and soybeans; the factory-style output of broilers (young chickens) is immense. A mercantile appendage of Wilmington and Philadelphia until the bay was bridged, the nine-county Shore has become a vacation and retirement spot for the affluent, who appreciate the privacy of its flat, wooded, little-posted estate areas serpentined with creeks, coves, guts, necks, and inlets.

Southern Maryland's five counties have built a way of life around state government, tobacco growing, military installations, and, increasingly, residential areas for Washingtonians. Thus, Prince George's county, almost one big suburb, has become Maryland's most populous county.

Central Maryland comprises the city of Baltimore and five counties. Four of the counties contain most of Baltimore's suburbs; the fifth is Montgomery, on Washington's northwestern edge. Only about one-sixth of Marylanders live outside metropolitan areas—and central Maryland is one long, contiguous metropolitan area.

The four counties of western Maryland owe much to road, railroad, and canal builders. The barging of coal and grain ceased in 1924, but the creation of the Chesapeake and Ohio Canal National Historical Park in 1971 assured a stream of excursionists. Interstate and national roads carry city dwellers to Garrett county, where mountainside ski runs complement water sports on Deep Creek Lake, the largest man-made body of water in the state.

The people. The white population, at first all from the British Isles, began to vary when German-speaking farmers and artisans moved from Pennsylvania into western Maryland during the 1700s. The process accelerated in the 1840s during Ireland's potato famine and as Germans and German Jews fled military conscription, and then Russian Jews, Poles, Czechs, Italians, Greeks, and others arrived at Baltimore, which was a major 19th-century immigration centre, and later fanned out into the countryside. Ethic diversity was one of the first characteristics that set Maryland apart from the regions below the Potomac River. Immediately after the American Civil War, this diversity of the state was countered by an influx of Southerners who despaired of life in a defeated and devastated homeland.

The ethnic mix

Maryland's Indian population had been killed off or pushed westward by about 1700. All that remains from their centuries of occupancy are campsite artifacts, still being unearthed; some impressive bayside oyster middens; and place-names, corrupted by uncomprehending whites, such as Chesapeake, Patapsco, Potomac, Wicomico, Patuxent, Piscataway, and Susquehanna.

African slaves were at work in Maryland under the first Calverts. The consciences of many Marylanders, particularly members of the Society of Friends (Quakers), were uneasy; from 1783 the importation of kidnapped Africans was heavily taxed. While Maryland did not formally outlaw slavery until 1864, it protected the liberty of more free blacks than any other state as slavery neared its end. After the Civil War, blacks found Maryland more congenial than the states of the Confederacy, in which, by 1900, systematic lifetime disfranchisement of blacks was under way. A similar effort in Maryland, led by the Democratic Party and coming to a head in 1910, was beaten in referendum by Republicans. Yet it took a U.S. Supreme Court decision in 1934 to force the University of Maryland to admit a black into its School of Law; and it was 1970 before Marylanders sent a black representative to Congress. The latter reflects in part the changing population of Baltimore, which now is more than half black.

The heaviest concentrations of population are centred around Washington, D.C., and Baltimore. Although the rural stretch between the two cities—only 40 miles apart—has been diminishing, the growth outward from the cities has been uneven, and thus a fusion into a common, uninterrupted cityscape has not occurred.

Planned city of Columbia

One attempt to fill in the Washington–Baltimore gap is the planned, nationally watched city of Columbia, in Howard county. Although it had been created only in the mid-1960s, Columbia was by the 1980s home to 70,000 people of an intended 100,000 or so. It is governed by a private association rather than an elected government. A community not only of cars and shopping malls but also with ample greenery and other amenities, Columbia is slightly closer to Baltimore, yet a majority of its residents work in or near Washington, D.C.

Baltimore is the largest city in the state, but it continues to lose people to the suburbs. Calculations for the next largest cities are impeded by the tendency of municipalities not to incorporate; thus, boundary lines are drawn arbitrarily by census takers.

The economy. Service businesses pervade the economy, and agriculture directly supports only a small number of people. The state has taken an increasing interest in the well-being of the private sector, particularly through its Department of Economic and Community Development. The agency encourages outside firms to locate in Maryland, promotes tourism, and keeps a close watch on the economy of the state.

Industry and agriculture. The major industries are primary metals, electronic and electrical equipment, transportation equipment, food and kindred products, and printing and publishing. Most large establishments are branches of out-of-state corporations, but wages for production workers are slightly above the national average. Electricity is generated at a nuclear plant at Calvert Cliffs; coal mining, in contrast, has ebbed. Maryland has one category of nationwide preeminence: crabs. The state's crab haul consistently surpasses that of other states. Saltwater staples also include other shellfish, as well as perch, flounder, and other finfish.

Transportation. Maryland offers the traveler from the north three trunk highways into Baltimore and four highways south to Washington, D.C. It is possible to swing around Baltimore and Washington on beltways or to avoid them altogether by Eastern Shore routes. The Harbor Tunnel Thruway and Francis Scott Key Bridge provide routes across Baltimore's harbour.

Amtrak provides passenger rail service to Baltimore, and residents of Montgomery and Prince George's counties commute to Washington, D.C., on the Metro subway. Baltimore, too, has built a Metro subway between Owings Mills, a northwest suburb, and downtown. Baltimore–Washington International Airport is augmented by numerous public airports throughout the state. For freight shipments, the port of Baltimore has excellent facilities and is one of the nation's busiest ports. The port, supervised by a state agency, is especially adapted to bulk commodities, container shipments, and foreign-made automobiles.

Administration and social conditions. *Government.* In spite of a provision for statewide voting every 20 years on whether to summon a constitutional convention, repeated attempts to scrap the 1867 document, with its unnecessary detail, obsolete concerns, and silence on points of modern interest, have been failures. The document has been amended some 200 times.

The form of state government in Maryland is like that of most other U.S. states. The governor, who serves for four years, may be reelected to an immediately succeeding term only once. A 1969–72 reorganization of the state government brought together in 12 departments several hundred separate agencies, boards, and commissions.

Executive, legislature, and judiciary

Members of the General Assembly serve four years and may be reelected indefinitely. Reorganization in the 1960s ended rural domination of the legislature and passed power to the counties adjoining Baltimore and Washington, D.C. The General Assembly consists of the Senate, with 47 members, and the House of Delegates, with 141 members.

Below the seven-member Court of Appeals, the highest judicial body, is an intermediate Court of Special Appeals and echelons of circuit, district, and other courts. Judges are appointed by the governor and must run against their records (and, in Baltimore alone, circuit court judges must run against anyone filing in opposition) in the election following appointment. Appeals judges are elected to 10-year terms, and circuit judges to 15-year terms.

To avoid the greater costs that would be entailed by incorporation as governmental bodies, many of Maryland's most populous areas remain unofficial entities, their services provided on a countywide basis. A constitutional amendment allows home rule for counties, under special charter. A charter county is governed by a county executive and county council, both elected to four-year terms;

the council is empowered to enact all local laws. Eleven mostly rural counties are still governed by elected boards of county commissioners. The city of Baltimore, a separate jurisdiction, is run by an elected mayor and city council and a mostly appointive Board of Estimates.

Maryland politics

In the 20th century, although free of the intimidation, poll tax, and other evils practiced in some places in the South, many Maryland elections have been machine dominated. The larger group of voters consistently has registered as Democrats, and the Democratic Party usually, but not always, wins. Exceptions have occurred when the Democrats split internally, when a nominee's platform was considered extreme, or when a Republican presidential candidate carried the local candidates. A Republican nominee can expect to do well in the western counties, one or two southern enclaves, the Eastern Shore, and some affluent parts of suburban Maryland. Baltimore, with its party-boss tradition, is a Democratic stronghold, and college-educated and consciously ethnic voters are dependably Democratic. The black vote, long staunchly Republican, switched parties during the New Deal of the 1930s.

The General Assembly remains heavily Democratic, as it has been for decades. Thus, Democratic primary elections are often more important than the general election. When the nation's axis of balance was North–South, numerous presidential nominating conventions assembled in Baltimore, but none has been held there since 1912. In 1969 Marylanders chose their first Jewish governor, in 1970 their first black congressman, and in 1986 their first woman U.S. senator. All were Democrats.

Through its revenues, largely derived from an income tax, federal aid, retail sales taxes, and a lottery, the state has assumed responsibilities that are no longer within the capacity of the local government. The state pays all full-time judges; the governor appoints the police commissioner of Baltimore; a state committee approves the construction of all public school buildings; and the state and the city jointly operate Baltimore's airport. The state provides annual subsidies to Baltimore's leading orchestra, art museum, public library, and resident theatre. In many instances, as population and industry have moved outward, as old structures have been demolished and property tax assessments and yield have dipped, municipalities have taken the initiative in this transfer of responsibilities to the state.

Education. Control of public education in Maryland is vested in a state board of education, a board of higher education, and Baltimore and county school boards. All positions are appointive, except in 10 counties where board members are elected to office. State supervision and the support of county public school systems began in 1870, but not until 1951 were 12 years of schooling uniformly required in all counties of the state. The state supports local systems, particularly as regards library services, vocational and rehabilitational instruction, and utilization of federal aid.

Colleges and universities

There are two-year community colleges in Baltimore and in several other locations. Crowning the state's system of higher education is the University of Maryland, with its main campus in College Park and branches located in Baltimore, Catonsville, and Princess Anne. The university's origins date from the College of Medicine of Maryland (opened in 1807) and Maryland Agricultural College (1856). The several graduate and undergraduate schools of the University of Maryland were consolidated in 1920, and in enrollment it has become one of the nation's largest universities.

Maryland has several private institutions of higher learning. The most prominent of these are Johns Hopkins University (founded 1876), with several campuses, a world famous medical school, and the Peabody Institute, a music school; the U.S. Naval Academy (1845) in Annapolis; St. John's College (1784) in Annapolis, noted for its emphasis on the great books of the Western world; Goucher College (1885) in Towson; and the Maryland Institute, College of Art (1826), in Baltimore.

Health and welfare. The Department of Health and Mental Hygiene, in addition to supervising county services, provides hospital therapy and care for alcoholism, drug abuse, mental retardation, mental illness, and other afflictions. The department is also active in preventive medicine, for instance, providing programs of education against drug abuse. The state's Shock Trauma Center at University Hospital in Baltimore conveys the severely injured from anywhere in the state by helicopter within one hour for immediate life-saving treatment.

The Department of the Environment is the centre for state efforts to prevent or reduce pollution. It monitors the state's water supply and sewage, air quality, and solid-waste disposal.

The Department of Human Resources is in charge of state welfare activities. Direct aid to families with dependent children is its largest outlay, followed by general public assistance and foster care.

Cultural life. Maryland has its share of such nationwide phenomena as institutional uniformity, ebbing respect for local and regional distinctiveness, and transitoriness of residence, and such problems as graffiti, smog, illiteracy, unwed teenage parenthood, drunken driving, and street crime. The setting includes considerable decay, particularly in residential neighbourhoods; the garish commercialization of streets and highways; the urban waste spaces where torn-down buildings have not been replaced; and the degradation of the slums—some of them rural.

Distinctive cultural remnants

At the same time, however, the high proportion of Marylanders who are long-time residents (some represent many generations) retain an unusual number of distinctive interests and activities. Echoes of landed gentry of England and of early Maryland are stirred by riding to hounds in pursuit of the fox and taking part in the Grand National Steeplechase and Maryland Hunt Cup races, by breeding Chesapeake Bay retrievers, and by jousting on horseback with a spear at rings dangling from a crossbar. The outdoors looms large in the life of Marylanders. Popular activities in the state include sailing and crabbing on the bay, trolling for ocean marlin, shooting wild ducks and geese in the marshes of the Shore, playing the game of lacrosse (derived from an Indian sport), beachcombing for fossil shark teeth at Calvert Cliffs, exploring mountain caves, and hiking the 38-mile (61-kilometre) stretch of the Appalachian Trail across Maryland in a single day.

Sailboat and sports-car racing attracts both participants and spectators. Audiences that extend beyond Maryland's borders are drawn to the horse racing at Laurel and Pimlico (notably the Preakness, which is run in May), Baltimore's baseball team, the Orioles, and Washington's ice hockey and basketball teams, both based in Prince George's county. Restaurants present a plethora of cuisines, but the traditional gastronomy of Marylanders tends to be centred on terrapin soup, steamed crabs, padded oysters, fish and crab cakes, and beaten biscuits, often washed down with Maryland rye and beer. In almost a different world are the tiny communities of watermen located on various islands of Chesapeake Bay, in which isolation has worked to preserve distinct attitudes and ways of life.

Prominent Marylanders

Among Marylanders who have made major contributions to artistic and intellectual traditions are the critic H.L. Mencken, the black abolitionist-statesman Frederick Douglass, the antislavery activist Harriet Tubman, the philanthropists George Peabody, Johns Hopkins, Enoch Pratt, and Henry Walters, and, in the 20th century, the novelists Anne Tyler and John Barth and the filmmakers Barry Levinson and John Waters.

Baltimore is the centre of much of the state's activity in the arts. It is the home of the Baltimore Symphony Orchestra and of several professional theatres, including the Morris A. Mechanic Theater, Center Stage, Theatre Project, and numerous suburban dinner theatres. The Baltimore Museum of Art, Walters Art Gallery, Maryland Historical Society, City Life Museums, National Aquarium, U.S. Frigate *Constellation,* Maryland Academy of Sciences, B&O Railroad Museum, Baltimore Streetcar Museum, and Baltimore Museum of Industry (all in Baltimore city) are supplemented by galleries and display centres elsewhere, as well as by fairs and festivals. The most notable among them is the Maryland State Fair, which takes place in late summer at the State Fairgrounds in Timonium, north of Baltimore. Tourists throng to Bal-

timore's Inner Harbor year-round. Developed mostly in the 1980s, the Inner Harbor is a waterside array of high-fashion stores, new hotels, performances, moving boats, docked ships, and venues for eating and drinking.

HISTORY

The area's earliest human occupation is accepted as having been by roving hunters in about 10,000 BC, as the ice sheet made its final retreat. The records of this pre-Archaic, fluted-blade culture, which left only the points of its weapons, remain imprecise. Later, the numerous Eastern Archaic and then Woodland Indian populations practiced agriculture and feasted on seafood; by AD 1000, permanent villages were established. During the early European settlement the tribes were Algonquian in language and politics, but they were under pressure from the Iroquois to the north. The English promise of support in these wars greatly smoothed relations in the early colonial years.

The colony. Leonard Calvert, the younger brother of Lord Baltimore, landed the founding expedition on St. Clement's (now Blakistone) Island in the lower Potomac in March 1634. The first settlement and capital was St. Mary's City. Aware of the mistakes made by Virginia's first colonists, Maryland's settlers, rather than hunt for gold, made peace with the local Indians and established farms and trading posts, at first on the shores and islands of the lower Chesapeake. The field hands included indentured labourers working off the terms of their passage and, after about 1639, African slaves. The most important crop was tobacco. Roads and towns were few, and contact with the English-model manor houses was largely by water.

Religious toleration in the colony

The religious latitude stipulated by the Calvert family was formalized by the General Assembly in 1649 in an Act Concerning Religion, later famous as the Act of Religious Toleration. It granted freedom of worship, though only within the bounds of Trinitarian Christianity. Commercial disputes with Anglican Virginia and boundary quarrels with Quaker Pennsylvania and Delaware did not affect this tolerance. Puritan ascendancy in England (1648–60) caused only brief turmoil, and during an interval of crown rule in Maryland (1692–1715) the Church of England was formally established. Maryland nonetheless remained a haven for dissidents from sectarian rigidity in other colonies.

As the population centre shifted to the north and west, the capital was moved to Annapolis, and in 1729 Baltimore was founded. Maryland's dominant "country party" early resisted British efforts to make the colonies bear more of the costs of government. Frederick county repudiated the Stamp Act in 1765, and in 1774, the year after the Boston Tea Party, a ship loaded with tea was burned at an Annapolis dock.

Marylanders took an active part in the American Revolution. The Continental Congress, often on the move to avoid British troops, spent a winter in Baltimore. At the close of the war it convened in Annapolis, where it accepted George Washington's resignation from the army and ratified the Treaty of Paris (1783), which acknowledged the independence of the colonies.

Postwar problems included the disposition of confiscated loyalist property, the struggle for paper money, and debtor relief. Maryland's controversy with Virginia over the use of the Potomac and lower Chesapeake Bay, resulting in the Compact of 1785, led toward the Constitutional Convention, as did the Annapolis Convention of 1786, at which Maryland was not represented. Luther Martin distinguished himself as a representative of Maryland at the Philadelphia Convention of 1787. Maryland ratified the Constitution on April 28, 1788, the seventh state to do so. It also ceded territory and advanced money for public buildings to help form the District of Columbia (1791).

The state. When harassment on the high seas and other factors brought on the War of 1812, Baltimore clippers, sailing as privateers, dealt more than equal punishment to British skippers. In 1814 the British troops who had burned the principal government buildings in Washington, D.C., were repulsed in their attempts to inflict similar punishment on Baltimore. Francis Scott Key, a Georgetown lawyer and an eyewitness of the futile bombardment of Fort McHenry in Baltimore harbour, wrote the four eight-line stanzas that, set to existing music, became the national anthem, the "Star-Spangled Banner."

With peace, Maryland and the nation occupied themselves with improvements in transport and communication. The National Pike, the first road to cross the Appalachians, was completed in 1818 and was followed by the first U.S. passenger railroad, the Baltimore & Ohio, and two canals, the Chesapeake and Ohio from Washington to Cumberland and, across the top of the Delmarva Peninsula, the Chesapeake and Delaware. The nation's first intercity telegraph line was constructed between Washington, D.C., and Baltimore. In 1845 the U.S. Naval Academy was founded on the Severn River in Annapolis.

Impact of the Civil War

The Civil War, however, arrested Maryland's progress. Landed gentry and residents of the Eastern Shore supported the secessionists, while workingmen and western Marylanders stood up for the Union; a third faction favoured neutrality. Federal troops occupied Baltimore and Annapolis, and martial law was imposed in this border state. Confederate armies mounted three major invasions in successive summers; they were checked at Antietam, they met full defeat at Gettysburg, in Pennsylvania, and their threat to Washington, D.C., was dissipated in 1864. The constitution of 1864 abolished slavery and removed power from the rural aristocracy. The more cautious constitution of 1867 remains in force.

Since 1865. After the Civil War Maryland prospered. The state was first an important entrepôt for raw materials from and consumer goods to the South and Midwest and became a growing centre of industry that rarely was controlled from within the state. Excesses that had won Baltimore the epithet "mob town" gradually were quieted. For many decades, until the mid-1900s, Marylanders consistently voted for Democratic candidates. Increasingly, however, the character of Maryland began to change because of its proximity to the seat of national government. The state became a major centre for federal installations, both military and civilian, during World Wars I and II and afterward—most famously, it found itself home to the presidential weekend retreat called Camp David, on Catoctin Mountain. But most important was the radically different face of the Maryland suburbs of Washington, D.C., which reflected change not only in the greater numbers of people but also in their unusually high educational and economic status.

When the nation took its first census, in 1790, the centre of population was found to be in Maryland. That site has long since moved far to the west. But Maryland remains, for all its lack of size, as close to a median or model for the whole 50 states as any other state can claim to be. (Ja.H.B.)

New Jersey

One of the original 13 U.S. states, New Jersey is bounded by New York on the north and northeast, the Atlantic Ocean on the east and south, and Delaware and Pennsylvania on the west. The state was named for the island of Jersey in the English Channel. The capital is Trenton.

Although it has major social, economic, and political force in its own right, New Jersey is often looked upon as a stepchild among the heavily industrialized and populated Eastern Seaboard states of the United States. Only four states in the nation have a smaller area than its 7,787 square miles (20,169 square kilometres), and only a few states have a larger population. Nonetheless, as a geographic entity and as a human collective, New Jersey suffers from a lack of identity among U.S. states. For the hundreds of thousands of its citizens who commute to New York and Pennsylvania, New Jersey is a vast bedroom state. Its transportation system, one of the busiest and most extensive in the world, is primarily a funnel for goods and people moving to New York City and other points north and to Philadelphia and Delaware and to points south. For hundreds of thousands of visitors it offers long stretches of fine beaches along the Atlantic Ocean, and the resort of Atlantic City may be better known than the state itself.

Contradictory nature of the state

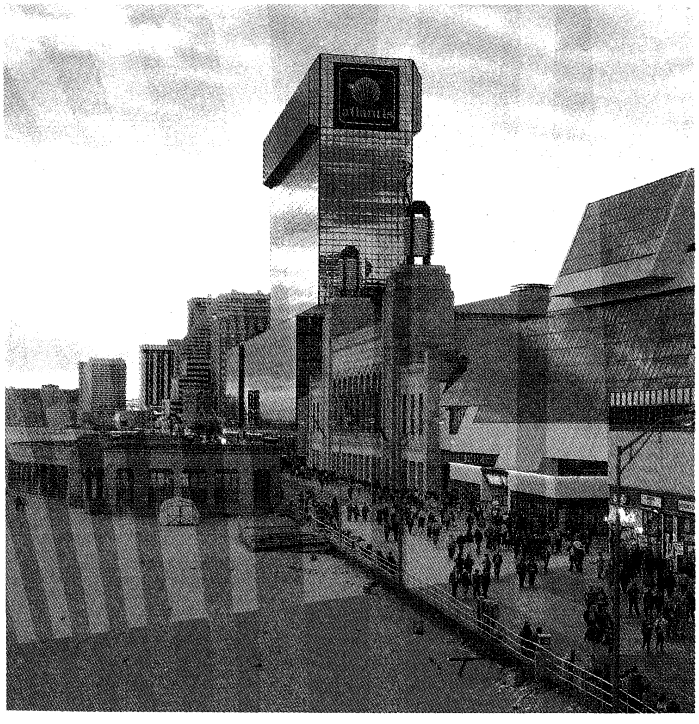
Above all, New Jersey is rife with contradiction and anomaly. Its people fiercely fight off attempts of state government to end home rule by powerful municipal administrations, which have contributed heavily to the almost uniform depression and decay of its cities. Whereas it has produced some of the most able and respected U.S. governors in the 20th century, its local politics have often been astonishingly corrupt, and it has achieved notoriety as a major haven for organized crime.

New Jersey is called the Garden State because it was famous for its fertility in the 18th century; it is now also among the most urbanized and crowded of states. The urban density of its northeast contrasts sharply, however, with the rugged hills of the northwest and with the enormous stretches of pine forest in the southeast and the rolling and lush horse country in the south central part of the state. New Jersey is an important industrial centre, but it has paid the price in environmental pollution, in dirt and noise, and in congested roads and slums. In sum, New Jersey is a curious amalgam of the urban and rural, the poor and wealthy, the progressive and backward, the parochial and cosmopolitan. Indeed, it is one of the most diverse states in the Union.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* New Jersey comprises four distinct physical regions: the Ridge and Valley section of the northwest, where the folded Appalachians slice across the state; the New England section, a southern extension of the ancient rocks of New England, which also trend across the state in a northeast–southwest direction; the rolling central Piedmont, where many of the major cities and suburbs are located; and the relatively level Atlantic Coastal Plain, which is divided into an inner and an outer portion. It is in the Outer Coastal Plain, with its relatively poor, sandy soils, that the Pine Barrens are located. The best soils are located on the Inner Coastal Plain and on the Piedmont and in valleys in the New England section south of the last glacial advance. Although suburbanization is rapidly devouring New Jersey's agricultural land, substantial estates and farms still exist in parts of the Piedmont, and truck farms still dominate many parts of the southern Inner Coastal Plain. Elliptic tracks reveal prosperous horse farms in the northern Inner Coastal Plain, and dairying still exists in the Ridge and Valley section. The most striking features of the state are its beaches, the Pine Barrens, the Palisades facing Manhattan, the broad marshes and swampland in the northeast, and the hills of the northwest, including the famous Delaware Water Gap.

Miro Vintoniv—Stock, Boston



The Boardwalk in Atlantic City, N.J.

Drainage. Lakes and ponds cover about 315 square miles of the state's surface. New Jersey's major river, which it shares with Pennsylvania, is the Delaware. The Hudson River separates the state from New York. Other major rivers are the Passaic and the Hackensack, both in the northeast, and the Raritan, which runs west to east and is generally regarded as the boundary between North and Central Jersey. Lake Hopatcong in Sussex and Morris counties is the state's largest.

Climate. The northwest experiences relatively cold winters, with average January temperatures of less than 28° F (−2° C). Relatively mild conditions prevail in the south, with average winter temperatures above freezing (0° C). Summers are relatively hot throughout the state, with averages for July ranging from about 70° F (21° C) in the northwest to over 76° F (24° C) in the southwest. Moist conditions prevail, with seasonally well-distributed precipitation averaging from 44 inches to more than 48 inches (1,120 to 1,220 millimetres).

Plant and animal life. Virtually all of the plant life that is common to the northeastern United States can be found in New Jersey, and in the marshes and Pine Barrens of southern New Jersey many rare plant species grow, including some insect-ingesting plants. The Barrens are dominated, however, by oak and pine on the well-drained sites and by white cedar in the poorly drained bogs. Common trees elsewhere include oaks, sugar maple, hemlock, birch, ash, sweet gum, and other deciduous species. Common plants are wild azalea, rhododendron, honeysuckle, mountain laurel, wintergreen, and cardinal flower.

The Hackensack Meadows, west of the Palisades, and the Great Swamp of Morris county are relics of glacial lakes of the last ice age. The former is dominated by grasses, the latter by trees. The Hackensack Meadows are managed to encourage wise land use and pollution abatement. The Great Swamp, one of several poorly drained areas in the Passaic River basin, is a National Wildlife Refuge. Elsewhere, increasing suburban development has encroached on wildlife habitats, although bears and wildcats can still be found in some woodlands. Deer are common, even in many suburbs, as are opossums and raccoons. Other mammals, snakes, and birds common to the Northeast (including migratory species) are also to be found.

The Shore

Settlement patterns. The most distinctive of New Jersey's regions is its long shore, which stretches for 125 miles (200 kilometres). Much of it is composed of long and narrow barrier islands separated from the mainland by shallow lagoons and from one another by tidal inlets. Cape May, at the southern tip of the state, was the first U.S. summer resort, and both that community and Long Branch in Monmouth county were known as the playgrounds of presidents during the 19th century. The quality of the Shore, as it is called, ranges from the urban garishness of Asbury Park to the opulence of Deal and Mantoloking. In such resorts as Wildwood and Atlantic City, the nightlife plies its revels until dawn, whereas such other towns as Avalon, Ocean City, and Beach Haven are family resorts. The Jersey Shore at its best can be found in its two state parks, Sandy Hook in the north and Island Beach in the south. The dunes there are still topped with coarse but fragile grass, and the osprey still build their nests. The marshes teem with wildlife, and the trees are bent and twisted by wind and salt spray.

Five northeastern counties in the New York City metropolitan area—Essex, Hudson, Passaic, Bergen, and Union—contain about two-fifths of New Jersey's population. Four of the six largest cities in the state—Newark, Jersey City, Paterson, and Elizabeth—are located there. The Newark–Hudson county–Elizabeth complex appears to many travelers as one endless industrial city: dingy, smelly, but throbbing with commercial life.

Beyond the cities lie the suburbs. Most are pleasant and prosperous, but some are old and show signs of urban blight. Industrial construction in suburban communities has increased, but many suburban towns, especially in Bergen county, remain bedroom communities of New

York City and of the New Jersey cities. Newark's population doubles every day as the work force pours in. New Jersey remains dominated, however, by the two giant cities just beyond its borders. Hundreds of thousands of Jerseyans cross the Hudson to New York City on the average weekday. North Jerseyans watch New York television, root for New York athletic teams, and patronize New York theatres and restaurants. The same situation exists in Camden, Burlington, Gloucester, and Mercer counties, where residents cross the Delaware to jobs in Philadelphia.

South Jersey

South of Trenton begins South Jersey, comprising most of eight counties. It includes roughly one-half of the state's area but only about one-fourth of the population. The loamy soil of the Inner Coastal Plain is well suited to vegetable farming, and most of the land not covered by forest or marsh is farmed. Less than 1 percent of the state's population is engaged in farming, but farm income per acre is among the highest in the nation. The Pinelands National Reserve, covering about 1,700 square miles (4,400 square kilometres) in the Outer Coastal Plain, was established in 1978; it was the country's first national reserve, in which the federal government provides funds for the purchase of a core of undeveloped land while state and local authorities are responsible for resource evaluation and economic planning in surrounding developed areas.

Central Jersey, all of five counties and part of three, is largely a plain, but hilly areas occur in Hunterdon and Somerset counties. Middlesex and Mercer counties, especially the former, are industrialized. Princeton University is located in Mercer county and the borough of Princeton, which combines the charm of the campus with a rich colonial past to create one of the nation's loveliest towns. Rutgers, the state university, is in nearby New Brunswick. Hunterdon and Somerset counties are a mixture of suburban development, farmland, and woodland.

The four counties of northwestern New Jersey take in a mixture of small town, affluent suburb, and rugged countryside, although two major cities, Passaic and Paterson, are located in Passaic county. The area contains numerous dairy farms and parks and recreation areas.

Hackensack Meadows

Although relatively small, the Hackensack Meadows are immensely valuable because of their location in the centre of the world's busiest metropolitan area. By the late 1960s, New Jersey had put together the machinery to develop this area through the Hackensack Meadowlands Development Commission. In 1976 the Meadowlands Sports Complex began operations with the opening of a racetrack. In 1977 the Giants (football) Stadium was completed, and the Brendan Byrne Arena opened in 1981. Warehouses and corporate buildings have also been constructed in the area.

The people. New Jersey's population reflects the immigration patterns of the 19th and 20th centuries, including Germans and Slavs, Russian and European Jews, Irish and Italians. It was a prime destination for the waves of blacks that left the South during and after World War II. The state has a sizable Puerto Rican population, and many Cubans who left their country after Castro's revolution settled in North Jersey, mainly Hudson county.

Italian-Americans are the state's largest ethnic group. They are the predominant white bloc in the cities, although the cities also contain sizable Polish, Hungarian, and other eastern European groups. Italian-Americans and blacks dominate the political and cultural life of the cities—a situation that often has brought the two groups into competition and conflict.

The most striking demographic trend in New Jersey is the movement of the white population away from the cities and the concurrent proportional growth of the urban black population, accompanied by an emigration of industry and commerce. Old, outmoded factories are left behind for sleek new buildings outside cities, and huge suburban shopping centres have replaced the downtown department stores. This shift means more jobs in the suburbs, and the jobs create a commensurate demand for housing. The cities bordering these areas have become increasingly black, a situation that is potentially explosive because of the continuing decay of the cities and the poor quality of housing and services the cities can supply.

The economy. Alexander Hamilton's attempt in 1791 to build the nation's first industrial town at Paterson was a failure. He had the right idea, however, for New Jersey was destined to become an industrial giant.

The State Division of Economic Development, along with the major utilities and business organizations, conducts an effective program of selling New Jersey to industry. The state has attracted many industries, especially corporate headquarters from New York City, largely through its greater space, better transportation, and favourable tax rates. The spread of industry and housing, however, has cost New Jersey much of its farmland, the most valuable per acre in the nation.

Research. New Jersey has a large and prestigious research industry. The state has one of the highest numbers of engineers and scientists per capita in the nation. The great inventor Thomas Edison established a research laboratory in Menlo Park in 1876. There he created the electric light and the phonograph and pioneered the technology of the motion-picture industry. Edison's primitive laboratory has been succeeded by the David Sarnoff Research Center, the Bell Telephone and Western Electric laboratories, the Institute for Advanced Study at Princeton, the James Forrestal Research Center, and many other modern centres dotting the landscape of Central Jersey.

Leadership in scientific research

Industry. New Jersey's major industry is chemicals. The next largest is the manufacture of electrical and electronic equipment. New Jersey also ranks high in the production of clothing, food, toys, sporting goods, and stone, glass, and clay products. The state is not noted for its mineral deposits, but North Jersey provided much of the iron for the shot and cannon used during the American Revolution. New Jersey's mineral products now largely consist of clays, sand, and gravel.

The resort industry is a big factor in New Jersey's economy, especially in the south, where a bad year at the Shore hurts the economic well-being of the entire region. In 1976, residents of New Jersey approved a constitutional amendment to permit gambling casinos at Atlantic City, which, with its huge Convention Hall and fine hotels, does a thriving winter convention business. Casinos are taxed, and tax moneys go to senior citizens and the disabled.

Transportation. Since colonial days, when New Jersey's toll roads linking Philadelphia and New York City were a major industry, transportation has been the lifeblood of New Jersey's economy, and its role in New Jersey can best be appreciated in the Newark area. There, 12 lanes of the New Jersey Turnpike converge with the main line of the Penn Central Railroad, Newark Airport, Port Newark, and Port Elizabeth to provide a steady stream of goods and people. The economy of northern New Jersey is bound tightly to that of New York City, and the commercial traffic between the two states is the nation's heaviest.

Port Authority of New York and New Jersey

In 1921 the states of New York and New Jersey formed the Port of New York Authority, now called the Port Authority of New York and New Jersey—a bistate commission empowered to finance and operate transportation facilities in the New York metropolitan area. The Port Authority, as it is known, is a public corporation operating Newark and Teterboro airports in New Jersey and La Guardia and Kennedy airports in New York, as well as the Lincoln and Holland tunnels, three bridges, huge piers, bus and rail lines, and truck terminals.

A similar but much smaller transit complex exists in the Camden area, linking the South Jersey area with Philadelphia. There is a deepwater port at Camden on the Delaware and high-speed transit to Philadelphia. The turnpike runs 132 miles (212 kilometres), the length of the state from the George Washington Bridge in the north to the Delaware Memorial Bridge in the south; the Garden State Parkway stretches 152 miles (245 kilometres) north to south; and the Atlantic City Expressway runs 44 miles (71 kilometres) to connect Atlantic City with the Camden area. Even before World War II, New Jersey had one of the finest road networks in the nation.

Administration and social conditions. *Government.* New Jersey has had three constitutions. The current constitution dates from 1947 and has been amended more than 30 times. New Jersey governors serve terms of four years, and they are permitted reelection to a second term. Be-

cause no other state official runs on a statewide basis, the governor has no rivals in the executive branch. The governor appoints, with the advice and consent of the Senate, virtually all top state officers and members of state boards, authorities, and the judiciary, and he has the authority to supersede county prosecutors. By virtue of his broad executive and administrative powers, the vast patronage at his disposal, and his unequalled access to the press, radio, and television, the governor of New Jersey is as strong a chief executive as there is in the nation. A two-thirds majority of both legislative houses is necessary to override his veto. The governor can be checked and occasionally defeated by a rebellious legislature or by political leaders, but New Jersey governors seldom lose wars, only battles. In 1966, in response to a U.S. Supreme Court decision, the New Jersey legislature adopted the "one man, one vote" principle. Voters in each of the state's 40 districts choose one senator and two general assembly members. Assemblymen are elected every two years, while senators run for one two-year term and two four-year terms each decade.

Local government

New Jersey's 21 counties are administered by boards of freeholders elected countywide. The boards vary from three to nine members, depending on the size of the county. In addition to these elected officials, local governments are supplemented by service commissions, boards, and authorities, many of which enjoy wide independence and even autonomy. Attempts to merge municipalities, reduce the number of school districts by consolidation, or strengthen county government have not been successful; local power and prerogatives are important determinants in New Jersey politics.

In national elections New Jersey is a swing state. It tends to lean Republican, but the Democrats frequently control the state legislature. Although there have been signs that civic bossism is declining, New Jersey's political system was long dominated by powerful county leaders who drew their power from the patronage and contracts that they dispensed through control of the municipal courthouse or city hall. The most notorious of these bosses was Frank Hague, who ruled Jersey City and Hudson county from 1917 to 1947. For three decades Hague dominated the Democratic Party and heavily influenced the Republicans. His philosophy of government was best summed up in his famous reply to critics: "I am the law."

Education. Public elementary and secondary schools in New Jersey are largely locally funded and controlled. New Jersey's reputation for shirking its responsibilities to higher education is improving. There was no state university at all until 1946, when New Jersey took over full responsibility for Rutgers University, now the State University of New Jersey composed of three campuses and a wide variety of colleges and programs. The process of converting the state colleges from strictly teacher-preparation institutes to full-fledged liberal arts schools was begun in the 1960s. Princeton University (1746) is one of the nation's most prestigious private institutions.

Health and welfare. Most of the services rendered the citizens of New Jersey come from the state, although most of the major counties maintain institutions of one kind or another and much funding comes through federal agencies. The state has a shortage of hospitals, prisons, and various other institutions, although much has been done to correct the situation.

Cultural life. New Jersey is well served by the cultural amenities of New York City and Philadelphia, and therefore few were developed in the state. New Jersey helped correct its reputation as a cultural backwater with the opening, in 1968 in Monmouth county, of a beautiful amphitheatre at the Garden State Arts Center, home of the New Jersey Symphony Orchestra. The facility has proved to be a success, and its programming of music, drama, and dance has been well received.

There are several summer theatres in New Jersey, most of them located near vacation areas. The McCarter Theater, on the Princeton University campus, is open all year and offers high-quality plays and musical presentations.

New Jersey has more than 100 nationally listed museums, many of them operated in conjunction with historic sites or buildings. The New Jersey State Museum, which includes a planetarium, is located in the state capitol complex in Trenton. The Rutgers University Art Gallery, the Newark Museum, and the Princeton University Art Museum are among other well-known museums.

Historic sites

New Jersey's rich traditions are manifested in such historic homes and sites as the Rockingham State Historic Site, Washington's winter headquarters near Princeton, where he wrote his farewell address to the Continental Army; Morven, the home of Richard Stockton, signer of the Declaration of Independence, in Princeton; the restored colonial villages of Batsto and Allaire; and the Camden home of poet Walt Whitman. These and other historic sites attract thousands of tourists each year. The state also operates a system of parks, forests, recreation areas, natural areas, and marinas.

HISTORY

The colony. Before the Europeans came, the Delaware (or Lenni Lenape) Indians had long occupied the region. In 1524 the Italian explorer Giovanni da Verrazzano, was the first European to reach New Jersey. Almost a century passed before colonization began with the arrival in 1609 of the English navigator Henry Hudson, who sent a party to explore Sandy Hook Bay. The first permanent European settlement was at Bergen (now Jersey City) in 1660. The colony was brought under English rule in 1664, although for the next five years the Dutch disputed that claim. In 1676 the province was divided into East and West Jersey, the former going to Sir George Carteret and the latter to a group of Quakers. The division continued until 1702, when all of the province reverted to the crown.

Unlike other colonists, who suffered from the harshness of English rule, the early Jerseyans were of such an independent nature that it was the royal governors who did much of the suffering. Until 1738 New Jersey and New York were ruled by a single governor. When Lewis Morris took office as the first governor of New Jersey alone, one member of the Assembly advised his colleagues on how governors should be treated: "Let us keep the doges poore and wee'll make them do as we please."

Revolution and statehood. Considerable division occurred within the state over the American Revolution, and Tory activity was heavy. The most significant battle of the conflict was fought in New Jersey on Dec. 26, 1776. General George Washington and his hungry, ragged troops crossed the Delaware River from Pennsylvania in so-called Durham boats (shallow-draft freight vessels), surprised the garrison of German mercenaries in Trenton, and captured the city. A week later, Washington won another vital battle at Princeton, routing the British forces under Colonel Charles Mawhood. The victories breathed new life into the Revolution, and an army of colonials near despair was transformed into an effective fighting force.

Revolutionary War battles

Between the Revolutionary and Civil wars, New Jersey underwent tremendous industrial development, largely abetted by the construction of canals and, later, railroads. The railroads, in particular the Camden and Amboy (forerunner of the Penn Central), played a crucial role in the state's political life, dominating and controlling legislators and governors during the "robber baron" era of industrial expansion in the 19th century. Accommodating tax laws of that era gave New Jersey the epithet The Mother of Trusts: one-half of the nation's largest corporations made their headquarters in the state. Public dissatisfaction with the power of the trusts and public utilities helped pave the way for the election of Woodrow Wilson as governor in 1910. Industrial growth continued during and after World Wars I and II, and, until the mid-20th century, the growing decay of the cities was overlooked amid the general prosperity brought by industry. (J.Ke./J.McL./P.O.W.)

New York

One of the 13 original U.S. states, New York was until the 1960s the nation's leading state in nearly all population, cultural, and economic indexes. Its displacement by California about the middle of the 1960s was caused by the enormous growth rate that has persisted on the West Coast of the United States, rather than by a large decline in New

York. It remains the second most populous state in the nation, and New York's gross economic product exceeds that of all but a handful of nations throughout the world.

New York is situated across a region of contrast—from the Atlantic shores of Long Island and the skyscrapers of Manhattan through the rivers, mountains, and lakes of upstate New York to the plains of the Great Lakes. With canals, railroads, and highways, New York is a principal gateway to the west from the Middle Atlantic and New England states and a hub for travel to and from much of the nation. The cities of the state—from New York City through Albany (the state capital), Utica, and Syracuse to Rochester and Buffalo on the Great Lakes—and their suburbs are home to more than four-fifths of all New Yorkers.

Both the New England and Southern colonies had a great deal more to do with the movement toward revolution and with stabilizing the new nation during its early decades than did New York, but, once it got under way, New York's growth attained a breakneck pace. The state—and New York City in particular—remains the centre of much of the nation's economy and finance, as well as of many formative impulses in American art and culture.

Role of New York City

The overwhelming presence of New York City has tended to divide the state socially and politically, causing long-standing problems for both the city and state, but the influence and image of the state is a major element in national political life.

The 49,108 square miles (127,190 square kilometres) of New York are bounded, from west to north, by Lake Erie, the Canadian province of Ontario, Lake Ontario, and Quebec province; on the east by the New England states of Vermont, Massachusetts, and Connecticut; on the southeast by the Atlantic Ocean and New Jersey; and on the south by Pennsylvania.

PHYSICAL AND HUMAN GEOGRAPHY

The land. Although New York state and New York City are synonymous to many people, the state has a wide range of geographic and climatic conditions. During at least a part of the Ice Age, most of New York was covered by glaciers; the only exceptions were southern Long Island, Staten Island, and the far southwestern corner of the state.

Relief. The movement of the glaciers left New York with nine distinct physiographic regions. Each has its own characteristic landforms, with distinctive geologic structures and patterns of erosion. In the northeast the Adirondack upland is characterized by the highest and most rugged mountains in the state, reaching 5,344 feet (1,629 metres) at Mount Marcy and 5,114 feet (1,559 metres) at Algonquin Peak of Mount McIntyre. With the exception of some forestry activities, the region has little economic value other than for recreation. A large part of it has been designated as a wilderness preserve by the state.

The St. Lawrence Lowlands extend northeastward from Lake Ontario to the ocean along the boundary with Canada. Within this area are three subdivisions: the St. Lawrence Marine Plain, a flat to gently rolling strip of land along the St. Lawrence River; the St. Lawrence Hills south of the plain; and, farther south, the Champlain Lake Plain.

The Hudson-Mohawk Lowland follows the Hudson River from New York City to Albany and then turns west along the Mohawk River. The Hudson valley, between the Catskill Mountains on the west and the Taconic Range on the east, is from 10 to 20 miles (15 to 30 kilometres) wide; the Mohawk valley is 10 to 30 miles wide. These routes provided access from New York City and New England into the hinterland of New York. Cutting pathways through the mountains of central and western New York, these rivers became the state's avenues of commerce, serving first as the basis of the Erie Canal and later as the route of the New York Central Railroad and of the New York State (Thomas E. Dewey) Thruway.

To the east of the Hudson River lies the New England Upland, extending eastward into Massachusetts and Connecticut and southward across the lower Hudson valley into Pennsylvania.

Two small regions complete the geographic picture in southeastern New York. The Atlantic Coastal Plain, which extends from Massachusetts to Florida, takes in Long Island and Staten Island. A small finger of the eastern Piedmont region juts up from New Jersey for some distance along the west bank of the Hudson.

The Appalachian Highlands

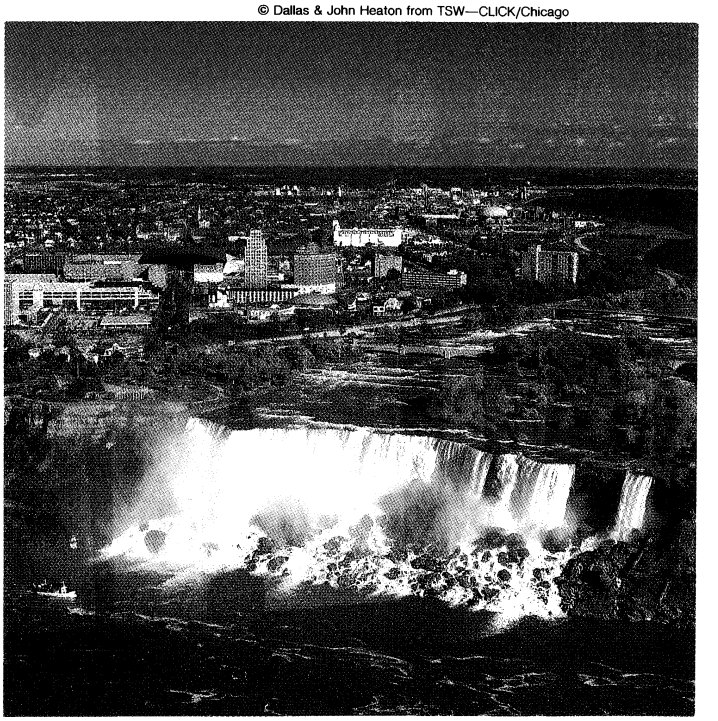
The Appalachian Highlands, the largest region in New York, comprises about one-half of the state, extending westward from the Hudson valley to the state's southern and western boundaries. The Catskill Mountains (the peaks of which reach 2,000 to 4,000 feet), the Finger Lakes Hills area, and the Delaware River basin are located in this region. The Catskills, with their mountains and lakes, are primarily a recreation area. The Finger Lakes area also provides many opportunities for summer and winter sports, and its valleys provide excellent grasslands for dairying. The Delaware basin is a mixed-farming area.

A plateaulike region known as the Erie-Ontario Lowlands lies to the north of the Appalachian Highlands and west of the Mohawk valley and extends along the southern shores of the Great Lakes. It is composed of lake plains bordering the Great Lakes that extend between five and 30 miles inland from the lakes. Because of the moderating influence of the lakes on the weather, the region has become an important fruit-growing area. Between the lake lowlands and the western reaches of the Adirondacks and north of Oneida Lake lies the Tug Hill Upland, which is one of the least settled parts of the state because of its poor soil and drainage and its excessive winter snow conditions.

Drainage. Among New York's special geographic features are its two major shorelines: 127 miles (204 kilometres) bordering the Atlantic and 371 miles (597 kilometres) on Lakes Erie and Ontario. In addition, the state has some 8,000 lakes and nine major rivers. The Hudson and Mohawk rivers have played the most important roles in the state's history, but the Genesee and Oswego, flowing northward into Lake Ontario, also have been important. The Delaware, Susquehanna, and Allegheny drain the southern and western portions of the state and provide a large part of New York City's water supply. The East River connects Long Island Sound with New York Bay and separates Long Island and Manhattan. The most dramatic of the waterfalls that dot the state is Niagara Falls, a source of much hydroelectric power as well as one of the major scenic attractions of the Northeast.

Soils. New York soils can be grouped into categories based on their parent material. One of the most productive groups is found in regions of lime-rich glacial till. Where drainage is good and the terrain not too steep, these soils

© Dallas & John Heaton from TSW—CLICK/Chicago



Niagara Falls, with the city of Niagara Falls, N.Y., to the upper left.

are excellent for agriculture. They occur in a broad belt across the state and into the Hudson valley. Another lime-rich soil group is found in areas that were formerly glacial lake beds, such as the Erie-Ontario Lowlands and large parts of the Hudson and St. Lawrence valleys. Soils of this group are fine-textured and are characterized by level topography. Where drainage is not a problem, these soils are quite suitable for agriculture. Alluvial soils, formed from the sediments of glacial meltwater and the floodwaters of present-day streams, are found in many valley bottoms, especially in the Appalachian Highlands and along the Mohawk and Hudson rivers. Most of Long Island is also covered by alluvial soils, which often have excellent productive potential. Other soils less suitable for agriculture are derived from lime-poor glacial till, such as those north of the major limestone outcroppings in the Ontario Lowland, or from material that is too shallow or coarse, such as those in the rugged mountainous areas of the state or in the sandy region west and north of Albany.

Climate. New York's climate was a great disappointment to the early Dutch settlers. Since Manhattan is actually Mediterranean in latitude, these early settlers were rather bewildered to find that "it freezes and snows severely in winter." If Manhattan was uncomfortably cold and wet in the winter months, the rest of the state must have been an even greater disappointment.

Average temperatures

Average July temperatures range from 77° F (25° C) in New York City to 64° F (18° C) at Indian Lake in the Adirondacks; averages in January range from 33° F (0.5° C) on Long Island to 14° F (−10° C) at Stillwater Reservoir in the Adirondacks. These figures represent the extremes, but there are substantial differences in climate between New York City and upstate Albany, Buffalo, Rochester, and Syracuse. A tendency to cloudiness across the state results in few completely clear days. New York City has about 100 such days each year, Syracuse and Buffalo 72, Binghamton 68, and Albany 73.

Precipitation ranges from 32 to 45 inches (810 to 1,140 millimetres) a year, with the Catskills receiving the greatest amount, while the Erie-Ontario Lowlands receive the least. The region around Buffalo receives an unusual amount of snow because it is on the eastern shore of Lake Erie.

Plant and animal life. More than half of New York state is forested woodland. Some 150 kinds of trees, including such southern species as tulip tree (yellow poplar) and sweet gum, are found in the state. Most woodland, however, is dominated by a small number of northern hardwoods, chiefly beech and sugar maple in association with ash, basswood, cherry, birch, red maple, various oaks, and, occasionally, with conifers such as white pine and hemlock. The spruce–fir association found in extensive parts of the Adirondacks and the largely oak-dominated forests in southeastern New York are the major exceptions to the northern hardwood forests.

Small mammals like the deer mouse, eastern cottontail, snowshoe hare, woodchuck, gray squirrel, muskrat, and raccoon are common. Larger mammals include the white-tailed deer, beaver, and black bear. New York is host to numerous migratory birds. Year-round residents include the eastern meadowlark, American goldfinch, cardinal, eastern bluebird, cedar waxwing, bluejay, several kinds of woodpeckers and owls, the red-tailed hawk, ruffed grouse, mallard, and the common house sparrow, introduced to North America from Europe at Brooklyn in 1850.

Settlement patterns. The cultural and social distinctions among various parts of New York state have diminished.

Diverse influences

Upstate cities, for example, are nearly as ethnically varied as New York City. Certain cultural and social characteristics brought by early settlers remain visible and, to some degree, still influence life-styles. During the colonial period and for a number of years after the American Revolution, New England was a major source of immigrants, and there are traces of the New England influence, particularly in the architecture and small-town planning of the northern shore of Long Island and in northern Westchester county. The Dutch influence around Albany remains in little more than place-names and street names, plus some preserved or rehabilitated Dutch architecture. German and Scottish settlers have left their mark in the Schoharie valley and parts of the Hudson and Mohawk valleys (German), in Orange and Ulster counties, and in the Cherry valley area (Scottish).

The distinction between upstate and downstate is normally along political lines—upstate, conservative; downstate, liberal. Political differences are matched by social differences. Downstate is divided between New York City and the suburbs, and, within the city, differences among the boroughs are important. Although Manhattan has many low-income residents, it is the centre for sophisticated life-styles and liberal politics. In the outer boroughs are relatively stable ethnic neighbourhoods and communities in the process of changing their ethnic or racial makeup; they are more conservative than those in Manhattan but are oriented toward the Democratic Party. The suburbs are dominated by white middle- and high-income families living in detached houses, though the income spread in the suburbs has increased, and the inner suburbs are beginning to resemble the city's outer boroughs.

The rural upstate areas must be distinguished from the upstate cities and their suburbs. Rural New York remains conservative both politically and socially. The city regions vary from relatively sophisticated Rochester, with its heavy concentration of white-collar technical and managerial employees, to the more conservative Syracuse–central New York area. Buffalo, with its emphasis on heavy industry, has a large blue-collar population.

The people. Since the colonial period much of New York's growth has resulted from immigration, both from other states and from abroad. Before the American Revolution the Dutch, English, Scots, and Germans were the primary settlers; they were followed in the first half of the 19th century by New Englanders spreading across developing parts of upstate New York and into Westchester and northern Long Island. The influx of European immigrants came first from the northern and central parts of the Continent and later from southern countries.

Patterns of immigration

Many New Yorkers either are foreign-born or have parents who were born abroad. The primary nations of origin are Italy, the Soviet Union, Poland, Germany, Ireland, the United Kingdom, and Canada, but a significant percentage of people are also from a great diversity of other countries.

Related to the concentration of foreign origin is the state's religious composition. About one-third of the population is Roman Catholic, while more than 10 percent is Jewish.

The growth of the nonwhite portion of the population in the 20th century is significant. The first large-scale influx of blacks from the Southern states occurred during World War I, but it was small compared to the migration that occurred during and after World War II. In 1940 only 4.4 percent of the population was nonwhite, but by the late 20th century the proportion had increased to about 20 percent, concentrated in the state's metropolitan areas and, within those areas, in the central cities.

Puerto Ricans are another immigrant group that has had a significant impact on the economy and culture of New York since World War II. Several hundred thousand reside in the state, mostly in New York City. After a heavy immigration of Puerto Ricans during the 1950s and early '60s, the growing economic strength of Puerto Rico led to a considerable reduction, with those entering the state being largely offset by those returning to Puerto Rico.

Considerable internal migration in the 20th century, far from being random, represents a sorting out of the population, with higher- and middle-income whites moving to the suburbs, leaving low-income whites and blacks within the central cities. Many economic divisions, particularly manufacturing and the headquarters of corporations, have also moved to the suburbs. This movement of people and economic activity has resulted in the urban crisis that is familiar across the United States: there is an increasing need for the cities to combat crime and other symptoms of the poverty concentrated within them, while their social and economic resources to do so are removed. Although the economic strength of the large metropolitan areas of New York is growing, the cities proper are increasingly unable to participate in the prosperity and seem likely to slip still further behind.

The economy. New York's economy is similar to those

of the other Northeastern states. The services sector predominates, though manufacturing is also important. Although the economies of other states are growing more rapidly, New York still has great economic strength. The state has, for example, a complex network of nearly every form of transportation. Its resources of electrical power for domestic and commercial use are enormous, including conventional coal- and oil-burning plants, hydroelectricity from the Niagara region, and a large nuclear capability.

New York is represented in nearly every economic category designated by the federal government. In comparison with the rest of the United States, however, New York's economy has a disproportionately large share of wholesale and retail trade, finance–insurance–real estate, transportation–communications–public utilities, manufacturing, construction, and government. It is, on the other hand, underrepresented in farming and mining.

The state's role in the economy

The state plays both regulatory and promotional roles in the economy. The Public Service Commission controls the rates charged by public utilities, and the Division of Housing and Community Renewal encourages the production of housing and other community facilities. The Department of Commerce aids in attracting new economic activity to the state, providing information and assistance to industries seeking to locate there, giving financial support to local communities interested in developing industrial parks, and offering other incentives to encourage the location of more industries within such areas.

New York tends to have somewhat lower unemployment rates during downturns in the national economy than does the rest of the nation, but it also recovers less rapidly. This is largely a result of the state's economic mix and its heavy dependence on nonmanufacturing activities.

New York has one of the most highly unionized work forces of any state. Unionization has grown rapidly in the service sector among such government employees as teachers, sanitation workers, police, and fire fighters. The nature of labour–management relations varies considerably from industry to industry, with workers in construction and the garment and apparel industries wielding great power. Nearly every session of the state legislature devotes attention to the field of labour relations, particularly public-sector employee relations.

Services and manufacturing. New York has a declining proportion of its population engaged in manufacturing. In 1947 more than one-third of its employed population was in manufacturing, while by the late 20th century the figure had shrunk to about one-fifth.

There is some economic specialization within different parts of the state. Services and finance–insurance–real estate activities are more concentrated in the New York City metropolitan area than in upstate New York. Buffalo is strong in heavy industry, while Rochester dominates the manufacture of photographic and optical equipment and is primarily responsible for the state's strong position in instrument production.

Syracuse ranks high in the state in the production of primary metals, machinery, and paper and allied products, as well as in educational employment. The Utica–Rome area specializes in machinery and primary metals, while the Albany–Troy–Schenectady area is strong in the production of paper and allied products. Albany, as the state capital, leads in government employment. The smallest of the state's metropolitan areas, Binghamton, is the original site of the International Business Machine Corporation (IBM) and therefore has a concentration of employment in the computer and business-machine field.

Agriculture. Dairying is the most important source of farm income, providing more than one-half of the total. Other important sources of farm income are poultry and eggs, livestock products, fruit, vegetables, and field crops. The state raises a variety of horticultural specialties, including nursery products, crops grown under glass, flower bulbs, and seeds, and it competes with Vermont in the production of maple sugar. The fruit and vegetable farms supply the food-processing industry with such products as apples, cherries, peaches, currants, strawberries, tomatoes, peas, beans, sweet corn, and cabbage.

Transportation. A great part of New York's economic advantage is its location on important natural transportation routes and facilities that connect urban centres within and without the state.

Waterways and railways

The Erie Canal, opened in 1825, tied New York City and its port to Buffalo and the westward-expanding nation. The main railroad system followed the route of the canal, with feeder lines that jutted north and south into the remainder of the state. After World War II the limited-access Thomas E. Dewey Thruway stretched from New York to the Pennsylvania state line, passing through Albany, Utica, Syracuse, Rochester, and Buffalo. The basic paths of these main transportation routes are not substantially different from those that were used by the state's original settlers.

With the completion in 1918 of the New York State Barge Canal System, which incorporated the old Erie Canal, New York had the country's most extensive inland waterway system. Although this system is an important means for moving bulk goods—particularly petroleum products, a major share of the tonnage—the annual tonnage carried on the system has dropped considerably.

The railways first challenged the supremacy of the canal as a carrier of goods. Beginning in the mid-19th century with the establishment of the New York Central Railroad, a system was built that tied New York's major cities to Chicago, Boston, Montreal, and other urban centres. Although the number of passengers carried has declined, the railroads remain important handlers of freight. Much of this freight originates at the port of New York, the largest port in the United States, which handles about one-tenth of all the nation's imports. Nearly one-third of all immigrants to the United States pass through this port.

Highways

Central to the highway system are the limited-access highways. The Thruway connects at Albany to the Northway, which extends northward to Canada. In central New York a major highway runs from the Pennsylvania state line to Canada, passing through Binghamton, Syracuse, and Watertown. At Syracuse this route intersects with the Thruway, causing the city to remain a transportation hub and accounting in large part for its economic viability. Another limited-access expressway extends across the southern tier of the state. On Long Island a set of east–west highways ties the island to New York City, New England, and upstate New York.

The New York metropolitan area, with its combination of subways, buses, and railroads, has the most complex commuter system in the nation. The New York transit system, nearly 800 miles (1,280 kilometres) long, provides intracity passenger transport. Commuter railroads serve suburban Long Island, New Jersey, Connecticut, and Westchester county. Many of these transportation networks were brought under the control of a single agency, the Metropolitan Transportation Authority, in 1968.

The three largest airports in the New York City metropolitan area are John F. Kennedy International, La Guardia, and Newark (New Jersey).

Administration and social conditions. *Government.* New York's constitution prescribes the distribution of powers among the branches of state government as well as the system of local government throughout the state. The constitutional framework

However, the document is excessively detailed and includes provisions that most constitutional scholars consider more appropriately treated in legislative statutes than in a constitution. Because of the detail, articles tend to become quickly outdated, necessitating frequent conventions for revision. Since the first convention in 1777, others have been held at roughly 20- to 25-year intervals. The constitution requires that the question of holding a convention be placed before the voters of the state every 20 years.

The first constitution established a bicameral legislature and provided for the first popularly elected governor in the United States, but it restricted the suffrage to male property holders. Veto power was vested in a Council of Revision, comprising the governor, the chancellor of the Supreme Court, and the justices of the Supreme Court, while the appointment of nearly all state and local officials lay with the Council of Appointment, comprising the governor and four senators. These councils represented an effort to avoid the autocratic rule that New Yorkers had experienced at the hands of the colonial governors. The

convention of 1821 abolished the two councils, extended the franchise, and introduced a formal bill of rights.

The convention of 1846, influenced heavily by the populist spirit of Jacksonian democracy, imposed financial limitations on the legislature, which often had extended state credit to private ventures in such areas as railroad and canal building. Many state offices, including the judiciary, were made elective.

The constitution of 1894 has been amended more than 200 times. Its provisions include a merit civil-service system, limitations on disposal of the state's forest preserve, a commitment to public education, and the first constitutional definition of state–local relations.

The state government is led by a strong governor, with power over appointments and budget. The governor is restricted, however, by a number of independently appointed or elected officials. The Board of Regents, for example, which presides over education, is appointed by the legislature. An independently elected comptroller acts as auditor for both state and local governments.

The legislature comprises a Senate of 61 members and an Assembly of 150. Members of both houses are elected for two-year terms. Each house has standing committees concerned with issues of public policy. Several committees composed of both senators and assembly members study policy issues and make recommendations to the legislature. The state also uses numerous nonlegislative commissions—appointed by the governor, by the legislature, or by both—to investigate such problems as education-aid formulas, state–local relations, the judicial process, welfare administration, and governmental organization.

New York is divided into 12 judicial districts. Each district has several elected judges, and together they form the Supreme Court. Four judicial departments act as appeal divisions from the supreme and inferior courts. The highest court is the Court of Appeals. The governor appoints the judges to the appellate departments from those elected to the Supreme Court, and the seven justices serving on the Court of Appeals are appointed by the governor with the approval of the Senate for 14-year terms. The Court of Claims hears cases against the state. There is a variety of local courts, including county courts, family and surrogate courts, and the court system of New York City.

Much legislative debate revolves around the allocation of state aid to local jurisdictions. The constitution has contained a home-rule provision since 1896, but court interpretations of the provision, which gives the state the power to act in any matter in which there is a state concern, have tended to weaken the home-rule concept and continue Albany's domination of local governments. Moreover, the increasing interdependence of the state and its parts caused by urbanization and industrialization inevitably has reduced the autonomy of local jurisdictions.

Local government

The state has 62 counties, which are divided into towns. Urban areas may be incorporated as either cities or villages. Cities are not subject to a town's jurisdiction, but villages remain a part of the town in which they are located and residents pay town as well as village taxes. New York City comprises five counties.

Unlike those states in which either town or county government is weak, New York has strong local governments of both types. This situation often leads to overlapping in the provision of governmental services outside of the cities. Special districts include port and bridge, health, and fire districts, as well as regional market authorities. The Port Authority of New York and New Jersey is one of the largest special districts, operating bridges, harbours, and related facilities throughout the New York City metropolitan area, including those in northern New Jersey.

The decision of the U.S. Supreme Court in 1965 requiring legislative districts to be roughly equal in population brought new life into New York's county government, since town supervisors were no longer able to have dual responsibility as county supervisors. Many counties, including urban counties outside New York City, have opted for single-executive systems.

Cities and villages generally are governed by a mayor and a council; only a few cities, the largest of these being Rochester, use the city manager plan. Some of the larger cities have a second legislative body, often called the Board of Estimate. In New York City the mayor, the president of the city council, the comptroller, and the five borough presidents serve on this body. In other cities, membership usually includes the mayor, the president of the city council, and one or more high-ranking fiscal officers.

The state–local governing system of New York places heavy responsibilities on local governments, and more than one-half of the state budget consists of aid to local government. Most of the aid is for public schools; other allocations include welfare, health, highways, and housing and urban renewal.

To finance such services, New York's relatively healthy economic base provides the source of one of the highest taxation per capita systems in the United States. State impositions include income, sales, business, and excise taxes. Federal funds are a major source of revenue. Local revenues are derived mainly from property and sales taxes. The broad state base plus the widespread use of local sales taxes allows New York to rely less on local property taxes than do other large or heavily populated states.

Systems of taxation

Political life. New York state politics has generally been defined by strong Democratic Party control in New York City and dominance by the Republican Party upstate and on Long Island. Since 1920, both Democratic and Republican governors have held power in New York. During these decades the legislature has tended to be Republican, although since 1975 the Assembly has tended to be Democratic-controlled and the Senate Republican-dominated.

Although the Democrats and Republicans have strong statewide party organizations, New York is one of the few states in which third and fourth parties have thrived and often played important roles in elections.

Education. New York spends more money per pupil for public education than any other state except Alaska. This public school system, with compulsory schooling between the ages of six and 16 (17 in Buffalo and New York City), had its beginnings in the colonial period. Schools were established by churches with government support as early as 1638 in New Amsterdam. It was not until 1791, however, that the state's first public school was established. Some state support was granted in 1795 to elementary schools, and in 1812 a permanent system of public schools was established. Parent-paid fees provided a part of the support until all elementary schools became free in 1867. During the 1850s a few cities established public secondary schools, and during the second half of the 19th century they spread across the state.

The University of the State of New York, one of the most comprehensive educational organizations in the world, was established in 1784 and its governance placed under a Board of Regents. From 1812 to 1904 the educational system in New York was administered by two departments. In 1904, however, the state legislature placed all educational activities under the direction of the Board of Regents. The board selects the state commissioner of education, approves the establishment of new colleges, licenses entry into professions, approves new degree programs, and advises the legislature on all educational issues. Standardized exams used in all secondary schools are called regents' exams. Scores on these exams provide a measure for determining school performance and form the basis for the awarding of a wide range of scholarships.

The Board of Regents

In 1948, public institutions of higher education, primarily teachers colleges and two-year agricultural and technical institutions, plus newly established institutions, were incorporated into the State University of New York, an institution distinct from the University of the State of New York but a part of that larger entity. Before then, private institutions had dominated higher education. Although private institutions in New York enroll a higher proportion of college students than in many states, the state system has been since its founding the fastest growing public institution of higher education in the nation.

The state university system comprises four general types of institution. Major university centres are located at Stony Brook, Albany, Binghamton, and Buffalo. The teachers colleges and several campuses have become general colleges concentrating on undergraduate education but pro-

viding some graduate training. Two-year state institutions and community colleges are supported in about equal parts by the state, the local community, and student fees. City University of New York is supported by the state and by New York City and provides a great variety of programs ranging from those offered by two-year community colleges to graduate education.

Private colleges and universities

There are more than 200 private institutions of higher education ranging in size from a few hundred students to more than 30,000 at New York University. This group includes some of the best-known universities in the nation. Columbia University, founded in 1754 as King's College, is known for the high quality of its graduate instruction and for the national influence of its teachers college. Cornell, the base for the agriculture, human ecology, veterinary medicine, and industrial- and labour-relations units of the State University, is a member of the Ivy League, as is Columbia. Fordham and St. John's are perhaps the best-known of the state's many Roman Catholic colleges and universities. The University of Rochester, known for its programs in music and the natural sciences, and Syracuse University, home of the Maxwell School of Citizenship and Public Affairs, the first university unit established for training students for public service, are also well-known private institutions. Other well-regarded institutions include Colgate, Hamilton, Union, St. Lawrence, Bard, Skidmore, Barnard, and Vassar.

Health and welfare. The life expectancy of New York state residents is shorter than the national average, but the state has some of the finest hospital and medical education facilities in the United States. The great majority of New Yorkers are covered by hospital and surgical insurance. Social welfare is a major enterprise in New York state. State benefits range from aid to dependent children to the support of nursing homes. State residents also participate in the federal food stamp program.

Culturai life. Much of the style and tone of life in the United States is set in New York City, which remains the artistic, cultural, and economic capital of the nation. The fashion industry is headquartered in its garment district. The main live theatre in the nation is located on and off Broadway; many television programs originate in New York City, where the three major networks have their home offices, and many motion pictures are filmed on its streets. The city's museums—particularly the Metropolitan Museum of Art, the Museum of Modern Art, and the American Museum of Natural History—set the pace for similar institutions across the country.

Several major publishing houses have their headquarters in New York City, as do a large number of national magazines. Serving their needs and those of the central offices of many of the country's largest corporations are banks, public-relations and advertising firms, management consultants, and legal firms. Because of this concentration of business and culture, New York City maintains a leading national position in many aspects of American life.

Regional museums and institutions

Cultural activities are not confined to New York City, however. Many art museums are located in the state's large and small cities. Among them the Albright-Knox Art Gallery, in Buffalo, has collections of contemporary paintings and sculptures. In Rochester are the Memorial Art Gallery of the University of Rochester, the Rochester Museum and Science Center, the Strong Museum, and the International Museum of Photography at the George Eastman House. The Everson Museum of Art of Syracuse and Onondaga County is considered an outstanding example of modern architecture, while the city's Erie Canal Museum is devoted to canal history. The New York State Museum in Albany is the oldest and largest state museum in the United States. Symphony orchestras outside New York City include those of Buffalo and Rochester, while the Eastman School of Music at Rochester is internationally known. Fine architecture is found across the state, and the performing arts are pursued by professional and amateur groups. The cultural life of the state's many college and university towns often is centred on these institutions.

The Saratoga Performing Arts Center in Saratoga Springs is the summer home of the Philadelphia Orchestra and the New York City Ballet. Theatrical performances also are held at this modern cultural centre. The Chautauqua Institution, located on Chautauqua Lake in southwestern New York, was founded in 1874 as a training school for Sunday-school teachers. The name Chautauqua has since been adopted to include a wide range of cultural and educational activities, including concerts, opera, drama, and lectures.

Cooperstown, founded by the father of James Fenimore Cooper, is known as the village of museums, the best known of which is the National Baseball Hall of Fame and Museum. At the Angel Moroni Monument on the Hill Cumorah near Palmyra, an annual pageant depicts the founding of the Mormon church. South of Palmyra are located the Greyton H. Taylor Wine Museum in Hammondsport and the Corning Glass Center in Corning. Historic homes, forts, and battlefields dot the state. More than one-third of all the battles of the American Revolution, including the Battle of Saratoga, were fought in New York state.

New York was the first state in the Union to establish a program for continuing financial support of the arts. The Council of the Arts, which administers the program, funds organizations in the fields of the performing arts, visual arts, film and media, and special programs.

Recreation

The variety of New York's geography provides not only great beauty but also vast opportunities for recreation, relaxation, and a study of the past. With the cool summers of the Adirondacks, the snowy slopes of the Catskills, the ocean beaches and lakes, and an almost unlimited repertory of water sports, New York state has a broad recreational base.

HISTORY

Two major groups of Indian tribes were living in the New York region when Europeans first arrived: the Mohegan (Mohican) and Munsee tribes of the Algonquian family near the Atlantic coast and, farther inland, the five tribes of the Iroquois—Mohawk, Oneida, Onondaga, Cayuga, and Seneca—which formed the Iroquois League in about 1570. This confederacy of Indian tribes, with its advanced social and governmental institutions, reached the height of its power in about 1700. The alignment of these tribes with the British against the French and the Algonquian probably enabled the British to emerge as victors in the nearly 150 years of struggle between the two European powers in North America.

Settlement and colonial period. New York was settled originally as a colony of the Netherlands following Henry Hudson's exploration in 1609 of the river later named for him. In 1624 the Dutch established Fort Orange at present-day Albany as the first permanent European settlement in New York. One year later New Amsterdam was established at the foot of Manhattan Island. To legalize the settlement, Peter Minuit, the Dutch governor, paid the Indians 60 Dutch guilders—about $24—in the form of merchandise. Although the Dutch established several settlements along the Hudson, their interest was more in trade than in permanent agricultural development. Thus, while the trading posts prospered and aided the general expansion of the empire of the Netherlands, no deep roots of permanent colonization were planted in New York. The likely explanation for this lies in the economic prosperity and social stability of the homeland. The Dutch citizens had no strong economic motivations to move overseas, nor were there sufficient religious quarrels at this time to promote any such movement. When an English fleet sailed into New York harbour in 1664, Peter Stuyvesant, the governor, surrendered without a fight. Although controversy ensued for several years, the colony was clearly in English hands by 1669. Under the English it was renamed New York after the Duke of York.

Slowness of growth in the colonial period

Despite this change in ownership and sovereignty, however, the colony did not develop rapidly. Like the Dutch, the English crown granted large tracts of land to private individuals. This system of landownership was not very attractive to settlers such as the farmer-colonists who had settled the New England area, and agricultural development, particularly in the areas along the Hudson valley, remained slight.

The European war between France and England had its counterpart in North America. The French, established along the St. Lawrence and in Quebec, made a number of forays into northern and central New York. The strong Five Nations federation of the Iroquois aligned itself with the English in New York and New England because of aid given earlier by the French to the rival Algonquian. This warfare discouraged settlement beyond Albany. The military situation was brought to a conclusion in 1763 by the Treaty of Paris, which confirmed English dominance of the New York region. A gradual but steady movement of settlers from New England was the beginning of New York's population explosion. The New Englanders moved across the borders of Connecticut and Massachusetts, some remaining on the east bank of the Hudson, others passing through Albany to the interior.

In 1698 the colony's population was about 18,000, two-thirds of it concentrated in and around New York City. By the eve of the American Revolution, it had grown to 163,000, with the concentration nearly exactly reversed, but New York still ranked only seventh among the American colonies. Dutch culture remained strong in New York City and in Albany, while most of the settlements in the interior had a flavour and dialect of the New England Yankee; there were also several German communities. This emerging pattern of cultural heterogeneity was later to have a considerable influence on the politics of the state, as were the waves of immigration from Europe that followed the war and continued well into the 20th century.

Revolution, statehood, and growth. New York contains many of the battlegrounds of the American Revolution. The war in New York took on many of the characteristics of a civil war, since the area probably had a higher proportion of residents who were loyal to the crown than did any other colony.

Following the war a part of New York's leadership aligned itself with leaders from other colonies to urge establishment of a strong central government for the new nation rather than the loose confederation that was then in power. New York delegates participated vigorously in the Constitutional Convention, one of the leaders of which was Alexander Hamilton. Despite the role played by Hamilton and by other New York delegates in drafting the Constitution, the politics of ratification within the legislature were intense and bitter, and New York was the 11th state to endorse the U.S. Constitution.

The American Revolution and the War of 1812 temporarily interrupted New York's expansion to the west, but thereafter the movement began in earnest. Turnpikes spread westward from Albany and from other locations up and down the Hudson River, and settlers moved across the state. The opening of the Erie Canal in 1825 confirmed New York's position as the gateway to the west from the Atlantic Coast. The railroads followed in quick order and tended to follow the pattern of trade that had been established earlier by the turnpikes and the canal.

Improvements in transportation

According to the census of 1800, New York state had become the second largest in the Union, trailing only Virginia; 10 years later it had surpassed all other states. Its leadership was not only in population, size, and growth but also in the areas of manufacturing, trade, and transportation—and in the increasing heterogeneity of its population.

Growth and change also were reflected in the political and governmental history of the state. The original state constitution restricted suffrage to property holders and established a governing system that was dominated by large property holders and leading commercial interests. The change in population composition, as well as shifting political attitudes in the nation, soon caused New York to move in a more democratic direction. During the 1830s a vigorous campaign was launched against the system of landownership in the Hudson valley, with renters eventually being given the opportunity to own the land they tilled. The constitutional convention of 1846 confirmed these democratic moves by expanding suffrage and restricting the power of both legislature and governor.

Emergence of political divisions. New York continued to grow in virtually every dimension, but its political development became centred on the increasing chasm of interest and affection between New York City and upstate New York. The issue of home rule, the demands of the city for total powers of self-government, has remained central to the conflict.

During the 1780s an organization, eventually to be known as Tammany Hall, was formed in the city to combat attempts by propertied Revolutionary leaders to limit the franchise. Largely middle-class in membership, it did not extend its democratic principles to the lower classes or to the immigrants. By the mid-19th century, however, through workers' and equal-rights parties, Irish politicians came to dominate the Tammany organization and the office of mayor; the trend culminated in the control of the Democratic machine after 1868 by William "Boss" Tweed.

Well into the 20th century, the name Tammany was an international byword for municipal corruption at the highest levels. City–state antagonism was fueled by Democratic domination in the city and Republican domination of the areas upstate and, in most years, of the statehouse and the legislature as well. Investigations of Tammany Hall and city politics in general were highlighted by those of the Seabury Commission (1931–32), which brought about the resignation of Mayor James J. Walker and led to the reform administration of Mayor Fiorello H. La Guardia (1933–45) and the efforts of subsequent mayors to tread the line between the power of Tammany Hall in municipal elections and an image of political incorruptibility.

City–state conflicts

Much of Tammany Hall's power was based on its social services to the waves of immigrants that had inundated New York City until changes in immigration laws slowed the tide during the 1920s. When the state and federal governments began to take over such services as social security, workmen's compensation, unemployment, welfare, and health benefits, notably during the depression of the 1930s, Tammany's hold began slowly to erode.

Recent trends. Since World War II, New York's social and educational services have increased dramatically, while its industrial base has eroded. This has created a difficult financial situation for both the state and New York City, with the latter barely avoiding bankruptcy in 1975. While responsibility remains with the state to provide a growing number of services, a succession of governors and legislative leaders have been successful in gradually reducing taxes and broadening the state's economic base.

(A.K.Ca./P.J.Sc.)

Pennsylvania

One of the original 13 American colonies, the Commonwealth of Pennsylvania has long been one of the most populous states in the nation. Geographically, it is the Keystone State, integrating first the older states of the Northeast and the South and later the states of the East and the developing territories and states of the Midwest. The state is polarized by two great metropolitan areas: Philadelphia lies athwart the vast population belt stretching along the seaboard from Maine to Virginia, and Pittsburgh is the beginning of the booming industrial belt reaching westward across the Great Lakes plains to Chicago and Milwaukee. The state capital, Harrisburg, nestles in the foothills of the Appalachian Mountains.

The Keystone State

The 45,308 square miles (117,348 square kilometres) of Pennsylvania are bounded on the north by Lake Erie and New York; on the east by New York and New Jersey; on the south by Delaware, Maryland, and West Virginia; and on the west by the panhandle of West Virginia and by Ohio. Although it is classified as a Middle Atlantic state, it does not touch the Atlantic Ocean at any point. Water, nonetheless, has been nearly as crucial in the state's growth as has the wealth of its earth. The boundary with New Jersey is formed by the Delaware River, on which Philadelphia is the major element in the Pennsylvania–New Jersey–Delaware harbour complex that is one of the world's busiest shipping centres. In the northwest a small panhandle separates Ohio and New York and forms a 40-mile (64-kilometre) waterfront on Lake Erie, giving the state access to the iron-ore barges and other commerce of the Great Lakes. Located at the point where the Allegheny and Monongahela rivers meet to form the Ohio

River, Pittsburgh has become one of the nation's busiest inland river ports.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The mountains and rivers that dominate Pennsylvania's natural landscape have had a major influence in determining its use. The Allegheny section of the Appalachian Mountain system, running across the state from southwest to northeast, leaves only a triangle of Piedmont and Coastal Plain in the southeast, with Philadelphia and environs at its extreme point. The Allegheny Front and Mountains form a diagonal spine across the centre of the state, while to the north and west a rugged plateau region falls almost at the lakefront into the Lake Erie lowlands, known as the Lake Erie Plain.

Surface features

In the eastern part of the state, the Delaware is fed mainly by the Lehigh and Schuylkill. The Susquehanna, draining the largest section of the state, is a wide, shallow stream that meanders finally into Maryland and Chesapeake Bay. The Ohio River is formed by the confluence of the Allegheny and Monongahela rivers at Pittsburgh, from where it flows westward to the Mississippi. Minor systems lead into Lake Erie in the northwest and the Potomac from the southwest.

Pennsylvania takes in parts of seven physiographic regions. Inland from the fertile Coastal Plain around Philadelphia, the Piedmont is an area of rich limestone soils and farms that are the mainstay of the state's agriculture. The Reading Prong, part of the New England Upland, is a small tongue of a ridge in the northeast. In the south central region, the northernmost finger of the Blue Ridge extends into the state; its topography was important in the Battle of Gettysburg during the American Civil War.

The Appalachian Ridge and Valley region, running southwestward across the state, to the west of the Piedmont, comprises a series of parallel ridges and valleys, including the long and wide Lehigh, Lebanon, and Cumberland valleys, which were important in America's early westward movement. From this rises the Appalachian Plateau, whose rough-and-tumble aspect of small valleys and broken, ragged ridge lines covers more than one-half of the state, from the southwestern and western borders to the Delaware in the northeast and on into New York. Plateaus up to 2,000 feet high cover northern Pennsylvania. Mount Davis, at 3,213 feet (979 metres), is the state's highest point; its landscape is dotted with coal mines and mining towns. The sandstone soils of the mountains and plateaus are poor agriculturally. In the northwestern bump lie the Lake Erie lowlands, a flat area of gravelly but fertile soils that once lay at lake or ocean bottom and now is a producer of grapes, wine, and potatoes.

Climate. Pennsylvania has a continental climate that is characterized by wide fluctuations in seasonal temperatures, with prevailing winds from the west. The frost-free period lasts the longest in the southeast, the Ohio valley, and the Erie lowlands. Higher areas are free of frost from three to five months a year. Coupled with some 39 inches (991 millimetres) of rain annually, this provides adequately for the cultivation of temperate-zone crops and vegetation. In the north, average temperatures range from about 25° F (−4° C) in January to about 69° F (21° C) in July. In the southeast and western portions of the state, averages are several degrees higher.

Plant and animal life. Pennsylvania marks the transition zone between the northern and southern forests of North America. About one-half of the state remains wooded, although only small areas are virgin forest. Hickory, locust, maple, cherry, beech, black walnut, elm, chestnut, poplar, ash, sycamore, willow, linden, white and yellow pine, and other conifers are all common in various parts of the state. The southeast has some trees common to the Mississippi valley.

Pennsylvania's abundant wildlife makes it a leading state for hunting. Unprotected predatory species include red and gray foxes, wildcats, and weasels. Most large animals gradually have been eliminated by humans. Species once common—*e.g.*, the black bear and the Canada lynx—are now rare, although skunks and raccoons, pheasants and other game birds, and, in some areas, deer are still found in abundance. More than 150 species of fish, including trout, salmon, and walleye, are found in the streams.

Settlement patterns. The Pennsylvania landscape under human settlement is as varied as its physical geography. Although nearly all the land is forested or under cultivation, Pennsylvania is considered an urban, industrial state. Urban areas in the state, however, comprise widely separated enclaves.

The rural areas of Pennsylvania, especially the farmlands, lie mostly in the southeastern counties. In the southwest and in the Susquehanna River valley, pasturelands dot the landscape. Generally, several small villages, each consisting of a general store, a church, a gasoline station, and a few houses, are located within a short distance of a larger town that serves the surrounding farms. In the mining regions the town structure was once completely dependent on the mines, but company ownership of stores and miners' homes is no longer prevalent.

Rural and urban Pennsylvania

For the most part, the urban centres of Pennsylvania developed along with specific industries. Reading is known as a textile and hosiery town, Hershey as a chocolate town, Pittsburgh and Bethlehem as steel producers, Erie as a Great Lakes port, the Wilkes-Barre and Scranton areas as centres of the coal industry, and Philadelphia as a centre of finance, industry, and commerce. In turn, the city influences the surrounding area as a supplier of goods and as the determiner of artistic and cultural life.

The people. Four major Indian groups with tribal settlements occupied Pennsylvania at the time of the European incursion. These included the Delaware, or Lenni Lenape, in the Delaware River basin; the Susquehanna along the Susquehanna River in Pennsylvania and Maryland; the Shawnee along the Susquehanna and near the present site of Easton; and various segments of the Iroquois League. The Indians were a powerful force in the early history of Pennsylvania, but all were gradually pushed to the north or west as European settlers prospered. Only about 11,000 Indians remain today in scattered locations.

Although the Swedes and the Dutch arrived in Pennsylvania first, the English Quakers were by far the most important of the early groups that settled along the Delaware River. With a secure hold on the region, the Quakers quickly turned the Delaware counties—especially William Penn's three original counties of Philadelphia, Chester, and Bucks—into a thriving commercial region.

Penn's policies of religious toleration and his experiments in democratic forms of government encouraged other groups to settle in Pennsylvania in large blocs. Germans from the Rhineland settled in the inland counties of Lancaster, Lehigh, Berks, and Northampton. By the time of the American Revolution, the Pennsylvania Dutch, as they had become known, constituted one-third of the colony's population. They turned their land into one of the richest and most productive farming areas in the world and made their customs, cooking, and artwork famous.

Beginnings of ethnic diversification

Small subgroups of the German community are important components of Pennsylvania's population. Largely religious, they include the Amish, Mennonites, Moravians, Schwenkfelders, and Dunkers. The Amish have attracted special attention because of their old-fashioned dress and educational methods and their preference for living apart and farming without modern machinery.

The third major ethnic group to settle in Pennsylvania were the Scotch-Irish. Pressing west past the areas of English and German settlement, this group settled along the western frontier during the middle of the 18th century. Along with the true Scots the Scotch-Irish composed about one-fourth of the entire population of the colony at the outbreak of the Revolution. In addition to the three main groups listed above, there were small numbers of French Protestants, or Huguenots; Welsh; Cornish; Irish; and, of course, the Dutch and the Swedes. Through the early years of the 20th century, immigration to Pennsylvania continued to be largely German and Irish.

Until the 1860s, there was open land on the western frontier of the state. During this period the religious character of the commonwealth changed: the Quakers and the Pennsylvania Dutch lost numerical strength to predomi-

nantly Roman Catholic immigrants, and Presbyterianism and Methodism gained footholds in the west. During the 20th century the most important European immigrants to enter the state have been of southern and eastern European origin, particularly Italians and Slavic peoples, who provided manpower for the state's burgeoning industry and manufacturing. During and after World War II, in response to the needs of the economy, black Americans from the rural South began to settle in Pennsylvania in increasing numbers. For the most part, the newly arrived black Americans, like the earlier Italians and Slavs, have tended to settle in urban, industrial areas.

Although Pennsylvania is a melting pot of many ethnic groups, the Quakers and the Pennsylvania Dutch continue to play roles in the culture and tradition of the state out of proportion to their actual numbers.

The greatest population density is found in Philadelphia, Delaware, and Allegheny counties, the latter including Pittsburgh. In 1790, when the first federal census was taken, 89.8 percent of the population was classified as rural. In 1900 the urban population, at 54.7 percent, outstripped the rural for the first time.

The economy. With fertile farmland, large areas of commercial forest land, seemingly inexhaustible supplies of coal, many navigable waterways, and an economically strategic location on the Eastern Seaboard, Pennsylvania has historically been one of the most prosperous states in the nation. As the exploitation of its primary wealth—soils, minerals, and forests—proceeded during the 18th and 19th centuries, Pennsylvania held a place that was second only to New York in wealth and population.

Changes to the economy

Although its steel industry began to contract in the mid-1960s, Pennsylvania still produces more specialty steel than any other state, and its manufacturing industries are as diversified as those of any state in the Union. In the 20th century, however, the changing nature of modern industry, the use of fuels other than coal, and the growth of the rest of the country have had an adverse effect on the Keystone State both in production and in population, and there are pockets of poverty and unemployment.

In 1956 the Pennsylvania Industrial Development Authority was set up to assist in the redevelopment of areas plagued by chronic unemployment. In 1987, the Pennsylvania Economic Development Partnership was created to direct the development of the state economy through both public and private channels.

Mining. The coal industry of Pennsylvania has declined steadily since World War II, but it remains important nationally. Pennsylvania's reserves of anthracite coal are the largest in the nation. The state produces only a tiny amount of the nation's crude oil, but natural gas deposits are a major asset. Pennsylvania is among the nation's highest-volume producers of power from nuclear energy, despite an accident at the Three Mile Island nuclear power plant, near Harrisburg, in 1979. Small deposits of iron ore have been discovered, along with limestone, silver and gold, copper, cobalt, zinc, and salt.

Industry. Pennsylvania's first ironworks, for the production of wrought iron, was built in Berks county in 1716. The eastern region long was the centre of production, especially after large-scale exploitation of its anthracite fields began. The Pittsburgh area became increasingly important as bituminous coal and coke replaced anthracite and as the Great Lakes became the route for barges from the rich iron ore beds around Lake Superior. Western Pennsylvania, however, is at a geographic disadvantage when compared to the Great Lakes–Midwest steel region, and its older plants are less efficient than those of the Southern United States and some foreign countries.

Pennsylvania has always been known for the variety of its manufactures. It is a leader in the cement industry and produces many clay products. In terms of value added by manufacture, the once dominant iron and steel industry has been surpassed by food products, chemicals (primarily pharmaceuticals), machinery, and electrical and electronic equipment. Industry is unevenly distributed, its greatest concentration being in Philadelphia, Allegheny, and Montgomery counties.

The retail sales industry is also important. John Wanamaker opened the nation's first department store in Philadelphia in the 1870s. Frank Woolworth opened his first successful five-and-ten in Lancaster in the same period. Other famous Pennsylvania retailers include S.H. Kress, S.S. Kresge, G.C. Murphy, J.J. Newberry, and W.T. Grant, all of whom founded chains bearing their names.

Agriculture. Pennsylvania has the largest rural population in the United States, and nearly one-third of the state is under cultivation. The Pennsylvania Department of Agriculture, various vocational agricultural programs, and farm and commodity organizations offer assistance to the state's farmers. Livestock, including dairy and beef cattle, hogs, and sheep, and livestock products are the major components of farm income. Pennsylvania is a major producer of milk, eggs, and poultry; fruits, including peaches, grapes, cherries, and apples; hay; corn (maize); mushrooms; and Christmas trees. Ice cream and sausage are major processed food products.

Major farm products and crops

Transportation. Pennsylvania's three major ports—Philadelphia, Pittsburgh, and Erie—are supplemented by others along the Delaware, the Susquehanna, the Allegheny, the Monongahela, and the Ohio. Petroleum is Philadelphia's leading cargo, but it has facilities for handling ores, coal, molasses, and general cargo as well. Part of Erie's facilities have been converted to the building of ore carriers for Great Lakes service.

Pennsylvania's highway system, one of the most extensive in the nation, includes the Pennsylvania Turnpike, a four-lane toll road joining New Jersey and Ohio that was a model for the nation in modern superhighway construction. It is paralleled to the north by Interstate 80, known in Pennsylvania as the Keystone Shortway.

Although the bankruptcy of the Penn Central Railroad in 1970 reduced passenger and freight service, the rail lines continue to serve as important economic links for the Middle Atlantic region. Philadelphia and Pittsburgh have the state's largest airports, but others also have international status.

Administration and social conditions. *Government.* Under the constitution of 1968 and subsequent amendments, the executive branch comprises the governor and lieutenant governor, the attorney general, the auditor general, the state treasurer, and the members of the cabinet. The secretary of state and the secretary of education are appointed by the governor, subject to Senate approval.

The governor is elected for a four-year term and may be reelected for one additional term. Among the main powers of the governor are the right to return bills for reconsideration and to veto portions of appropriation bills. The General Assembly is made up of a Senate of 50 members and a House of Representatives of 203 members. Senators are elected for four-year terms and representatives for two-year terms.

Functions of the executive

A unified judicial system comprises the Supreme Court, Superior Court, Commonwealth Court, and a modern system of lower courts. The seven justices of the Supreme Court are elected for 10-year terms; justices of the peace and of the municipal and traffic courts of Philadelphia are elected for six-year terms.

Under the constitution, local government is provided for by the General Assembly, which classifies cities, counties, boroughs, school districts, townships, and special authorities by population. All Assembly provisions must be uniform throughout each class, and no legislation may be directed toward one locality. As the only first-class city, however, Philadelphia becomes the target of legislation aimed at this class. Along with Pittsburgh and Scranton, which are second-class cities, it has a strong-mayor government. Smaller cities have several governmental forms, while boroughs have elected councils and weak mayors.

Pennsylvania has two classes of townships and five classes of school districts. Local authorities are special units that were set up in 1933 to circumvent constitutional restrictions preventing local units from incurring debt. The cities, boroughs, and townships, as governmental units, carry out the usual public services. A constitutional amendment in 1968 allowed local governments to opt for home rule.

Between the Civil War and the 1920s, state government was, with a few breaks, in the hands of a succession

of Republican administrations that made the statehouse a seat of "boss rule." A similar dynasty persisted in Philadelphia until the late 1940s. Since the mid-20th century the membership of the General Assembly has usually been about evenly divided between Republicans and Democrats. Pennsylvania generally exhibits the characteristic American pattern of rural conservatism and urban liberalism, although locally such labels may bear no necessary relation to particular parties.

Education. Public school districts provide physical facilities, teachers, and textbooks and levy taxes and issue bonds. The state Department of Education oversees statewide standards for teacher certification and curricula and apportions money to local school districts.

Institutions of higher learning

Pennsylvania is the home of more than 100 colleges and universities in addition to a state system of two-year community colleges. Pennsylvania State University, in University Park, is the publicly supported land-grant institution; it has 21 branch campuses throughout the state. The University of Pennsylvania, in Philadelphia, was originally a charity school at its founding in 1740 but was transformed by Benjamin Franklin and others into an academy. In 1765 it opened the first medical school in North America. Today, among its most notable divisions are the Wharton School of Finance and Commerce and the University Museum, a leading sponsor of archaeological expeditions and research throughout the world. Though privately endowed, it receives considerable state aid.

State aid is given also to Temple University (1884), in Philadelphia, whose programs are geared to its urban setting, and to the University of Pittsburgh (1787). Philadelphia is a major centre of medical education, while in Pittsburgh the Carnegie Mellon University, formed in 1967 by the merger of the Carnegie Institute of Technology (founded in 1900 as the Carnegie Technical School) and the Mellon Institute (1913), makes that city a centre of scientific studies. Other schools with major reputations are Bryn Mawr College (1880); Haverford College (1833); Swarthmore College (1864), a Quaker school; and Villanova University (1842), a Roman Catholic institution—all near Philadelphia. Dating from the 18th century are Dickinson College (1773), in Carlisle; Franklin and Marshall College (1787), in Lancaster; and Washington and Jefferson College (1787), in Washington. Carlisle was the site of the Carlisle Indian Industrial School from 1879 to 1918 and is now the home of the U.S. Army War College.

Health and welfare. More than one-half of Pennsylvania's budget is allotted to education, welfare, and highways. Hospitals and centres throughout the state deal with problems of chronic illness, old age, mental retardation, and the like, and local governments are assisted in welfare payments in a broad range of categories. Urban renewal programs also receive state and federal funding. Pennsylvania created a Department of Aging in 1979, and a new Department of Corrections was established in 1984.

Cultural life. In colonial times, Philadelphia was the focus of the nation's intellectual, cultural, and political life. As Pennsylvania grew and prospered, Pittsburgh and other, smaller cities also became centres of the arts.

Two of the nation's major symphony orchestras are located in Pennsylvania. The Philadelphia Orchestra, under such conductors as Leopold Stokowski and Eugene Ormandy, has become world famous, as has the Pittsburgh Symphony Orchestra, under such conductors as Victor Herbert, Fritz Reiner, and William Steinberg. Philadelphia's Academy of Music provides a home and concert hall for its orchestra, and the world famous Curtis Institute of Music, founded in 1924, is one of the world's leading conservatories. In addition, many community orchestras perform throughout the state. Also notable is the religious music of the Moravians. At their cultural centre, Bethlehem, the Bach Choir's Bach Festival attracts music lovers every May from many states.

The visual arts: painting and sculpture

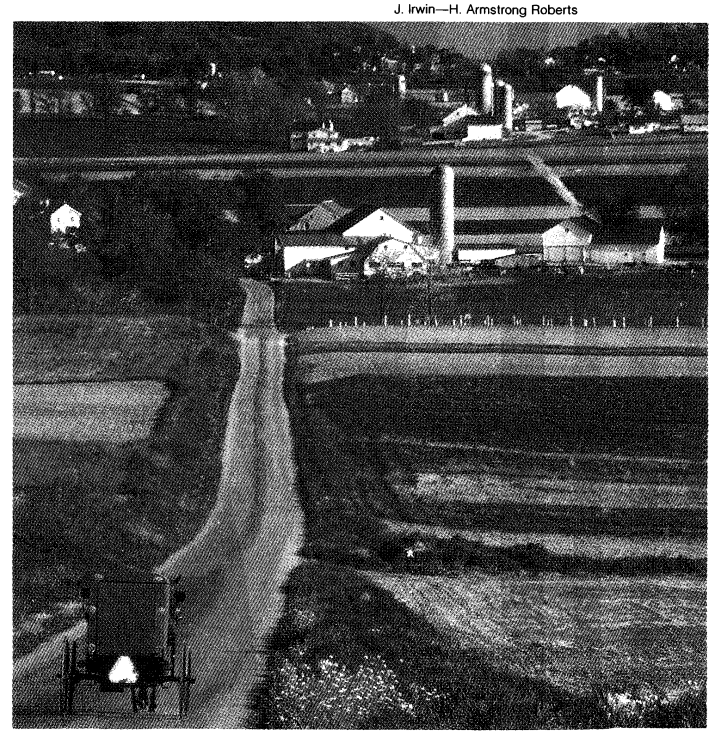
Philadelphia is the home of one of the world's finest art museums, the Philadelphia Museum of Art, and of the Rodin Museum. The Pennsylvania Academy of the Fine Arts offers a base for teaching and study as well as for display. The Carnegie Institute's museum and library in Pittsburgh also houses many famous paintings. Among famous artists whose homes were in Pennsylvania are Charles Willson Peale and his two sons Raphaelle and Rembrandt, Benjamin West, Mary Cassatt, Thomas Eakins, and the Wyeths, N.C. and his son Andrew. The Calder family of Philadelphia is famous in sculpture. Alexander Milne Calder's giant statue of William Penn stands atop City Hall, and his grandson, Alexander Calder, gained renown for his free-form mobile sculptures.

Philadelphia long vied with New York City as the theatrical capital of the young nation, though today it is basically a tryout town for shows on their way to Broadway. Summer theatres have proliferated across the state, especially in the many resort areas. Temple and Carnegie Mellon universities offer major programs in theatre. The Pennsylvania Ballet gives performances within the state and also maintains an active touring schedule. Among well-known writers of the 20th century with Pennsylvania origins are Pearl S. Buck, Rachel Carson, James Michener, Christopher Morley, John O'Hara, Mary Roberts Rinehart, and John Updike.

In addition to its art museums, Pennsylvania enjoys many other museums. The Franklin Institute in Philadelphia is a museum of applied science. The William Penn Memorial Museum and Archives Building in Harrisburg, the Pennsylvania Farm Museum of Landis Valley near Lancaster, the Annie S. Kemerer Museum in Bethlehem, and the Mercer Museum of the Bucks County Historical Society (crafts and craftsmen's tools) in Doylestown house fine collections. There are also small museums in Allentown, Reading, Scranton, and Greensburg and historical society museums in Lancaster, York, Reading, Wilkes-Barre, and West Chester, to mention a few. The Pennsylvania Historical Commission, established in 1913 and known since 1945 as the Pennsylvania Historical and Museum Commission, is active in conservation of the state's historical heritage; the Philadelphia Historical Commission oversees that city's many historic shrines.

Pennsylvania has retained strong elements of folk culture among its diverse ethnic groups. The Plain People—the Amish, the Mennonites, and other small sects—have kept their old customs and dress and often resist the use of machinery. In Lancaster county, where the Amish still farm with horses and oxen, their wagons are a familiar sight. The folk art and the cooking of the Pennsylvania Dutch are famous, and their brightly painted hex signs adorn their large barns. At Czestochowa, in Bucks county, the

J. Irwin—H. Armstrong Roberts



Amish horse and buggy in Lancaster county, Pa.

largest Polish church in America is evidence of the area's Polish population.

Public figures such as Benjamin Franklin were responsible for the emergence of newspapers, magazines, and libraries in the colony. Later, the Lippincott and the Curtis publishing families became giants in the publishing world, the Lippincotts as book publishers and the Curtis family as publishers, until recently, of such magazines as *The Saturday Evening Post, Ladies Home Journal,* and *Country Gentleman.* The *Farm Journal,* a Philadelphia publication, is one of the nation's leading farm magazines. Several other major publishers are also headquartered there.

Historic eminence in publishing

All of Pennsylvania's cities and larger towns have newspapers, and many ethnic newspapers are published throughout the state. The major cities have their own television stations, and most local communities have their own radio stations. Station KDKA, the first commercial broadcasting station in the world, opened in Pittsburgh in 1920.

HISTORY

Swedes were the first European settlers in the area that was to become Pennsylvania. Traveling up the Delaware from a settlement at the present site of Wilmington, Del., Governor Johan Printz of the colony of New Sweden established his capital on Tinicum Island in 1643, within the boundaries of modern Pennsylvania. Other Europeans, principally the Dutch, established trading posts within Pennsylvania as early as 1647, although the Swedes remained at Tinicum until 1655. In that year, rivalry and fighting between the Dutch and the Swedes led Peter Stuyvesant, governor of New Netherland, to seize New Sweden. Dutch control of the region ended in 1664, when the English seized all of New Netherland in the name of the Duke of York.

The Quaker colony. In March 1681 King Charles II of England signed a charter giving the region to William Penn in payment of a debt owed by the king to Penn's father, Admiral Sir William Penn. The charter, which was officially proclaimed on April 2, 1681, named the territory for Admiral Penn and included also the term *sylvania* ("woodlands"), as the younger Penn requested.

William Penn intended that the colony should provide a haven of religious tolerance for his fellow Quakers. While still in England, he drew up the first of his "frames of government" and sent his cousin, William Markham, to establish claim to the land and also to establish the boundaries of what became the city of Philadelphia. Penn arrived in 1682 and called a General Assembly to discuss the first Frame of Government and to adopt the Great Law, which guaranteed freedom of conscience in the colony. Under Penn's influence, fair treatment was accorded the Indians, who responded with friendship in return. When Penn returned to England in 1684, the new Quaker province had a firmly established government based on religious tolerance and government by popular will.

Colonial growth. The century that followed was a period of great expansion and turmoil for Pennsylvania. Its interior included land that was claimed by the French, and, as time went on, the Indians became increasingly hostile to the expansion of settlements to the west and north. Much of the fighting during the French and Indian War (1754–63) took place in Pennsylvania. There the young George Washington began his journey into the Ohio valley to warn the French to leave; later, it was in Pennsylvania that the English general Edward Braddock suffered defeat at the hands of the French forces and their Indian allies.

Site of battles and Revolutionary ferment

For many Pennsylvanians, the period following these conflicts marked growing dissatisfaction with British rule. Limitations on westward expansion, especially as established by proclamation in 1763, were imposed to pacify the Indians, but Pennsylvanians pressed westward over the Alleghenies. Outposts such as Fort Pitt (Fort-Duquesne under the French; now Pittsburgh) became settlements vital to the flow of trade from the opening lands to the west.

By the eve of the American Revolution, Pennsylvania had become a centre of military, economic, and political activity. The first (1774) and second (1775–76) Continental Congresses met in Philadelphia, the Declaration of Independence was signed there, and after the war the city became the capital of the short-lived Confederation and of the fledgling U.S. government.

Early years as a state. In 1790 a new state constitution was adopted that replaced the unicameral legislature of the Revolutionary period with a bicameral one and a fairly strong governor. During the next 70 years, farm equipment was mechanized, roads were improved and extended, canals were built, and railroads spanned the state, all combining with the economic strength of the thrifty Philadelphians to make Pennsylvania a major commercial power. Beginning in 1820, important mining companies were formed to exploit Pennsylvania's deposits of hard and soft coal, and in 1859 Edwin L. Drake drilled the nation's first successful oil well at Titusville. During this same period the state became a leading producer of textiles, ships, lumber, tobacco, and, most important, iron and steel.

The Pennsylvania Emancipation Act of 1781 had pledged the gradual abolition of slavery in the state. Once the Civil War broke out, Pennsylvania became once again a centre of military and political activity. The southern boundary of Pennsylvania, ratified in 1769, was the Mason and Dixon Line. It became the dividing line between the slave and the free states during the Civil War. At Gettysburg the Union army won one of the most decisive victories of the war, against a Confederate force led by General Robert E. Lee.

The Civil War and its effects

Emergence of the modern state. With the end of the Civil War came a period of great economic, industrial, and population expansion in Pennsylvania. Until well into the 20th century, Pennsylvania was the second most populous state in the nation. In 1873 the state passed its fourth constitution; with amendments, the document survived until 1968, when it was so fundamentally reshaped that it became known as the constitution of 1968. In 1898 construction of a state capitol building (replacing a structure that had burned the previous year) was begun at Harrisburg, the capital since 1812. The new building was completed in 1908.

In both world wars, Pennsylvania's heavy industries were major suppliers of iron and steel, arms, and machinery. After World War II, however, the many changes taking place in the global economy began to affect Pennsylvania's emphasis on heavy industry. A relative decline in the state's manufacturing occurred between 1965 and 1984, and Pennsylvania has come increasingly to rely on a variety of high-technology and service industries. (C.L.T.)

THE SOUTH

Alabama

Admitted as the 22nd state in 1819, Alabama comprises 51,705 square miles (133,915 square kilometres) forming a roughly rectangular shape, elongated in a north–south direction. Tennessee is the bordering state to the north, Georgia to the east, and Mississippi to the west. The Florida panhandle blocks Alabama's access to the Gulf of Mexico except in the state's southwestern corner, where Mobile Bay is located. Montgomery is the state capital.

The state offers much topographical diversity. The rich agricultural valley of the Tennessee River occupies the extreme northern part of the state. In northeastern Alabama the broken terrain of the southern fringe of the Appalachian Highlands begins and continues in a southwesterly progression across the northern half of the state. Below that the band of prairie lowland known as the Black Belt has rich soils that once cradled a rural, cotton-producing way of life central to the state's development. Further south stretch piney woods and then coastal plains until one reaches the striking ranks of azaleas blossoming in the Gulf breezes and the moss-draped live oaks of Mobile.

The Black Belt

The landscape of Alabama has been the scene of many of the major crises in the settlement of the continent and in the development of the modern nation. It was a battleground for European powers vying for the lands of the New World, for the fights between the white settlers and the Indians, for the struggles between North and South during the Civil War, and for the forces of economic and social change that have extensively altered many aspects of the Deep South in the years since World War II. Although Alabama continues to trail near the bottom of the states in many significant social rankings, there has been improvement in race relations, particularly in school desegregation and in the election of blacks to political offices. The state's economy has also shown marked improvement. Yet Alabamians and outsiders alike tend to agree that the state's troubled heritage is often still apparent.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Although the average elevation of Alabama is about 500 feet (150 metres) above sea level, this represents a gradation from 2,407 feet (734 metres) atop Cheaha Mountain in the northeast down across the Black Belt to the flat, low, southern Gulf Coast counties. Within this gradation, several relief regions may be distinguished.

Major relief regions

The southern extremities of the Appalachians cover almost half the state. In the far north the Cumberland Plateau region, segmented by river action, thrusts south across the state line. Altitudes rise to 1,800 feet in the more rugged eastern portions. The Great Appalachian Valley forms another marked division to the east. A small triangular portion of the Piedmont Plateau juts across from Georgia at an altitude averaging 1,000 feet.

The character of the state changes markedly as the rugged, forest-clad hills and ridges of the Appalachian extremities give way to the lower country of the Coastal Plain. The plain has a number of subdivisions: in the north lie the rolling Fall Line Hills, while farther south the pine and hardwood belts add irregularity to the flat landscapes. Arcing into the heart of the lowlands of Alabama, the Black Belt has been distinctive from its association with the cotton production that long dominated its rich soils. The 53 miles (85 kilometres) of coastline have occasional swamps and bayous, backed by timber growth on sandy soils and fronted by stretches of white-sand beaches.

Drainage. The Cumberland Plateau region drains to the northwest through the Tennessee River and the often deep valleys of its tributaries, with much water retained in three large, scenic lakes formed in the 1930s by the Tennessee Valley Authority. The rest of the state is drained southward through broader, lazier valleys. The Coosa and the Tallapoosa rivers join north of Montgomery to form the Alabama, which unites with the Tombigbee, draining the state's western portion. Their waters then part again and are discharged into Mobile Bay through the Mobile and Tensaw rivers.

Soils. There are four main soil zones found in Alabama. In the far north, the Tennessee valley contains the dark loams and red clays that add vivid dashes of colour to the landscape when exposed. Farther south lie the varied soils of a mineral belt, and these are succeeded by the rich limestone and marl soils of the Black Belt. Along the coast of Alabama there are sandy loams and deep porous sands.

Climate. The Alabama climate is temperate, with an average annual temperature of about 64° F (18° C), mellowed by altitude to some 60° F (16° C) in the northern counties and reaching 67° F (19° C) in the southern counties, although summer heat is often alleviated somewhat by the winds blowing in from the Gulf. Occasionally the temperature may rise to 100° F (38° C) in the summer, whereas snow may occasionally fall in the northern counties and frosts are periodic. The average summer temperature is 79° F (26° C); the winter average is 48° F (9° C).

Rainfall is fairly evenly distributed throughout the year, with an annual average of 56 inches (1,420 millimetres) and a concentration on the coast. Droughts are infrequent. These favourable conditions have given the state a long growing season, ranging from 200 days in the north to 300 days in the south.

Plant and animal life. The warm climate of Alabama has nurtured a rich plant cover, including more than 125 tree varieties. Most of the thick forests are in the north and northeast. Pine trees predominate, and live oaks are also found statewide, lending grace to the streets of the older towns and cities. Sweet gum and black walnut are also common, while the colourful red cedar is most abundant in the Tennessee valley and the Black Belt, with the stately black cypress clustering around rivers and ponds. There are many varieties of shrubs and grasses, and bamboo, large canes, and mistletoe are widespread. Muscadines, scuppernongs, and blackberries also flourish. Beardlike Spanish moss gives coastal woodlands a distinctive charm.

Birdlife, too, is rich. Bluebirds, cardinals, blue jays, mockingbirds, doves, woodpeckers, owls, hawks, yellowhammers, and an occasional eagle are found here. Other wildlife includes rabbits, squirrels, opossums, foxes, wildcats, raccoons, muskrats, deer, and even a few bears. Coyotes and armadillos have spread into Alabama from the west. Snakes include poisonous rattlers, water moccasins, copperheads, and corals as well as some nonpoisonous types, such as black snakes. Alligators still exist in some of the swamps and bayous of the coastal regions.

Settlement patterns. By 1980 less than two-fifths of the population of Alabama was classified as rural, and the proportion was declining still further as rural residents sought homes in the state's cities. Eight counties had a population loss during the 1970s. Much of the loss was from the old cotton region of the Black Belt, which had been largely depopulated by the 1980s. By the 1970s, however, the state had ceased to be a net loser of population to Northern migration. A large part of Alabama's rural population lived near, and worked in, metropolitan areas.

By 1960, for the first time, more than half of Alabama's population was concentrated in 10 counties, and the dominance of the urban areas has since continued.

Urban growth

The sprawling city of Birmingham continues as the major metropolitan area of the state, with its increasingly service-oriented economy employing about one-fourth of the state's workers and producing one-fourth of all its manufactures. Birmingham's population growth has taken place in outlying suburbs, however.

Mobile, the state's port city, is the second largest metropolitan area, the growth of which, after a surge in the 1970s, slowed in the 1980s, partly as a result of a recession in the oil industry. Cities such as Gadsden and Florence, with manufacturing-based economies, experienced stagnation in the 1980s, while Huntsville grew rapidly as a result of national defense installations and new high-technology industries. Montgomery's population expanded, owing partly to the growth of state government.

The people. Three-fourths of the state's population is white. The white population is significant for its deep roots in the state: the number of foreign-born residents is very small, and most whites are descendants of 19th-century settlers who came from adjoining regions to the east and north.

Black roots in the state

Black Alabamians have equally deep roots in the state, dating to the days of chattel slavery and the African slave trade. Other ethnic minorities make up only a very small percentage of Alabama's population. Religious affiliations in the state are predominantly Protestant, with the various church groups in the black community having played an unusually prominent social role since the days when other outlets for such activity were denied.

The birth rate in Alabama, especially in the areas of rural poverty, was somewhat above the national average for many years, but with the depopulation of rural areas and the expansion of the urban economy, the ratio has come nearer to the national average. Some rural areas of the state continue to be plagued with very high rates of infant mortality, attributable to inadequate health care.

The economy. Among the 50 states, the relative status of Alabama may be indicated by its income per capita: it has ranked close to the bottom of the economic scale for a number of years. This low status results, in part, from the depressed state of agriculture, which employs a large segment of the population. Rural poverty thus drags down the state average, concealing more promising developments and the much stronger economic base that exists in the

urban areas. An important development in the Alabama economy has been the emergence of Birmingham as a financial and commercial centre, especially as the home of major state banks, regional utilities, national insurance companies, and international construction concerns.

Alabama spends a high percentage of its total revenue on education, health and hospitals, welfare, and highways. The state has generally low taxes on property and relatively high taxes on consumption. Federal funds support programs affecting agriculture, public education, a wide range of health and welfare projects, conservation, urban development and public works, and highway construction. The federal government maintains the Air University in Montgomery and the George C. Marshall Space Flight Center and the Redstone Arsenal at Huntsville, as well as several veterans' hospitals and a part of the Tennessee Valley Authority operations.

Agriculture. The Alabamian rural economy challenges the traditional view of a dependency on cotton. Although cotton still remains of local importance, it suffered a heavy blow with the onset of the boll weevil blight in 1915, and acreage continued to decline. Mechanization and consolidation increased the average farm size after the 1930s. The diversification of agricultural production then brought a great increase in the acreage devoted to forestry, and cotton fields were given over to pasture for dairy and beef cattle. Poultry has become a major farm product in the state. The principal crops are cotton lint, peanuts (groundnuts), soybeans, and corn (maize). Farm income continues to rise, and the average value of a farm has multiplied many times since the end of World War II.

Diversification in agriculture

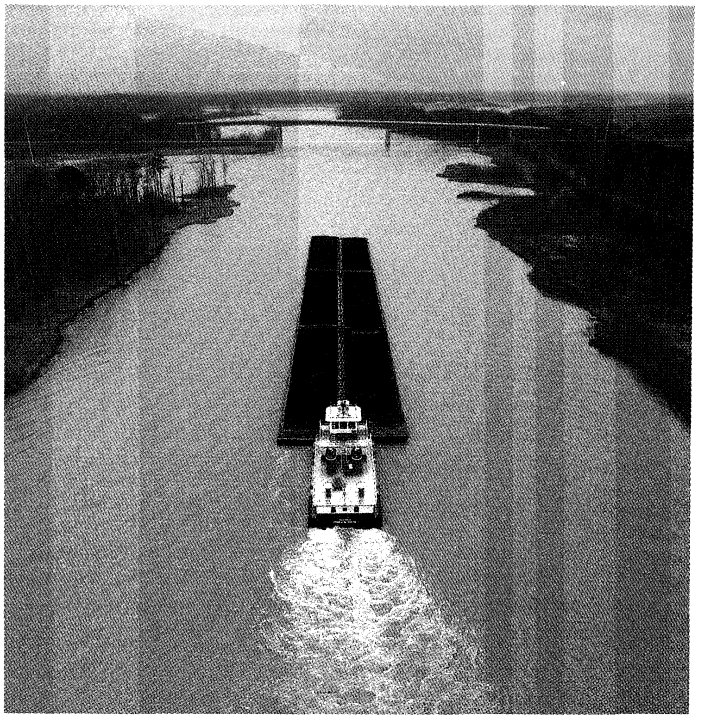
Industry. Industrial development in Alabama has long been based on the iron and steel industry of Birmingham, the development of which was facilitated by accessible deposits of iron ore, coal, and limestone. Other minerals include the state's well-known white marble. Oil production in commercial quantities dates from 1944, and there are a number of wells in the coastal regions. World War II defense industries gave an impetus to the industrial economy of the state. Although iron and steel production continues to be important to Alabama's economy, the growing chemicals and plastics industries have reduced the reliance on primary metals. Since 1960 the Marshall Space Flight Center at Huntsville, notable for producing the Saturn booster rockets, has been a major contributor to the state's economy. In the 1970s Huntsville experienced the growth of high-technology industries, including computer design and production. At the same time, Birmingham had changed to a predominantly service economy, in which the city's major employer was the University of Alabama and its regional medical centre.

Transportation. Together, the six major rivers of Alabama provide about 1,300 miles of navigable waterways, while Mobile Bay has been deepened by a ship channel. Mobile developed as a modernized port and ranks among the top dozen seaports of the nation. The Tennessee–Tombigbee Waterway, a 234-mile canal opened in 1985, connected two of the state's main river systems. Although railroad transportation, as elsewhere in the United States, has suffered a relative decline in Alabama, bus, truck, and airline traffic have increased in the state.

Administration and social conditions. *Government.* Alabama is governed by a bicameral legislature and a governor and cabinet. The legislature consists of 35 senators and 105 representatives who meet annually in regular sessions and are elected for four-year terms. The constitution, a complex and, some have claimed, outdated and often inadequate document, dates from 1901. After various poll tax and other tax provisions aimed at restricting black voter registration were declared illegal by the U.S. Supreme Court, voter registration among blacks made great progress. The chief administrative officers of the state, ranging from the governor to the state Board of Education, are all elected for four-year terms. The state Supreme Court of nine elected members is the highest judicial body.

State and local government

At the county level the chief elected officials in Alabama are the county commissioners, judges of probate, tax assessors and collectors, and boards of education. In the municipalities there is no uniform system of government: the mayor–council form is most common, but some cities have a commission, and some employ a city manager.



Tugboat pushing barges on the Tennessee–Tombigbee Waterway, western Alabama.
© Ed Malles/Photo Options

The Democratic Party of Alabama has long held political control of the state government, although there are signs of an increased Republican showing. In 1986 the state elected its first Republican governor since Reconstruction, and a few Republicans, usually from suburban areas, won places in the state legislature and in local government bodies during the 1980s. Numerous black political organizations have helped increase black participation in the political process. Black office holding was commonplace beginning in the 1970s; several towns and cites have elected black mayors, including Birmingham in 1979.

Education. Elementary and secondary education in Alabama have improved substantially since the 1960s, though public schools in the state suffer from very weak local funding resulting from the state's low taxation on property. Teachers' salaries have risen since the early 1970s. Rural schools get less support than urban ones. Most public schools are integrated racially, though a substantial minority of white children attend private academies.

Alabama has 15 state-supported four-year colleges, many junior colleges and trade schools, and more than a dozen private colleges and universities. The University of Alabama system comprises the state's original college at Tuscaloosa and newer campuses in Huntsville and Birmingham, the latter being home to a nationally renowned medical centre. Auburn University, one of two land-grant institutions in Alabama, has the largest enrollment in the state and provides the headquarters for agricultural extension work.

Most black college students are enrolled in traditionally black institutions, although an increasing number attend formerly all-white colleges. Among the several public and private black institutions, Tuskegee University, founded in 1881 and the home of Booker T. Washington and George Washington Carver, is the most well-known.

Higher education in Alabama suffers from duplication of effort caused by the overabundance of institutions, which dilutes resources. This duplication was the product of both a dual system for racial separation and a tendency to build schools as political favours. The state historically has not developed a strong mechanism for coordinating higher education.

Health and welfare. In rural areas and among nonwhites, educational and economic opportunities are fewer, and health and medical resources and services are less available. Welfare payments in Alabama rank low by national standards. Penal institutions include several prisons and camps for youthful offenders. Controversy over conditions in these institutions has occasionally existed, and substantial reform has been introduced since 1970.

Folk culture

Cultural life. Alabama enjoys a rich vein of folk culture in the rural areas, especially among blacks. Story telling has attracted the attention of folklore specialists. The experiences of rural life contributed important elements to the development of modern American popular music, including ragtime and jazz. Sacred music, in the form of gospel quartets and shape-note, or "fa-so-la," singing, remains a vital part of Alabama's cultural life. Quilt making is a highly developed folk art among rural dwellers, black and white.

Several Alabama writers have won attention through the exposition of themes in the state's settings. Johnson J. Hooper, John Gorman Barr, and Joseph G. Baldwin were popular local-colour writers in the 19th century. Booker T. Washington and Helen Keller wrote powerful and popular autobiographies in the early 20th century. The novelist William March made a distinguished literary contribution in his stories and novels in the 1930s and '40s, particularly *Company K* and *The Looking Glass.* T.S. Stribling, in a trilogy of realistic novels in the 1930s; Harper Lee, in *To Kill a Mockingbird* (1960); and Mary Ward Brown, in *Tongues of Flame* (1986), have explored social conditions, especially racial problems, in critically acclaimed works.

Art museums are found in Huntsville, Montgomery, Mobile, and Birmingham, the latter containing an especially impressive collection of American art. The George Washington Carver Museum at Tuskegee University has unique material on black history. The Sloss Furnace Museum focuses on Birmingham's industrial history. The Alabama Space and Rocket Center in Huntsville explores the development of space travel. The Alabama Shakespeare Festival in Montgomery offers professional productions of classic and modern plays.

Special library collections include those on medical history at the University of Alabama Medical Center in Birmingham; the Booker T. Washington Collection of black history material at Tuskegee University; and the Alabama and Southern history material at the Alabama Department of Archives and History, founded in 1901 as the first such department established in the United States.

Several historic places in Alabama are supervised by the state, including the Mound State Monument in Hale county, an important site for Mississippian Indian culture; and Fort Morgan, a Confederate fortress standing at the entrance to Mobile Bay. Alabama boasts many surviving examples of 19th-century residential architecture, perhaps most notably Gaineswood Mansion in Demopolis. The U.S. Park Service has made parts of Tuskegee University a national park for the study of black history.

Distinctive festivals are celebrated in various Alabama places. Mobile's Mardi Gras is a major event in February, as are its springtime Azalea Trail garden tours and the annual Junior Miss Pageant. Birmingham explores international culture in its annual spring Festival of Arts. The town of Opp hosts a yearly Rattlesnake Rodeo that draws large participation. Most Alabama towns and cities sponsor historical pilgrimages in April to celebrate architectural survivals.

The state maintains many parks and several large public lakes. Waterskiing, boating, and stock-car racing rank among the most popular recreational activities among Alabamians. The Alabama International Speedway at Talladega attracts hundreds of thousands of auto-racing enthusiasts each year. Three dog-racing tracks draw many bettors. College football, especially the teams fielded by the state's two major universities, elicits avid devotion from a large proportion of the state's residents.

HISTORY

The Indians. The earliest and longest established inhabitants of the present-day state of Alabama were Indians. Visible traces of their occupancy, which spanned almost 10,000 years, may be seen in the great mounds that snake across the landscape near the river valleys. Many placenames in the state also indicate an Indian origin. The name Alabama itself derives from an Indian word meaning "thicket clearers." The principal Indian groups at the time of the initial European exploration of the region were the Chickasaw, in the northwest; the Cherokee, in the northeastern uplands; the Upper Creek, or Muskogee, in the centre and southeast; and the Choctaw, in the southwest.

European rivalry, settlement, and growth. The first known European explorers were of Spanish descent and arrived at Mobile Bay in 1519. The main thrust of exploration came in 1540, when Hernando de Soto and his army of about 500 men entered the interior from the valley of the Tennessee River to search for gold. His expedition, which extensively crisscrossed the area, was important because of his discovery of the Mississippi River, the knowledge he gained of a wide band of southern Indian cultures, and his role in opening up the whole region to European settlement. A battle with the ill-equipped warriors of the Indian chief Tuscaloosa, however, resulted in the slaughter of several thousand Indians; it may have been the bloodiest single encounter between whites and Indians in North America. De Soto found no gold, and subsequent Spaniards failed to establish settlements in Alabama.

The de Soto expedition

The ensuing 250 years were characterized by struggles among the French, British, and Spanish for control of the region, often in shifting alliances with the Indians of the area. In 1702 the French founded the first permanent European settlement in Alabama at Fort Louis, north of present-day Mobile. The British had also made a number of trips to the region from the Carolinas, but the French settlements—part of a string of forts arcing down from Canada and designed to contain the British—were more numerous. Port Dauphin, on Dauphin Island, received the first Africans when a slave ship landed there in 1719.

The Treaty of Paris (1763) gave to Britain what was then the only settled part of Alabama, the Mobile area. In another Treaty of Paris (1783), which officially ended the American Revolution, Spain gained Mobile, and the new United States received the rest of the present-day state. Then in 1813 the United States, claiming Mobile as a part of the Louisiana Purchase of 1803, drove the Spanish out of the area and established authority throughout the state. As for the Indians, the Cherokee, Chickasaw, and Choctaw had ceded some land by 1806. In 1814 General Andrew Jackson inflicted a decisive defeat on the Creek at the Battle of Horseshoe Bend. The influx of white population following these actions and the institution of the cotton economy caused a rapid removal of the Indians to the west. The Creek cession of 1832 virtually ended the Indian claims to territorial rights in Alabama. Most descendants of Alabama Indians live in Oklahoma; only a few hundred Creek remain in the southern part of Alabama.

The antebellum period. Alabama was established as a separate territory in 1817 and became a state in 1819. By 1820 Alabama's population was more than 125,000, including about 500 free blacks. By 1830 there were 300,000 residents, 38 percent of them slaves, and cotton was the principal money crop. Until the Civil War, domestic politics centred on the removal of the Indians, land policy, the banking system, and the question of slavery. The state suffered severely for almost a decade in the economic depression that followed the panic of 1837. During the late 1840s and '50s many efforts were made to create a more modern, industrialized economy. Railroads, cotton manufacturing, and some mining were begun, though such efforts often suffered from a shortage of capital. The vast majority of investment remained in cotton and slaves. By 1860 the population was almost 1,000,000; nearly half of the people were black, and all but 5 percent of the state's population was rural.

The Civil War and its aftermath. In 1861 Alabama seceded from the Union and joined the Confederate States of America, which established its first capital in Montgomery. The state legislature conscripted soldiers and appropriated several million dollars for military operations and for the support of the families of soldiers. Some 35,000 of the

122,000 Alabamians who served in the war died. Following the collapse of the Confederacy and the refusal of the state legislature to ratify the Fourteenth Amendment to the U.S. Constitution, Alabama in 1867 was placed under military rule. The next year the state ratified a new constitution that protected the civil rights of blacks, and Alabama was readmitted to the Union.

Reconstruction and after

From 1868 to 1874 the state was in political turmoil. To many whites the Reconstruction period was tragic, but to many blacks it was a period of opportunity and hope. The Huntsville *Advocate* asserted, "This is a white man's government and a white man's state," and the Ku Klux Klan used terror to enforce that view. Among white Alabamians, a struggle ensued between those who defied the notion of blacks having political rights and power and those willing to cooperate with blacks and their Northern allies. Blacks demanded access to education and were given it, but most whites insisted that schools be racially separate. Although blacks participated in the constitutional conventions and in the state legislatures, their political power was not as strong as that of blacks in South Carolina, Mississippi, or Louisiana. In 1874 the white Democrats of Alabama, most of whom had been supporters of the Confederacy, regained control of the state political machinery. Blacks were rendered almost powerless until the Civil Rights movement of the 1960s. Throughout this period, however, some blacks worked diligently to stimulate political activity, to enlighten and influence their white fellow citizens, and to influence the state and federal governments to guarantee their political and social rights.

In 1875 a state constitutional convention was held and a new conservative constitution ratified. Subsequent conservative political efforts centred on restricting the participation of blacks in government, reducing expenditures and state services, and fostering the expansion of railroads and industry. By 1901, when another state constitution was ratified—this one disfranchising blacks—there was virtually no black participation in government, and a tide of social and political reaction was in full flood.

Emergence of sharecropping

The economy recovered slowly from the devastation of the war. Sharecropping as a system of land tenure and labour relations emerged, and with it came even greater dependence on a single crop, cotton. Depressed agricultural conditions fanned a populist revolt among small farmers in the 1890s. After 15 years of delay from depression and capital shortages, cotton manufacturing and pig-iron production began to grow steadily in the state in about 1880, with a long interruption during the depression of the 1890s. By the turn of the century, however, Alabama was one of the more highly industrialized Southern states.

The 20th century. In 1900 the state remained largely rural. The onset of the boll weevil in 1915 seriously damaged one-crop agriculture, forcing a diversification of the rural economy. Rural dwellers, mostly poor and black, began an exodus to Southern cities and to the North, where cheap foreign labour supplies had dried up during World War I. A factor in encouraging the out-migration of Alabama blacks was the pattern of racial segregation under the "Jim Crow" system, which was enforced legally and extralegally. The level of the state's black population began a slow decline, which reduced their numbers to less than one-third of the total population by mid-century.

The Great Depression of the 1930s made suffering virtually universal in the state. Many thousands of tenant farmers lost their credit when the price of cotton fell to its lowest point. Birmingham's industrial economy almost came to a standstill. Federal relief programs alleviated some problems, and the Tennessee Valley Authority created new economic activity in northern Alabama.

The defense buildup lifted the Alabama economy out of depression in the World War II years. Statewide, the war did more to encourage industrialization than any other historical factor. After the war the contributions of the federal government in support of agriculture and national defense, including the space program, and the provision of such services as road building, education, and welfare, helped to transform the state's economy. The mechanization of agriculture in the 1940s and '50s completed the revolution in the state's agricultural economy.

The Civil Rights movement

Segregation, nevertheless, continued to give rigidity to the social framework of Alabama and effectively excluded the black population from social and economic power. The 1954 U.S. Supreme Court decision declaring segregation in public education unconstitutional encouraged black Alabamians to work to improve race relations. Progress was nevertheless slow and bitter. The state acquired international significance as the site of such noteworthy civil rights actions as the bus boycott of 1955–56 in Montgomery, which introduced Martin Luther King, Jr., to the nation; the "freedom rides" of 1961; street demonstrations in Birmingham in 1963 in which commissioner of public safety Eugene Connor turned fire hoses and police dogs on black protesters; Governor George C. Wallace's defiant attempt to stop the desegregation of the state university that same year; the death of four black Birmingham children in an explosion that destroyed their Sunday school, also in 1963; and the march from Selma to Montgomery in 1965.

This period of black activism precipitated major revisions in American law. The Civil Rights Act of 1964 finally ended segregation in public accommodations and provided protection against some forms of employment discrimination. The Voting Rights Act of 1965 outlawed most means of limiting the political rights of blacks.

As a result of these activities, black citizens have attained better access to public services, broader educational and economic opportunities, and freer political participation. By the late 20th century the percentage of blacks registered to vote had increased fivefold. Blacks have been elected in small but increasing numbers to state and local government positions. Job opportunities in some professions and in government have improved markedly for blacks, though poverty in the state is still disproportionate among blacks. Many professional and civic bodies and most schools have achieved a good measure of desegregation, though churches have not. Progress has been sometimes slow and incomplete but nevertheless significant.

(C.G.G./R.J.No.)

Arkansas

Ever since Arkansas was admitted as the 25th member of the United States in 1836, its people have maintained a remarkable homogeneity, and today most of them are native to the state. Striking cultural contrasts exist within Arkansas, however, with the long-isolated mountain people who eked out subsistence livings in the north and west counterposed to the people to the east and south who created a Southern environment in which cotton growing and sharecropping long were the dominant modes of economic life. Between the two regions lies Little Rock, the capital and the urban and economic centre of the state. Its location and increasingly cosmopolitan character are symbolic of Arkansas's growing unification and urbanization.

Arkansans are concerned about the state's relative poverty and lack of development. Although Arkansas remains among the lowest-ranking states in income per capita and other economic indicators, the overall economy in recent years has gained faster than the national average, and the population has increased, reversing a long decline. Programs have been developed to increase these trends and to continue the process of equalizing the educational, economic, and social opportunities of the state's citizens.

Arkansas's 53,187 square miles (137,754 square kilometres) make it 27th in area among the states, but, except for Louisiana and Hawaii, it is the smallest state west of the Mississippi River. Its neighbours are Missouri to the north, Tennessee and Mississippi to the east, Louisiana to the south, Texas to the southwest, and Oklahoma to the west. Arkansas has the high Ozark and Ouachita mountains in the north and west and a heavy tracery of rivers that cut through its rich agricultural lands. Nearly all of the rivers flow from northwest to southeast and empty via the Arkansas and the Red into the Mississippi, which forms the major eastern boundary. The state's name was used by the early French explorers for the Quapaw Indians and the river along which they settled. It probably was a phonetic spelling of the Illinois term for "downriver" people, a reference to the Quapaw.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* A line drawn from the southwestern corner to the northeastern corner of the state approximates the division between the highlands lying west and north and the lowlands lying south and east. The highlands are divided by the Arkansas River valley into the Ouachita Province on the south and the Ozark Plateau on the north. The lowlands include the Mississippi alluvial plain in the east and the western Gulf Coastal Plain in the south and extreme southwest. The highlands are covered with the dense pine and hardwood forests of the Ouachita and Ozark national forests.

Mountains and plains of Arkansas

The Ozark Plateau is broken by broad, flat-topped ridges and steep valleys with fast-flowing streams. The more rugged southern edge, known as the Boston Mountains, contains the highest elevations. Excellent farmland, producing a wide variety of crops, lies in the northern part. The Arkansas River valley contains the highest point in the state, Mount Magazine, at 2,753 feet (839 metres) above sea level. The western section has extensive coal and natural gas deposits. Several peaks in the Ouachita Province reach 2,500 feet. The mountains are eroded, exposing faulted rock, and the ridges extend west and east. The famous Hot Springs National Park is in this area.

The western Gulf Coastal Plain has gentle hills suitable for livestock grazing and farming. Much of this area consists of pine and white oak forests, which sustain lumbering industries. Petroleum and natural gas deposits have been developed in the Smackover and El Dorado area. The Mississippi alluvial plain, much of which was once a vast swamp, is now well drained and protected against flooding. It contains the state's richest and most fertile farmland. Rice and soybeans have replaced cotton as the major crops. A long, narrow chain of hills, Crowley's Ridge, runs north–south through the centre of the plain.

Climate. The climate generally is mild in winter and hot in summer. Normal high and low temperatures in Little Rock in January are 51° and 29° F (11° and −2° C); in July they are 93° and 71° F (34° and 22° C). The normal annual precipitation of 49 inches (1,250 millimetres) is distributed about equally during the year, though summers tend to be drier than the other seasons.

Plant and animal life. A great variation in soils and elevations in Arkansas supports a large number of plant species. There are more than 200 species of trees, of which pine, oak, hickory, maple, gum, ash, cypress, and elm are the most important. In fall and spring the woodlands are colourful with dogwood, flowering fruit trees, redbud, and innumerable wild flowers.

© Terry Donnelly from TSW—CLICK/Chicago

Flowering dogwood (*Cornus florida*) in the Ouachita National Forest, Hot Springs National Park, Arkansas.

Arkansas is situated on the Mississippi flyway; migratory water birds and some 300 native species attract hunters to the rice fields and reservoirs of eastern Arkansas. Deer, quail, opossums, turkeys, squirrels, and rabbits are among the more abundant game animals. Bobcats and wolves are not uncommon in the hill country. The lakes and streams of the state offer an abundance of fish, including crappie, bass, drum, catfish, buffalo, gar, and trout.

Settlement patterns. The inhabitants of the Ozarks and Ouachitas once lived in rural isolation, which bred an independence of spirit and a suspicion of strangers. Hunting and fishing were essential to supplement the limited produce of their farms. Since a plantation economy was impracticable in the uplands, few slaves were brought into the region. Settlement of the westernmost regions was long discouraged by the lawless frontier border with Indian Territory. Much of this area remains timbered and lies within the Ouachita and Ozark national forests.

In eastern Arkansas the plantation economy produced a vast gulf between the sharecroppers and tenants on one end of the social scale and the managers and landlords on the other. The owners of small farms or businesses constituted another class. The croppers lived a bare and meagre existence. Handicapped by lack of economic resources and education, they accomplished remarkable results through the Southern Farm Tenants Union, which they organized in eastern Arkansas in the 1930s; this organization influenced the national farm policy of presidents from Franklin D. Roosevelt onward.

Contrast of cultures

Although changes in the economy were evident earlier, the rate of change since World War II has been dramatic. The Ozarks are no longer isolated. A network of paved highways brings tourists to enjoy the region's scenic beauty and varied recreational activities. Numerous "retirement villages" attract visitors and buyers from across the country. The tourist industry remains the economic mainstay, though small industrial plants have taken advantage of the climate and the ample labour supply.

Mechanization of farming in eastern Arkansas and the shift from cotton farming to rice and soybeans has virtually eliminated the sharecropper—though not the rural poor. As the pace of mechanization increased, so did the exodus of the tenant farmers to cities in the North and East. Farming is increasingly a corporate venture. Eastern Arkansas is still, however, more Southern in character than the mountainous region. The shacks of the sharecroppers are gone, and much of the rural population has left the state or moved to nearby towns offering nonfarm employment. In eastern and central Arkansas reside the majority of the state's blacks, many of whom still work the land as their ancestors did.

Little Rock, the major port on the Arkansas River, lies among the easternmost foothills of the Ouachita Mountains. A marketing centre and the site of manufacturing facilities, the city has completed or undertaken several urban renewal projects in its downtown area, including the construction of a pedestrian mall, the renovation of historic buildings, and the expansion of convention facilities. At the western boundary of the state lies Fort Smith, the second largest city on the Arkansas River. It is one of the most industrialized cities in the state and serves as a regional business and service centre. The economy of Pine Bluff, some 50 miles (80 kilometres) downriver from Little Rock, depends primarily on the surrounding agricultural area. Texarkana, contiguous with the Texas city of the same name, is an important regional rail centre.

The people. Prior to the Civil War, Arkansas's population came largely from Kentucky and Tennessee, a part of the westward movement of Scottish, Scotch-Irish, and English stock from Virginia and the Carolinas since early colonial times. The black population in 1860 was about 110,000, or 25 percent of the total; by the late 20th century there were more than 350,000 blacks living in Arkansas, making up a decreasing percentage of the total population. A few counties in eastern Arkansas are

more than 50 percent black. The heaviest concentrations of population are in the fertile eastern alluvial plain, in the river valleys, and on the plateaus in the northwest.

The largest religious denominations are the Baptist, United Methodist, Presbyterian, Church of Christ, and Roman Catholic. The religious atmosphere is one of conservative fundamentalism, and Arkansas is considered a part of the Bible Belt. Fundamentalism underlies many characteristic attitudes of Arkansans. The sale of alcoholic beverages is subject to local option; many counties and cities prohibit their sale or permit it only in private clubs and certain other establishments in major cities. The right-to-work amendment to the state constitution in 1944, which prohibits compulsory union membership, was sponsored by the Christian Association. Harding College in Searcy is the site of an annual Freedom Forum, which advocates a blend of religious fundamentalism, extreme patriotism, and free-enterprise capitalism.

The economy. Cotton is no longer king in Arkansas, and the state is no longer primarily agricultural. Industrialization and urbanization are major factors in Arkansas's recent record of economic progress. Labour unions are strong in transportation, utilities, construction, and heavy industry, but most of the state's labour force is unorganized. In both political and economic policy-making, labour is less influential than business.

The demand for increased revenue has led to cooperation between leaders in the private sector and public agencies in the promotion of economic growth. Progress in overhauling the state tax structure and in improving methods of tax collection has been slow but steady. The state attempts to generate more revenue by raising income per capita through increasing employment opportunities and developing human resources to their maximum.

Mineral wealth

Resources. Oil fields in southern Arkansas yield natural gas and bromide salts. Coal of a nearly smokeless quality, as well as natural gas, are found in the Arkansas River valley. Arkansas's aluminum industries have reduced substantially their bauxite mining operations and have closed some of their reduction plants. Experimental use of lignite in coal-fired electrical generating stations offers the possibility of extensive commercial development of the widespread lignite deposits in southern Arkansas. Magnet Cove, near Hot Springs, contains more than 40 different minerals in one small valley; barite and titanium are the most important. Arkansas whetstones made from novaculite are regarded as among the finest in the world. Near Murfreesboro, in southwestern Arkansas, is the only diamond mine in the nation, now operated only as a tourist attraction. Almost one-half of Arkansas is covered with forests, including extensive stands of pine and white oak.

Hydroelectric power is produced at dams erected by the U.S. Army Corps of Engineers and by private companies. Two nuclear power plants have been constructed near Dardanelle, and coal-fired stations have also been built.

Agriculture and industry. Cotton remained the major source of agricultural income into the 1960s. Since then rice and poultry, of which the state is a leading producer, have dominated. Other important crops are soybeans and grains. Commercial fish farming has begun to take advantage of the extensive rice paddies of eastern Arkansas. Farms have followed the national trend of increasing in size while decreasing in number.

Manufacturing chiefly involves the production of consumer goods. Major industries include food processing and the manufacture of clothing, furniture, electrical and nonelectrical machinery, electronic equipment, and fabricated metal products.

Transportation. Several major railroads provide freight service within Arkansas, as well as to major cities in the central United States. Airline service is provided by national carriers from a number of airports to any point in the nation. By interstate highways more than half of the nation's population is within a two-day driving radius of Arkansas. Motor-fuel tax revenues are reserved for public highways and streets. As funds are available, major routes are being rebuilt as four-lane, limited-access roads.

The McClellan-Kerr Arkansas River Navigation System for navigation and flood control is the largest civil works project ever undertaken by the U.S. Army Corps of Engineers. The project provides access to more than one-half of the nation's navigable inland waterways. Annual freight tonnage along it has exceeded estimates, and significant industrial growth has been attributed to the project.

The Arkansas River Navigation System

Administration and social conditions. *Government.* Adopted in 1874, Arkansas's constitution has been amended more than 70 times. The governor, who is elected to a four-year term, has the authority to summon the legislature into special session and to veto acts, though a veto may be overridden by a simple majority vote in each legislative house. The Senate has 35 members with four-year terms; the House of Representatives, 100 members with two-year terms. The judicial branch includes the Supreme Court of seven popularly elected members who serve eight-year terms, a court of appeals, and circuit and chancery district courts. Where established, municipal courts have jurisdiction throughout the county.

Elected officials of the 75 counties include county judge (chief executive), clerk, treasurer, sheriff-collector, surveyor, and coroner. In each county, elected justices of the peace make up a quorum court, which serves as an advisory body to the judge and exercises some legislative functions. There are many local improvement districts and school districts. Although a number of incorporated cities have a city-manager form of government, the traditional mayor–council form is most common.

Political life

Arkansas is a predominantly Democratic state. Since Reconstruction, few Republican governors or congressmen have been elected, although the Republicans have gained strength. Unless a candidate receives a majority of votes cast in a preferential primary, a runoff is required. Permanent voter registration replaced the poll tax in 1965.

Education. The public school system functions under the state's department of education and district school boards. Specialized institutions include schools for the deaf and the blind. The state's facilities for retarded children and for the treatment of mental illness have attracted nationwide acclaim. Vocational-technical schools serve most areas of the state. In addition, private schools, most of them church-related, offer instruction from kindergarten through the secondary level.

The University of Arkansas has a main campus at Fayetteville and branches at Little Rock, Pine Bluff, and Monticello. The graduate schools of health sciences, technology, and social work are located in Little Rock. Several other universities and two-year colleges are supported by the state. Some of these were developed from institutions for the agricultural and mechanical sciences and for teacher training. There are also several private and church-affiliated colleges and universities.

Health and welfare. Problems of malaria, pellagra, and pinworm that once plagued the region have been virtually wiped out by widespread efforts of state and local health authorities. The state departments of health and human services administer many programs funded in part by the federal government. Emigration of young people over the past several decades has aggravated health and welfare programs, especially in declining rural areas. Welfare payments are among the lowest in the nation. The mild climate and attractive scenery has fostered the establishment of retirement villages in the Ozarks.

The wages of Arkansas's workers are among the lowest in the nation, and living costs approximate those of the south central region. Blacks live at distinctly lower economic and social levels despite improvements.

Cultural life. The people of eastern Arkansas are typically Southern in their speech and customs. Central Arkansas also reflects its Southern heritage, but the speech and manners of its people have been influenced by immigrants from other parts of the country. The rural areas of the Ouachitas and Ozarks have retained to the fullest degree an unchanged culture.

The fine arts are well served by semiprofessional orchestras, choral groups, and ballet, theatre, and opera companies in Little Rock and other urban centres. Most colleges and universities offer training and performance in the arts. A four-state opera workshop is held each summer in the Ozarks. Arkansas's richest contributions are in

Folk arts and culture

the folk arts of the Ozarks. A major folk art centre in Mountain View has been designed to provide a showcase for local and visiting performers in dance and music; to preserve traditional skills in ceramics, jewelry, wood carving, hooked rugs, and basketry; and to offer instruction in the native folk arts. Other aspects of folk culture include the gospel singing of rural areas. Black spirituals and soul music flourished in Arkansas long before they became popular nationwide.

The University of Arkansas has a fine collection of archaeological and historical artifacts. A collection of colonial glassware is featured in the Museum of Science and History, housed in the old federal arsenal in Little Rock. The Arkansas Arts Center and its branch for the decorative arts, located in a restored antebellum mansion, have attracted regional recognition. Historic sites include Arkansas Post, the first European settlement in French Louisiana; Washington, the Confederate state capital during the Civil War; and the Territorial Capitol Restoration and Old State House in Little Rock.

Arkansas devotes considerable effort to attract out-of-state vacationers, who annually contribute millions of dollars to its economy. State and national agencies stock lakes and streams with fish, and the state's preserves and conservation practices assure ample game in hunting seasons. The largest single attraction in Arkansas is Hot Springs National Park, which offers both outdoor recreation and luxury hotels throughout the year. The Buffalo National River, Blanchard Springs Caverns, and the resort town of Eureka Springs, known for its arts community and Victorian architecture, attract thousands of visitors annually. In addition to the five national park sites, there are numerous state parks affording a great variety of recreational activities.

Little Rock has the oldest newspaper west of the Mississippi, the *Arkansas Gazette,* founded in 1819; like the *Arkansas Democrat,* it serves the entire state. Central Arkansas is served by radio and television affiliates of the major networks. The state has numerous commercial radio stations. The Arkansas Educational Television network covers most of the state, and cable television serves urban communities.

HISTORY

Exploration and settlement. Arkansas's early inhabitants included bluff-dwelling Indians, whose farming and hunting culture flourished about AD 500. Later mound-building cultures left sepulchral mounds and other remains along the Mississippi.

Spanish and French explorers traveled the trans-Mississippi regions in the 16th and 17th centuries, and the Frenchman Henri de Tonty founded the Arkansas Post on the lower Arkansas River in 1686. The first permanent white settlement in what is now Arkansas, it served as a fur-trading centre and a way station for travelers between the Gulf of Mexico and the Great Lakes.

Following the Louisiana Purchase by the United States in 1803, Arkansas lay within the territories of Louisiana until 1812 and of Missouri until 1819, when it became a separate territory. Its northern boundary, latitude 36°30′ N, was the line of the Missouri Compromise in 1820 that later separated the slave and free states in the West.

Statehood and Civil War. By the time of statehood in 1836, all land titles of the Quapaw, Osage, Caddo, Cherokee, and Choctaw Indians had been withdrawn by the U.S. Congress, and the tribes were forced westward into the Indian Territory, the future Oklahoma. Violence broke out intermittently along the state's western border until the late 19th century, when the frontier atmosphere disappeared with the white settlement of the Indian Territory.

War and Reconstruction

Although a slave state, Arkansas did not secede from the Union until May 1861—five months after South Carolina did so. Arkansas took this action only after the Confederate capture of Fort Sumter and President Abraham Lincoln's call for volunteers. Union sentiment was strong in northern Arkansas; about 6,000 Arkansans joined the Federal forces. About 58,000, however, fought for the Confederacy. Little Rock fell to Federal troops in 1863, and for a decade the state was a legislative battleground between secessionist supporters and the imposed Republican government. Arkansas was readmitted to the Union in 1868, but internal strife approached open warfare. In 1874 the state returned to the fold of the Democratic Party, and remained there until Winthrop Rockefeller, a Republican, was elected governor in 1966.

The Civil War's chief long-range effects on Arkansas, as on most of the other former Confederate States, were a crop-lien sharecropping system, a race problem of new and formidable dimensions, a one-party (Democratic) political system, and widespread poverty. Economic development in Arkansas was severely handicapped by the collapse of state credit following repudiation in 1885 of bonded indebtedness, including interest of nearly $14,000,000.

Recent decades. Until World War II, Arkansas experienced slow economic development, remained predominantly rural, and was tied to a single cash crop—cotton. The depression of the 1930s was worsened by years of drought that turned many farm families into itinerant labourers. Defense-related activities during World War II and the postwar mechanization of agriculture greatly altered both the economy and social patterns. Women entered the labour force, the pace of urbanization increased, farm tenancy decreased, and civil rights of minorities were granted. In 1957 federal troops entered Little Rock to maintain order after the state militia had been ordered to prevent the desegregation of one of the city's high schools; the confrontation focused international attention on the state. Since then private enterprise has developed nationally outstanding state-based companies in retail merchandising, poultry products, computer software, and finance brokerage. (B.A.D.)

Florida

The geographic location of Florida has been the key factor in a long and colourful development, and it helps explain the striking contemporary character of the state. The greater part of Florida lies on a peninsula that protrudes southeastward from the North American continent, separating the waters of the Atlantic Ocean from those of the Gulf of Mexico and pointing toward Cuba and the Caribbean Sea beyond. Florida shares a border with only two states, Georgia and Alabama, both to the north. Its capital, Tallahassee, is located in the northwestern panhandle. The nearest foreign territory is the Bahamian island of Bimini, some 50 miles (80 kilometres) to the east of the state's tip. With the exception of Hawaii, Florida is the southernmost state of the United States, its northernmost point lying 100 miles farther south than California's southern border. The Florida Keys, a crescent of islands that forms the state's southernmost portion, lie within 1,700 miles of the equator. Florida, admitted as the 27th state in 1845, has an area of 58,664 square miles (151,939 square kilometres).

Strategic position

The state lies close to both the geographic and the population centres of the landmass of the Western Hemisphere, and its position not only commands one entrance to the Gulf of Mexico but also lies along a strategic crossroads between North and South America and historic routes to the European and Mediterranean worlds. Florida played a prominent role in the struggles of the European powers to control the New World, and it is fitting that St. Augustine, founded in 1565 on its northeastern coast, is the oldest European settlement within what were to become the boundaries of the continental United States.

Although agriculture and manufacturing continue to be important in Florida, the climate and scenery of the "Sunshine State" have attracted enormous numbers of visitors, and tourism is now a mainstay of a well-diversified economy.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Florida is a geologically young, low-lying plain, mostly less than 100 feet (30 metres) above sea level. The highest point is in Walton county, a mere 345 feet (105 metres) above sea level. Sedimentary deposits of sand and limestone cover most of the state, with areas of peat and muck marking locations where freshwater bodies

once stood. The contemporary topography has been largely molded by running water, waves, ocean currents, winds, changes in sea level, and the wearing away of limestone rocks by solution. These forces have produced enough variation in the state's surface to permit classification into seven basic physiographic regions: the coastal lowlands, the Lake Okeechobee–Everglades basin, the Kissimmee lowlands, the Marianna lowlands, the central highlands, the Tallahassee hills, and the western highlands, though these divisions are scarcely apparent to the naked eye.

The coastal lowlands occupy approximately 75 percent of the surface and vary in width from 10 to 100 miles. Generally, the region is very flat and is often less than 25 feet above sea level. Much of the area is swampy, and in the eastern part of the state numerous former beach ridges parallel one another. Offshore barrier beaches (bars) rim much of the region and account for most of Florida's excellent beaches.

The Lake Okeechobee–Everglades basin and the Kissimmee lowlands are actually subdivisions of the coastal lowlands, but their uniqueness justifies separate designations. The former is 150 miles long and 50 miles wide and is actually a shallow, slow-moving river (the "River of Grass"). The northern portion has been modified with canals, dikes, and pumping stations and is Florida's principal zone of sugarcane production. The southern portion retains much of its pre-European flavour and is protected within the confines of the Everglades National Park. The Kissimmee lowlands are approximately the size of the Lake Okeechobee–Everglades basin and include the broad valley of the Kissimmee River, the major source for water flowing southward to Lake Okeechobee. Much of this region is a flat grassland dominated by pastures and cattle ranches.

The Marianna lowlands is a small region in the northwestern panhandle, bounded on the east by the Apalachicola River and on the west by the Choctawhatchee River. The region is heavily eroded and has numerous sinkholes and caves.

The Marianna lowlands separates the Tallahassee hills, to the east, from the western highlands, which extend to the Alabama border on the west. Both of these regions are about the same size (40 by 100 miles), and both are ancient upland plains that have been dissected by streams. They form beautiful rolling country that, combined, is the most important zone for the production of field crops in Florida.

The central highlands lie between the Suwannee River, on the eastern side of the Tallahassee hills, and the St. Johns River, which separates the highlands from the eastern coastal lowlands. The region extends southward from the Georgia border to the area of Arcadia and Sebring, a distance of some 400 miles; the width varies from 50 to 75 miles. The region is rolling and dotted with thousands of lakes, and the central and southern portions contain most of Florida's citrus acreage.

Drainage. The flat Florida landscape is covered by a latticework of some 1,700 named streams (mostly in the north and northwest) and about 30,000 named lakes (mostly in the central region). The state also contains 17 of the 75 first-magnitude artesian springs in the nation, most of them located in the central region. There are some 39 drainage basins, with the Lake Okeechobee–Everglades basin (17,000 square miles [44,000 square kilometres]) being the largest. Lake Okeechobee (700 square miles) is the third largest freshwater lake entirely in the United States (after Lake Michigan and Iliamna Lake in Alaska). This vast water network is fed by the state's porous limestone substructure, which stores large quantities of water.

Climate. Climatically, Florida is divided into two regions. The tropical zone lies generally south of a west–east line drawn from Bradenton along the south shore of Lake Okeechobee to Vero Beach, while north of this line the state is subtropical. Summers are uniform throughout Florida. Freezing weather of short duration (but often crippling to agriculture) can occur as far south as Miami, but the Keys have never had frost.

Rainfall is heaviest in summer, with drier weather prevailing in the winter months. The average annual rainfall ranges from 40 inches (1,000 millimetres) in Key West to 62 inches in West Palm Beach. Snow falls occasionally in the northern areas and has been reported as far south as Miami. Hurricanes occur about once a year on the average, although Florida is no more vulnerable to hurricanes than are the other Gulf states or, indeed, the entire Atlantic coast as far north as Boston. The hurricane season is from June to November, though September is the month during which they are most likely to occur.

Hurricanes

Average annual temperatures show little variation, ranging from 68° F (20° C) in Tallahassee in the north to 77° F (25° C) at Key West in the south. Corresponding monthly averages range from the middle 40s F in the north to the middle 50s F in the south in January, and are in the lower 80s F in August.

Soils and plant life. In general, Florida's soils consist of sand, sandy loam, claylike marl, peat, and muck, but more than 300 soil types have been mapped. More than 300 species of trees have been identified; about half of the state is covered by forests. The dominant trees include pines, oaks, cypresses, palms, and mangroves. Many tropical trees thrive in southern Florida, while beech, red maple, sweet gum, tulip, magnolia, and hickory are common in the north. Almost half of the species of trees found in the United States grow in Florida. More than 3,500 other plants have been identified, including many plants imported into the state.

Six broad soil–vegetation regions may be described. (1) The flatwood lowland soils form the largest soil region in Florida, which corresponds to the coastal lowlands. The terrain there is level and underlaid by a hardpan that impedes drainage and encourages flooding; slash and longleaf pine, oak, sabal palm, and grass are typical vegetation. (2) Organic soils are found in many parts of Florida, particularly in the Lake Okeechobee–Everglades basin, where saw grass, cypress, sabal palm, myrtle, willow, elderberry, and gum are important vegetation. In this soggy environment submergence often prevents the oxidation, decay, and shrinkage of peat and muck, but when the soils are drained they deteriorate rapidly. (3) Southern limestone soils occur in the Kissimmee valley, the Big Cypress Swamp, and the Miami–Homestead area. Pines and oaks grow in some areas, but grasses, along with saw palmettos and sabal palms, predominate in the Kissimmee valley. The cypress, bay, and gumbo-limbo—a tall tree with a brown, brightly lacquered trunk—are typical of the extreme southern areas of this region. (4) Northern upland soils, from overdrained sands to well-drained loams, occur in the region stretching across the north of the state and support hardwoods, loblolly pine, and longleaf pine. (5) Northern slope soils, usually considered a distinct region, lie immediately to the south, with slash and longleaf pine, oak, and saw palmetto. (6) Central upland soils—with a vegetation similar to that found in northern slope soils—are located in the higher ridge area of central Florida, westward to the Apalachicola River. There are a number of other zones of soils and vegetation: dunes fringe the magnificent beaches of the state, while lush, dank mangrove swamps, along with tropical hardwoods and sand pine and oak, are found in the Ocala National Forest.

Major soil–vegetation regions

Animal life. Florida's rich and distinctive tropical and subtropical environment is inhabited by a huge and varied wildlife population; the rarer forms, such as panthers and crocodiles, are fully protected. Approximately 100 species of mammal are found in the state, including deer, puma, bobcat, boar, black bear, armadillo, otter, mink, and gray fox; smaller animals are also numerous. Manatees (sea cows) are found along the coast and in warm inland waters, and several species of porpoises and dolphins lend their distinctive charm to the clear coastal waters.

Huge and varied wildlife population

More than 400 species and subspecies of birds have been identified; land birds include the turkey, quail, dove, eagle, hawk, owl, and most smaller birds common to the southeastern states, while such coastal birds as the gull, brown pelican, sandpiper, osprey, and cormorant are also numerous. Freshwater and marsh birds include the gallinule (marsh hen), duck, goose, coot, egret, heron, ibis (stork), and flamingo. There are vast natural rookeries in the Everglades, and numerous wildlife refuges are maintained for the protection of migratory birds.

Heron and an American alligator, in the area of Shark Valley, Everglades National Park, Florida.
© Connie Toops

The alligator is the king of Florida reptiles, its role as a builder of water holes being vital to southern Florida's ecology, while the crocodile still inhabits part of the Everglades National Park. The 40 species of snakes in the state include the nation's four poisonous types: the coral, rattlesnake, moccasin, and copperhead (the latter restricted to limited areas of northern Florida). Turtles, tortoises, lizards, and frogs are also abundant.

Florida's 4,000,000 acres (1,600,000 hectares) of water (of which some 2,800,000 acres are inland) contain more than 700 species of fish and shellfish. Common saltwater varieties are bluefish, pompano, flounder, mackerel, mullet, trout, redfish, snapper, grouper, snook, sailfish, tarpon, shad, weakfish, bonefish, marlin, and shark. Others include crawfish, oysters, stone and blue crabs, clams, and shrimp. The largemouth black bass is the state's foremost freshwater species, while others include bream (bluegill), sunfish, speckled perch, and catfish.

Settlement patterns. Northern and southern Florida are often distinguished as separate regions. Northern Florida is generally a colder, historically older, rural and hilly area, oriented toward field agriculture and forestry. Southern Florida is a warmer, flat, urban area, the more recently settled region of the state, with an economy based on tourism, citrus fruits, vegetables, and livestock. A refinement would add central Florida, with its lake region and citrus belt. A more scientific classification might identify the east coast, the west coast, central Florida, the Keys, the Everglades, and the panhandle. More popular regional names are the Gold Coast (the Miami–West Palm Beach metropolitan sprawl), the Sun Coast (the Tampa Bay area), and the Big Bend (centring on Tallahassee). Lesser-known designations include the Silver or Platinum Coast (a term applied to the lower southwest), the Island or Mangrove Coast (the extreme southwest), Suwanneeland (the Suwannee River basin), Miracle Strip (the upper northwestern coast), Florida's Crown (the northernmost part of the peninsula), the Surf Coast (the middle east coast), and the Tropi-Coast (Miami–West Palm Beach).

To a stranger entering the state from the north, the Florida landscape may appear devoid of human imprint. It is being used, but the use is of a type unfamiliar to many visitors. More than half of Florida is covered by commercial, national, and state forests, state and federal parks, lakes, beaches, and military reservations. Less than two-fifths of the state is farmland, and about one-third of this is in either pasture or timber. Only about one-fifteenth of Florida's total land is used for harvested crops.

The effect on the landscape is striking: farmsteads are generally common in northern Florida, where field crops are important, but even there timber covers thousands of acres. Citrus groves occupy much of central Florida and the east coast, while along the west coast and north and south of the citrus belt spread vast expanses of cattle land. Around Lake Okeechobee, the cultivation of sugarcane and vegetables has produced the modern equivalent of plantation agriculture. The small, private farm has little place in these systems, having been superseded by mechanization and the use of migratory labour. The social conditions of migrant workers, which have occasionally given rise to national concern, remain one of the negative aspects of a generally affluent state. The rise of corporate agriculture has led to an inevitable increase in farm size and a corresponding reduction in farm numbers. Agricultural patterns

Even the small town seems to be disappearing, as more and more people find it relatively easy to commute from urban and suburban centres. The great majority of the people live in urban areas, and only a tiny percentage live on farms. The densest concentration is along the extensive Miami–Fort Lauderdale–Boca Raton–West Palm Beach urban complex in the southeast. This area appears to many observers to be duplicating the less desirable aspects of the great urban belts burgeoning in other parts of the nation. On the west coast the Tampa–St. Petersburg metropolitan area contains another concentration of population. Farther north the Daytona Beach–Cape Canaveral–Orlando triangle is central Florida's dominant urban area; Jacksonville is the major centre of the upper east coast and southeastern Georgia; and Pensacola dominates the western panhandle and part of Alabama. Lesser metropolitan areas—including Tallahassee, Gainesville, and Fort Myers—are hubs of local influence.

The people. *Ethnic groups.* Ancient Indian groups entered Florida from the north as early as 10,000 years ago, but farming did not appear much before 500 BC, and some southern Indian groups remained hunters, fishers, and gatherers until their extinction. Indians continued to arrive in small numbers after 500 BC, and contacts with Cuba, the Bahamas, and, possibly, the Yucatán reflected Florida's unique situation. By 1750 virtually all of these peoples had been destroyed by disease, slavers, and wars, largely the responsibility of English and Creek raiders from Georgia. The latter, accompanied by a few runaway black slaves and renegade whites, were collectively called Cimarrones. The name Seminole evolved from *cimarrón* (Spanish: "wild, unruly, runaway"). There were approximately 5,000 Seminoles in Florida when it came under formal U.S. jurisdiction in 1821. Within 25 years this population had declined to about 150, by which time most of the Seminoles had been removed to the territory of Oklahoma. By the late 20th century a small number of descendants of those who had remained were living in three reservations in southern Florida. Contacts with the Caribbean

The first Spanish settlement, established in 1559, was abandoned in 1561, but San Augustín (St. Augustine), founded in 1565 after the destruction of the French settlement nearby, became the first permanent European settlement in the United States. For the next 250 years Florida was little more than a wilderness in terms of any permanent European settlement, though its importance as a historical pawn was considerable.

A large population of European origin did not develop until the United States established effective civil control in 1822, and the great increases of the late 19th and 20th centuries owed much to economic factors. Immigrants from northern Spain came to Tampa at about the time of World War I, drawn largely by the expanding cigar industry and the prospect of living in a Spanish-speaking community. Italians also came in large numbers after World War I.

Tarpon Springs was settled about 1880, and by 1905 Greek immigrants, drawing on the traditions of their homeland, had established the nation's major sponge industry. Greek is still widely spoken, and the city is a centre for the Greek Orthodox religion. Other ethnic contributions lending character to the overall population of the state range from a Jewish community at Miami–Miami Beach to a Slovak settlement at Masaryktown.

It is not known when the first blacks arrived in Florida, but it is known that some accompanied the first Spanish expeditions. A few runaway slaves came with the Seminoles, but it was only with U.S. rule that the black population began to increase. By 1830 there were as many blacks as whites (about 11,000). The increase of the black population coincided with the development of the Southern plantation system. The Civil War overthrew slavery, but the agricultural patterns remained, and not until the end of the 19th century did an influx of new settlers cause the white population to increase faster than that of the blacks. The black percentage of the total population continued to decline in the 20th century, though large percentages are still found in the old plantation belt (north central Florida) and in the Everglades truck-farming region.

Cubans in Florida

Cubans came to Key West after 1868 when, as a result of revolutionary turmoil in Cuba, Vicente Martínez Ybor moved his cigar factories there from Havana. Labour troubles and a disastrous fire encouraged Ybor to move again in 1886, this time to Tampa, and again many Cubans followed the factories. A similar influx occurred after the Cuban revolution of the early 1960s, when more than 350,000 Cubans fled their homeland. A third of these settled in Florida (mostly in the Miami area) during the decade. Still another wave of Cubans arrived in Florida in 1980, and about 80,000 were integrated (albeit uneasily) into the Cuban community around Miami.

Demography. The continued migration of older, retired people into the state has given Florida the nation's oldest population. The northwestern counties continue to grow more slowly than the southern urban areas. Those areas heavily dependent on military and space programs continue to be directly affected by federal budgetary decisions that influence such activities. Key West, for example, a Navy-oriented city, suffered a population loss of more than 35 percent between 1960 and 1980. Cutbacks in the space program created one of Florida's highest unemployment rates in the Cape Canaveral area.

The economy. Florida experienced little economic development before 1821, when the United States took formal possession. The 60 years that followed were dominated by small-farm and plantation agriculture; the supplying of naval stores and the production of beef and hides, pork, salt, tobacco, and cotton were the main activities. The 1880s marked the beginning of a new era in Florida. In 1881 phosphate—the state's most important mineral—was discovered in the Peace River valley, while in that same year Hamilton Disston, a Philadelphia industrialist, bought 4,000,000 acres in the Everglades for 25 cents an acre. This freed the state from its post-Civil War debt and opened the way for the development of much of the peninsula. In the west a railroad reached Pensacola in 1883, and in the following year another reached as far as Tampa; the line was financed by a Northern capitalist, Henry B. Plant. On the east coast his counterpart, Henry M. Flagler, was building a rail and hotel empire that soon extended past Miami to Key West. Agricultural development, settlement, industry, and tourism all followed the rails. The 1890 population of the state was double that of 1870, and the total passed 500,000 by the turn of the century and continued its spectacular growth thereafter.

Role of Northern capital

Labour–management relations in Florida tend to follow the national pattern. Manufacturing, mining, communication, and transportation tend to be unionized; services are less so, while the professions and agriculture are generally nonunion. The cooperative movement is especially strong in citrus growing and commercial fishing. The Florida Citrus Mutual works closely with growers and the state in research, production, processing, marketing, and regulation, while there are several fishing cooperatives, primarily to help market the catch. The state government cooperates closely with the private sector of the economy; it is instrumental in inspection, licensing, regulation, and research and education, but the state rarely competes with private business.

Resources. Directly or indirectly, Florida's tropical and subtropical climate affects almost every aspect of the local economy, and it can be quite justifiably considered the state's chief resource. Together with land and water—both of which have a rich potential for economic development—climate forms the basis of the state's wealth. The water resources, important to the fishing industry and tourism alike, include not only 4,424 square miles of fresh inland water but also an even larger area of adjacent salt water. In the United States only Alaska has a tidal shoreline whose length exceeds that of Florida, which totals 8,426 miles (Gulf coast, 5,095 miles; Atlantic coast, 3,331 miles). On land, forestry activities are supported by the half of the state (mostly in the north) that is wooded, and livestock raising is supported by the state's large grasslands (mostly in the central and southern areas).

Florida also yields several important minerals. Phosphate, which is used in fertilizer and livestock feed and by the chemical industry, is found in the west central portion of the state, and production there accounts for some three-fourths of the national total. Titanium, zircon, and such other important heavy minerals as thorium and cerium are mined near Jacksonville, Starke, and Vero Beach and in west central Florida. Petroleum is produced in the northwest and the southwest. Kaolin is mined in Putnam county; fuller's earth comes from the Tallahassee region; and clay, sand, and gravel are mined in numerous locations, with pure silica sand extracted mostly in areas around the 100-foot contour line. Limestone, from the northern portion of the peninsula, is used as building stone and road-surfacing material and in cement, concrete, and fertilizer; peat, used as a soil conditioner, is dug in many areas. The versatility of the marine resources of Florida is indicated by a plant at Port St. Joe on the Gulf coast for the recovery of magnesium from seawater.

Production of citrus fruit

Agriculture. Florida produces about three-fourths of the nation's citrus fruit and is second only to California in vegetable production. Citrus fruits, vegetables, and livestock each account for about 25 percent of farm receipts. Tomatoes are the leading vegetable crop. Sugarcane is the king of Florida field crops, and the state produces approximately 40 percent of the nation's total. Florida is also well known for its thoroughbred horses raised near Ocala.

Industry. As might be expected in a state dominated by tourism, the service industry is the most important employer, accounting for some 30 percent of all Florida jobs. Another 20 percent is in retail trade, also closely associated with the tourist industry. Manufacturing ranks third in employment but on a value-added basis exceeds all sectors but tourism. The electrical and electronics sector leads in manufacturing, followed (in order of importance) by printing and publishing, transportation equipment, food processing, machinery, metal fabrication, clothing, chemicals, and wood products. Together the manufacturing industries employ some 13 percent of the labour force. Other major industries in Florida include construction, finance, government, transportation, communications, and wholesale trade.

Tourism. Tourism is the largest income-producing activity in the state and has developed into a year-round business. The Atlantic and Gulf beaches that form the basis of the state's popular "vacationland" image are seldom the only destination for the Florida visitor. Other attractions include the large theme parks, professional and collegiate sporting events, golf, hunting and fishing, and an abundance of parkland, including two national parks, two national seashores, a national preserve, five national monuments and memorials, and numerous state parks.

Transportation. Florida's transportation system is comprehensive, covering the entire state except for certain isolated areas in the Everglades. In general, highway arteries run across the north of the state, from Jacksonville to Pensacola; down the east coast, from Jacksonville to Miami; diagonally across the state, from Jacksonville to Tampa–St. Petersburg on the west coast, bisecting the state from Tampa–St. Petersburg to Daytona Beach; and through the southwestern portion, linking Tampa–St. Petersburg to Miami. Rail and air traffic also follow these patterns. Although there is no direct Miami–Tampa railroad, there is a heavily traveled air route between the two cities.

The most heavily traveled throughways are the interstate and state turnpike systems that connect all major cities. There are several airports with regularly scheduled flights

and numerous private airfields; the international terminals at Tampa, Orlando, and Miami are among the most modern in the world. An extensive rail network provides passenger and freight service to most areas. An integrated system for domestic and foreign shipping is provided by eight major deepwater ports and several lesser ports and harbours, while more than 1,000 miles of navigable channel are maintained by the federal government.

Cape Canaveral

Florida has a spaceport, the John F. Kennedy Space Center at Cape Canaveral, which occupies 88,000 acres. This is not only a major Florida industry but has also become a prime tourist attraction.

Administration and social conditions. *Government.* The basis of Florida's government was laid in the constitution of 1885, to which, through the years, about 100 amendments were ratified. Twenty articles of the constitution were formally revised in 1968 and ratified by the people in the 1968 general election. There are a number of important constitutional prohibitions and tax exemptions, including prohibition against an income tax and inheritance tax and exemptions that relate to homesteads.

The executive department comprises the governor and an independent cabinet of six officers (secretary of state, attorney general, treasurer and commissioner of insurance, comptroller, commissioner of education, and commissioner of agriculture). The governor and members of the cabinet each serve four-year terms, with a consecutive limit of two. The legislature is composed of the Senate of 40 members and the House of Representatives of 120 members. Senators serve four-year terms and House members serve two-year terms.

Judiciary powers are exercised through courts established by the constitution: the Supreme Court, five district courts of appeal, 20 circuit courts, and 67 county courts. Appellate judges are initially appointed by the governor; other judges are elected to office on a nonpartisan ballot.

Political trends

Although Florida has become a two-party state in some respects, Democrats continue to outnumber Republicans in the legislature. Top-level state executive offices and seats in the U.S. Congress are also most often held by Democrats. Nevertheless, Florida has elected Republican governors since the 1960s and typically votes Republican in presidential elections.

Florida has no state income tax on individuals but does tax corporate income, which produces about 9 percent of the total tax revenue. Approximately two-thirds of the total comes from the general sales and use tax. The remainder derives from special taxes on a wide range of items, including unemployment compensation, gasoline, motor vehicles and mobile homes, alcoholic beverages, cigarettes, utilities, insurance premiums, pari-mutuel wagering, and inheritances. A state lottery, established in 1988, offered a new source of revenue.

Education. Florida supports the public school system to the extent that no child is deprived of a minimum standard of education, while the counties are expected to supplement this minimum. Virtually the entire population of Florida is within commuting distance of a state-supported college or university, part of an extensive system of higher education. There are nine state-supported universities, several of which offer courses on multiple campuses. In addition to the university system, Florida maintains several community (junior) colleges, all of which offer the first two years of university-equivalent courses in addition to terminal programs in technical areas. Private sources support other institutions of higher learning, ranging from small, highly specialized schools to major universities, such as the University of Miami.

Health and welfare. Socially, Florida regards itself as a progressive state, and a major proportion of the state's financial resources go into those areas that serve the public, with education, welfare, health, and hospitals receiving about two-thirds of the total appropriations.

The Department of Health and Rehabilitative Services (HRS) is charged with programs of direct assistance to the public. Included are aid to families with dependent children, old-age assistance, aid to the blind, and aid to the disabled. HRS is also responsible for child welfare, including child neglect and abuse. Florida's expenditure per capita for health and hospitals exceeds the U.S. average and is among the highest of those of the heavily populated states. Public welfare payments, on the other hand, are among the lowest in the nation. Florida ranks at the bottom of all Sunbelt and populous states in this respect. In part, this is a reflection of the public's refusal to offer welfare to the needy of other states who seek refuge in Florida's tropical and subtropical environment.

Florida's favourable climate and geographic position has led to two types of migration: elderly persons come to Florida from the North, and political and economic refugees enter the state from Latin America. Both of these movements have severely taxed the state's ability to support the needy. A large percentage of the population of Florida is over 65 years of age, and the population is continuing to become older. This development requires that numerous support programs be increased. There is, however, also a burgeoning young population that has resulted from the mass immigration of peoples from Latin America. This group makes additional, but different, demands on the state for education, family assistance, and public safety.

In total personal income, Florida ranks among the top states nationwide and ranks relatively high in income per capita. There are no serious labour shortages in Florida except in a few highly skilled and professional occupations and, at certain times, in service occupations. Unemployment in Florida usually runs below the national average, and the diversified economy of the state has not been as subject to labour fluctuations as in many other areas where one industry dominates the economy. The cost of living in Florida is generally below the national urban average, and in some areas, such as clothing, it is considerably below the average—another small but significant benefit accruing from the favourable climate.

Employment and living costs

Cultural life. Florida is well endowed with a variety of cultural activities and institutions, a situation stemming partly from the importance of tourism and partly from the increasing leisure time available to the growing number of retired residents. The state itself maintains more than 800 parks and other areas, many of historic or natural interest. The Everglades National Park contains 1,400,533 acres in the heart of a unique natural region. Florida's rich history is preserved in such places as St. Augustine, the nation's oldest town, portions of which have been restored, while its famous fort, Castillo de San Marcos, has been made a national monument.

Sarasota is a centre for both art and theatre. The John and Mable Ringling Museum of Art possesses an internationally famous collection, and the city also contains a circus museum and hall of fame and the only 18th-century Italian theatre in the United States—the Asolo. The Florida State Museum is located in Gainesville.

Among the numerous festivals that fill the cultural calendar are the Greek Orthodox Epiphany (Tarpon Springs; January), the Orange Bowl Festival (Miami; December–January), the Florida Citrus Festival (Winter Haven; February), the Strawberry Festival (Plant City; March), the Festival of States (St. Petersburg; March), the Arcadia Rodeo (Arcadia; March and July), and the Fiesta of Five Flags (Pensacola; June). The Gasparilla Pirate Invasion, a festival comparable to the Mardi Gras celebration in New Orleans, La., is held in Tampa in February, in association with the state fair.

Sports are well represented in Florida, with the Orange Bowl and Joe Robbie Stadium in Miami, the Gator Bowl in Jacksonville, and Tampa Stadium attracting major university and professional events. The National Basketball Association has teams in Miami and Orlando, and the National Football League is represented by teams in Tampa Bay and Miami. The Basque sport of jai alai enjoys great popularity in the state's urban areas. The mild climate allows Florida to host outdoor sporting events year-round. Many major-league baseball teams hold spring training in the state, and there are major golf tournaments, as well as internationally known auto races at Daytona Beach and Sebring. Horse racing is also important.

Commercial attractions proliferate in Florida, many of them educational as well as entertaining. Walt Disney

World, near Orlando, is the largest single tourist attraction in the country; the facility's Experimental Prototype Community of Tomorrow (EPCOT) is a permanent international showcase. Nearby Sea World conducts important marine research with sharks, porpoises, and whales, as does Marineland (near Daytona Beach) and the Seaquarium in Miami. At Busch Gardens in Tampa, visitors observe hundreds of African animals in an open environment.

In addition to other cultural offerings, Florida's universities and educational television stations provide broad programs in continuing and adult education. Several cities maintain symphony orchestras, and there are a number of major performing arts centres that host concerts, Broadway shows, and other events.

HISTORY

Exploration and settlement. The early history of Europeans in Florida reflects the conflicts of the Spanish, French, and English crowns for empire and wealth. Juan Ponce de León's quest for the Fountain of Youth brought him to the peninsula in 1513 and 1521. Because he landed on the peninsula during the Easter season (Spanish: *Pascua florida* [season of flowers]) and because of the vegetation he found there, Ponce de León named the area Florida. Ponce de León made no attempt to found a settlement in 1513. In fact, he was under the impression that Florida was one of the islands in the Bahamas archipelago, and he does not seem to have ventured much north of present-day West Palm Beach. After an intermission of eight years, Ponce de León returned to establish a colony in the vicinity of modern Fort Myers. He was mortally wounded there in 1521 by the Calusa Indians and died the same year in Havana.

Spanish inroads and setbacks

In 1528 Pánfilo de Narváez landed on the shores of Tampa Bay with more than 400 men, with the intent of learning how this land was connected to Mexico. Within a year, and while still no closer to Mexico than northern Florida, the force was reduced to 15 survivors. Of this group, four Spaniards—including Álvar Núñez Cabeza de Vaca and Estebán, a Moorish slave who was the first black man known to enter Florida—reached Culiacán, Mexico, in 1536. Hernando de Soto came in 1539, landing somewhere between Fort Myers and Tampa, and led a devastating expedition through western Florida. Almost 20 years elapsed before Tristán de Luna y Arellano attempted to set up a colony at Pensacola Bay. It was abandoned in 1561, following devastation by a hurricane. In 1564 a group of French Protestants (Huguenots) who originally had been led by Jean Ribault established Fort Caroline on the banks of the River of May (St. Johns River), near modern Jacksonville. This group was seen by the Spaniards as a threat to their sea-lane from Havana to Spain. An expedition commanded by Pedro Menéndez de Avilés massacred most of the French colony in 1565 after founding St. Augustine nearby.

Shifting alliances and allegiances. During the following centuries there were frequent raids by English seafarers, including Sir Francis Drake in 1586, and clashes with French colonizers along the northern coasts of the Gulf of Mexico and with English settlers in the Carolina and Georgia colonies. Shifting alliances among the three powers reflected the vicissitudes of European politics, and St. Augustine and the English ports of Savannah and Charleston to the north of Florida were besieged at various times throughout the first half of the 18th century.

England received Florida in return for Havana in 1763 and replaced its military government with civilian officials. Expenditures for economic development brought prosperity as well as loyalty from most Floridians during the U.S. War of Independence, when the area was used as a base for attacks on colonial coastal cities. Three decades of political and social instability followed Florida's return to Spain after the war, with U.S. expansionist interests in constant conflict with the Spanish presence. Pensacola was a base for the British during the War of 1812, when Indians and runaway slaves were employed to harass U.S. settlements. The First Seminole War (1817–18) was the beginning of armed conflict between the U.S. government and the Indians in Florida. General Andrew Jackson's capture of Pensacola led to the cession of Florida to the United States in a treaty signed in 1819 and ratified in 1821. But the following territorial years were difficult and included the Second Seminole War (1835–42), the most costly of the conflicts arising from Indian resistance to giving up their valuable land.

Statehood. By 1845, when Florida was admitted as a state, all but a few hundred Seminoles had been removed to the territory of Oklahoma. The Third Seminole War (1855–58) was the final conflict with the Seminoles remaining in Florida.

Florida was a part of the Confederacy during the American Civil War, but action in the state was mostly limited to the capture of coastal cities by Union troops. In the 1880s an era of railroad building began that opened the state to permanent settlers and a new tourist trade and that brought it into the heart of the nation's economic life as the winter vegetable and citrus centre of the East.

Evolution of the contemporary state

The 20th century. The growth of Florida in the 20th century had been frantic, if not chaotic. Land booms and the burgeoning population have paralleled economic diversification and Florida's increasingly international focus. The state became the nation's fourth most populous in the late 1980s, not only welcoming a flood of new residents from northern states and Canada but also accommodating hundreds of thousands of immigrants from the Caribbean area. Miami has become the economic "capital" of the Caribbean, a city where English is now virtually a second language. Floridians take most of these developments in stride, though the problems of rapid growth have resulted in pressure on the natural environment and have taxed the state's social resources. Ironically, the nation's oldest region of European settlement has once again become a frontier. (R.H.Fu.)

Georgia

The largest of the U.S. states east of the Mississippi River and by many years the youngest of the 13 former English colonies, Georgia was founded in 1732, at which time its boundaries were even larger—including much of the present-day states of Alabama and Mississippi. Its landscape presents numerous contrasts, with more soil types than any other state as it sweeps from the Appalachian Mountains in the north (on the borders of Tennessee and North Carolina) to the marshes of the Atlantic coast on the southeast and the Okefenokee Swamp (which it shares with Florida) on the south. The Savannah and Chattahoochee rivers describe much of Georgia's eastern and western boundaries with South Carolina and Alabama, respectively.

Georgia has an area of 58,910 square miles (152,576 square kilometres). For most of the 19th century the state was the capital of the cotton empire of the South, but poultry products now account for many times the income from cotton. Although industry has far outstripped agriculture in economic importance, a high proportion of industrial workers remain in farm- or forestry-related jobs such as lumber products, food processing, and textile manufacture.

Atlanta, the capital, has long been the economic and cultural centre of the Southeast. Its name evokes the largely romantic legends of the pre-Civil War South, of the traditions of Southern gentility, and of white-columned mansions along Peachtree Street, its best-known thoroughfare. The history of the state is marked by events of the Civil War: the many major battles fought there, the Confederate prison at Andersonville, in which nearly 13,000 Union prisoners died, and the burning of Atlanta and the devastating March to the Sea by Union forces under General William Tecumseh Sherman.

The degree to which the wounds of this history have been healed in Georgia is most strikingly exemplified in modern Atlanta. Since the 1960s, black citizens have played an increasingly important role in the city's administration; the first black mayor was elected in 1973. During the same period, Atlanta has become a nationally oriented city, attracting major corporations as well as citizens from all parts of the United States.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The southernmost portions of the Blue Ridge Mountains cover northeastern and north central Georgia. In the northwest a limestone valley-and-ridge area predominates above Rome and the Coosa River. The higher elevations extend southward about 75 miles (120 kilometres), with peaks such as Kennesaw and Stone mountains rising from the floor of the upper Piedmont. The highest point in the state, Brasstown Bald in the Blue Ridge, reaches to a height of 4,784 feet (1,458 metres) above sea level. Below the mountains the Piedmont extends to the fall line of the rivers—the east-to-west line of Augusta, Milledgeville, Macon, and Columbus. Along the fall region, which is nearly 100 miles wide, sandy hills form a narrow, irregular belt. Below these hills the rolling terrain of the coastal plain levels out to the flatlands near the coast, the "pine barrens" of the early days.

Surface features

Drainage. About half the streams of the state flow into the Atlantic Ocean, and most of the rest travel through Alabama and Florida into the Gulf of Mexico. A few streams in northern Georgia flow into the Tennessee River and then via the Ohio and Mississippi into the Gulf. The river basins have not contributed significantly to the regional divisions, which have been defined more by elevations and soils. The inland waters of Georgia consist of some two dozen artificial lakes, about 70,000 small ponds created largely by the federal Soil Conservation Service, and natural lakes in the southwest near Florida. The larger lakes have fostered widespread water recreation.

Because of the region's bedrock foundation, Piedmont communities and industries must rely on surface runoff for their primary water supply. The Coastal Plain, underlain by alternating layers of sand, clay, and limestone, draws much of its needed water from underground aquifers. Increasing domestic and industrial use of underground water supplies in Savannah, St. Marys, and Brunswick threaten to allow brackish water to invade the aquifers serving these coastal cities.

Soils. From the coast to the fall line, sand and sandy loam predominate, gray near the coast and increasingly red with higher elevations. In the Piedmont and Appalachian regions these traits continue, with an increasing amount of clay in the soils. Land in northern Georgia is referred to as "red land" or "gray land." In the limestone valleys and uplands in the northwest, the soils are of loam, silt, and clay and may be brown as well as gray or red.

Climate. Maritime tropical air masses dominate the climate in summer, but in other seasons continental polar air masses are not uncommon. The average January temperature in Atlanta is 42° F (6° C); in August it is 79° F (26° C). Farther south, January temperatures average 10° F (6° C) higher, but in August the difference is only about 3° F (2° C). In northern Georgia rain usually averages from 50 to 60 inches (1,270 to 1,524 millimetres) annually. The east central areas are drier, with about 44 inches. Precipitation is more evenly distributed throughout the seasons in northern Georgia, whereas the southern and coastal areas have more summer rains. Snow seldom occurs outside the mountainous northern counties.

Plant and animal life. Because of its mountains-to-the-sea topography, Georgia has a wide range of natural vegetation. It ranges from maple and hemlock and birch and beech near Blairsville in the north to the cypress, tupelo, and red gum of the stream swamps below the fall line and to the marsh grass of the coast and islands. Throughout most of the Appalachians, chestnut, oak, and yellow poplar are dominant. Much of this area is national forest. The region that extends from the Tennessee border to the fall line has mostly oak and pine, with pines predominating in parts of the west. Below the fall line and outside the swamps, vast stands of pine—longleaf, loblolly, and slash—cover the landscape. Exploitation of these trees for pulpwood is a leading economic activity. Much of the land, which had at one time been cleared of trees for agriculture, has gone back to trees, scrub, and grasses.

Variety of natural life

Georgia's wildlife is profuse. There are alligators in the south; bear, with a hunting season in counties near the mountains and the Okefenokee Swamp; deer, with restricted hunting in most counties; grouse; opossum; quail; rabbit; raccoon; squirrel; sea turtles, with no hunting allowed; and turkey, with quite restricted hunting. In general, wildlife is in a period of transition. Deer have been seen in suburban counties and bears on golf courses near Atlanta, but solid stands of pine and unbroken pasture are not ideal for wildlife. There is extensive stocking of game birds and fish. The major fish of southern Georgia, except snook and bonefish, are in waters off the coast, and all major freshwater game fish of the United States are found in Georgia's streams and lakes. Some 50 species of plants and 20 species of animals are listed as endangered in the state.

Settlement patterns. Migrations and historical change have tended to blur the traditional regions of Georgia. In an earlier time the coastal mainland and the sea islands were distinct, being separated from the middle part of Georgia by the pine barrens of the lower coastal plain. Rice and sea island cotton were major crops of this area. There was usually a summer exodus of plantation families to the southern highlands and to the North: a Savannah family, for example, might well know more people in Boston, New York City, or Philadelphia than they did in Augusta, Macon, or Columbus. The coastal region and the sea islands of Georgia and South Carolina still support the unique culture of the Gullahs (also called Geechees), blacks who speak a creole language based on colonial English and various West African languages.

Early subsistence farmers, who were known as "crackers" (probably meaning boasters or liars), tilled the land in the thinly populated pine barrens just above the coastal area. There the soil is sandy and poorly drained, and the farms and plantations existed amid near-frontier conditions throughout much of the 19th century.

Beyond the pine barrens to the north and west, a prosperous cotton culture flourished for a period of more than a century. The classic period of cotton plantations and farms before the Civil War was brief in several counties. The white-columned houses generally were built in the county seats, or "courthouse towns," whereas the majority of the rural homes could be termed no more than substantial farmhouses. The mountainous area of northern Georgia was an area of small subsistence farms and few slaves.

Blacks were long identified with commercial agriculture, which produced rice and sea island cotton on the coast and upland cotton in middle Georgia. Some modern historians use the term Black Belt as a demographic description of middle Georgia, where slavery predominated, though its origin stemmed from the dark soils of Alabama.

Plantations were cultivated by supervised group labour before the Civil War. After the war a family-plot system, called sharecropping or tenant farming, replaced the larger labour groups and reduced immediate supervision. There were two basic arrangements in the sharecrop system with some variations: in middle Georgia the division-by-halves system gave the landowner control of the farm's management and the sale of the crops that were raised. The more independent landowner of northern Georgia, if he did not own a small farm himself, rented by thirds and fourths and had more control.

Land divisions and town growth

Rural Georgia was settled in a pattern of separate farms without unified communities. Area names suggested such centres, but these generally were derived from the names of creeks, mountains, and militia districts. Schools, churches, and stores often drew neighbours in different directions, producing a highly diffused community life. Urban settlements originally served political and commercial purposes as county seats and cotton markets. The fall line cities became railroad points, and later they and some courthouse towns in the upper Piedmont had cotton mills.

Elsewhere the country general store and small local cotton gins declined as the larger towns gradually absorbed the slight commerce and industry. Later, better roads enabled people to travel from these smaller locales to the larger towns and cities. Today shopping centres, neon-lighted restaurants, and service stations are scattered throughout both the rural and urban areas of the state.

The people. In 1752, the year Georgia came under direct rule of the British government, it had only about 3,000 inhabitants, most of them either English or black.

Salzburgers from Austria lived at New Ebenezer and Savannah, Scottish Highlanders at Darien, and New England Congregationalists at Sunbury and Midway. Settlement was concentrated on the coast and up the Savannah River to the Augusta area. Irish, German, and other massive immigrations into the United States during the 19th century affected Georgia little, and by 1910 only about 15,000 foreign-born whites resided in the state.

Patterns of migration

Georgia lost more people than it gained through migration in each decade from 1870 to 1960, but the total population increased steadily due to high birth rates. More whites than blacks left until 1910, generally moving to Southern states to the west; after 1910 the black exodus was greater, generally to the cities of the North. The boll weevil plague of the 1920s devastated the cotton economy and caused massive departures of both races. The white emigration loss almost stopped in the 1950s; thereafter there was a net migration gain of both races.

Mechanized and chemically controlled agriculture, along with abandonment of cultivated farmlands to pastures and pine forests, were major causes of black migrations. Perhaps as important, however, were blacks' hopes that elsewhere they would find fewer ingrained patterns of discrimination.

The economy. Georgia's rapid industrialization after 1940 caused a shift in the meaning of such words as "farm" and "industry," and "agribusiness" has become a term in wide use. More people work in industry than on farms, but a great majority of industrial jobs depend on farm or forest products. Income from poultry is regarded as farm income despite the fact that chickens are raised in factory-like structures; feed prices are quoted for 150-ton lots, and financing is done by complex base-and-incentive arrangements—all indicating an extensive and highly developed agribusiness.

More conventional industries, such as automobile assembly, exploit the nearness of Atlanta to southeastern markets. Several decades of prosperity have also increased commerce, and a large number of New York City stores have branch stores in Atlanta and other Georgia cities.

The federal government affects Georgia's economy through direct purchases from industry, but more through payrolls at the several major military installations in the state. It produces some hydroelectric power and regulates its sale, much of it to rural electric cooperatives.

The state government, on the other hand, functions in the economic sphere largely to promote further industrial development or financial investment in the state, which continues to rely to a large extent on outside money. Atlanta is the financial centre of the Southeast and the headquarters of the Sixth District of the Federal Reserve Bank. More than 60 percent of state revenue comes from taxes levied on sales, licenses, and personal and corporate incomes. Taxes on personal income account for more than one-third of all tax revenue. Local governments in Georgia rely mainly on general property and sales taxes.

Following the pattern in most Southern states, membership in labour unions in Georgia remains below the national average. Given the traditional regional antipathy toward unionization and Georgia's consistently lower-than-average unemployment rates, this condition appears unlikely to change in any significant degree in the future.

Quarrying and forestry

Resources. Georgia is one of the nation's major producers of building stone and crushed stone. Pickens county in northern Georgia has one of the richest marble deposits in the world. Kaolin is taken from vast pits in middle Georgia, and the state is the major producer in the United States. Phosphate deposits in the southern part of the state are largely unexploited, but this region's ample reserves of artesian well water are proving useful for agricultural irrigation.

Georgia's virgin timberlands have been cut over, but the state ranks first in the nation in number of acres of commercial forest. Taxation of the state's timber-growing lands is an internal political issue, with growth rates versus tax rates a crucial argument. Lumber, plywood, and paper are major products, but Georgia is especially known for its large production of naval stores from its pine forests.

Georgia lies to the south of the states that benefit from the many hydroelectric dams of the Tennessee Valley Authority, and waterpower contributes only about 6 percent of its electrical energy. Petroleum, natural gas, and butane pipelines come into the state from the Southwest. Coastal waters provide working grounds for a number of Georgia fishermen, with shrimp and crab being the main catches.

Loading slash pine, to be used for pulpwood, near Fitzgerald in the Coastal Plain area of south central Georgia.
Larry Lefever from Grant Heilman

Industry. In terms of manufacturing employment, the most important of Georgia's industries are textiles and apparel, food processing, transportation equipment, electrical and electronic equipment, and paper and lumber. Cotton textile production has occupied a major sector of Georgia's economy since the late 19th century. The continuation of specialization in textiles is shown in the great number of rug and carpet mills in northern Georgia. The concentration of looming and weaving skills there make the state one of the major textile producers in the nation. Manufactured items include airplanes, automobiles, mobile homes, chemicals, and processed foods.

Agriculture. With the continuing consolidation of farms into fewer but larger units and the advent of a pervasive agribusiness, Georgia has followed nationwide trends in agriculture. Much of the poultry industry is conducted by large companies that parcel out their work to small farmers and supply them with modern poultry-raising facilities. Cattle and swine raising are important, especially in the southern part of the state. Cash receipts from livestock and livestock products exceed those from crops. Cotton is still one of the major crops, although its value is far below the peak reached in the early 20th century. Georgia is the leading state in peanut (groundnut) production and ranks high in tobacco. Corn (maize) and soybeans follow peanuts in value. Peaches have become especially identified with Georgia, and pecans and watermelons are grown nearly everywhere in the state.

Major crops

Transportation. Water transportation determined the location of Georgia's first cities. As early as 1790 William Longstreet of Augusta was experimenting with steam-powered craft on the Savannah River. By the late 1820s, river steamers were carrying large cargoes of cotton downstream from collecting warehouses at the fall line to Savannah and other export centres. The *Savannah,* equipped with auxiliary steam power, sailed from its namesake port to Liverpool in 1819 to become the first steamship to cross the Atlantic Ocean.

Navigation on 500 miles of inland waterways has been

revived, and a state port authority has created barge service at Augusta, Columbus, Bainbridge, Savannah, and Brunswick for the distribution of chemical, wood, and mineral products. Savannah is the leading port on the Southern Atlantic seaboard in terms of tonnage of cargo handled.

Milledgeville was briefly the centre of an emerging road system for the settled counties in eastern Georgia and for the old military and post road running through Indian territories to Alabama. Atlanta, originally called Terminus on the early railroad survey maps, had a near-optimum location for all but water transport, thus making it a hub of railroad transportation for the Southeast. With the advent of highways and then of air traffic, the city maintained its focal position. Three interstate highways intersect in downtown Atlanta.

More than a dozen cities in the state have commercial air service. Atlanta's Hartsfield International Airport is by some measures the world's busiest. A rapid transit system began operating in the Atlanta area in 1979.

Administration and social conditions. *Government.* Georgia adopted its 10th constitution in 1982, a document characterized by a reduction in the number of local amendments. The structure of state government tends to sever governmental from political processes, thereby limiting the appointive powers of the governor. The executive branch nonetheless exercises considerable control over state agencies by virtue of its major role in shaping the state's annual budget. The governor is elected to a four-year term, but is limited to serving two terms.

The Georgia General Assembly consists of the 56-member Senate and the 180-member House of Representatives and meets annually in 40-day sessions; districts in 1972 replaced counties as units of representation. Like all other areas of government in Georgia, the courts reflect a marked shift from an earlier day of highly politicized complexity to a more centralized and professionalized system. Probate courts, magistrate courts, and municipal courts function at the lowest level, with superior courts, state courts, and juvenile courts forming the next tier. The Court of Appeals and the Supreme Court form the capstone of the state judicial system.

County, regional, and municipal government

Georgia has many levels of local government, including 159 counties, about 533 municipalities, and more than 385 special districts. Counties often perform municipal-type services. Independently and through multicounty cooperative districts, they operate forestry units, airports, hospitals, and libraries. Most counties are governed by an elected Board of Commissioners.

Georgia politics has been in flux since the 1960s. Voting patterns have changed, and distinctions among local, state, and national politics have increased. Republican or third-party candidates have carried the state in presidential elections since 1964, with the exception of 1976 and 1980, when Jimmy Carter, a Democrat who had been governor of Georgia in 1971–75, won the state's presidential vote. Local politics and the "courthouse rings" (local elites) remain almost universally Democratic; and Democrats, while remaining shy of identification with the more liberal national party, continue to win the state house offices. Republicans, however, have won seats in Congress.

The political idiom has changed as well. Explicit racial demagoguery is no longer in evidence. Campaigns are long, and, with only occasional captive audiences, gubernatorial candidates are forced to stump and shake hands with the voters in a way that was once unnecessary.

Black precincts in Atlanta and Macon that were 80 to 90 percent Republican in 1956 are now Democratic by that margin or more. Whites and blacks in rural and less-affluent urban areas may vote for candidates on racial issues, whereas rural elites and affluent suburbanites tend to vote for either party depending on the nature of the contest and the commitments of the candidates.

The Civil Rights movement in Georgia has been characterized by legal action, many nonviolent and a few violent confrontations, selective buying campaigns, voter registration, and education. The movement has touched life in rural Georgia only a very little. In cities and in settings of high visibility—offices, airports, stores, restaurants, and schools—blacks occupy roles dramatically different from those of earlier decades.

Education. Public education in Georgia dates from the passage of a public school act in 1870. Since 1945 the ages for compulsory attendance have been from six to 15 years. Since 1964 state support of public schools has been for a period of nine months per year. The racial integration of public schools increased private-school enrollments dramatically. In 1985 the General Assembly passed the Quality Basic Education Act, which substantially revised the formula for allocating state funds to local school systems.

Major colleges and universities

Public institutions of higher learning, under a unified Board of Regents, are headed by the University of Georgia (chartered 1785; opened 1801) in Athens; the Medical College of Georgia (chartered 1828; became part of the university system 1950) in Augusta; and the Georgia Institute of Technology (1885) and Georgia State University (1913), both located in Atlanta. Other state colleges and junior colleges are spread across the state so that 95 percent of the population is within 35 miles of a college or university. The four undergraduate colleges and the graduate and professional schools of Atlanta University Center, all located on a single campus, are at the forefront of black higher education in the United States.

Health and welfare. Georgia has a modern mental health program. Regional hospitals for evaluation, emergency, and short-term treatment have been established throughout the state to serve communities within a 50-mile radius. In addition, there are approximately 60 community health care centres for outpatient treatment. A number of general hospitals have been built through federal programs.

Georgia has imaginative programs in family and children's services. There are state and regional youth-development centres for persons under 17. The state aids colleges in training welfare workers, whose activities are supplemented by a widespread volunteer program.

Cultural life. Atlanta is the cultural centre of the Southeast. Its memorial arts centre includes the High Museum of Art and a school of the visual arts, with performing facilities for its symphony orchestra and a professional resident theatre, both of which have premiered new works. Atlanta also has cooperative galleries run by painters and sculptors, and an active group of filmmakers.

Elsewhere in the state, there are regional ballet companies, and community theatres perform in more than 30 localities. In addition to instruction in theatre, dance, the visual arts, and music in many colleges, Georgia Institute of Technology has a school of architecture, and the University of Georgia has a school of environmental design. Dozens of public museums and college galleries exhibit art, and Atlanta University has a notable Afro-American collection. In 1988 Atlanta hosted the first National Black Arts Festival.

Georgia is rich in traditional arts and crafts, especially in the mountainous north. The craft of tufted fabrics was a major factor in attracting the carpet industry that developed around Dalton. Other handicraft workers find sales opportunities through country fairs in Hiawassee and nearby Gatlinburg, Tenn., and at many other art festivals across the state. A mountain arts cooperative has a store in Tallulah Falls, and craft shops are attached to several art galleries. A bulletin of the state's agriculture department gives free advertising for crafts, and a quarterly publication describes many of the old craft techniques.

Folk music

Traditional music is sung by folk groups on the sea islands and in the mountains. Country music conventions are held in northern Georgia—with some tension between purists and users of electronic equipment. In rural churches of the northwest, unaccompanied shape-note, or "fa-so-la," singing from the *Sacred Harp* songbook remains strong, and throughout the area many prayers and sermons are delivered in singsong.

Georgia has a wide range of outdoor recreational opportunities. The mountainous north is dominated by the Chattahoochee National Forest, which includes the Cohutta Wilderness Area. On the coast is Cumberland Island National Seashore, which comprises part of that large barrier island. Numerous other national wildlife areas

and refuges are found throughout the coastal zone. The unique character of Okefenokee Swamp is nurtured and preserved through the administration of the Okefenokee National Wildlife Refuge and Wilderness Area, as well as the Stephen C. Foster and Laura S. Walker State Park facilities found there. Georgia also maintains a system of state parks that offer a wide range of outdoor recreational experiences, from ocean surf bathing to mountain hiking and climbing.

Though generally conservative, the press in Georgia has supported the more liberal statewide candidates in recent decades. *The Atlanta Constitution* has been recognized as one of the nation's outstanding newspapers. The Henry W. Grady School of Journalism at the University of Georgia oversees competition for the George Foster Peabody Awards for Distinguished Achievement in Broadcasting.

HISTORY

Prehistoric period. The first inhabitants of what is now Georgia found their way into the area during the period from 10,000 to 8000 BC. A migratory hunting people equipped with finely worked flint projectile points, these so-called Paleo-Indians appear to have built small, seasonally occupied camps as they followed the movements of their large animal prey. Members of the culture that arose between 8000 and 1000 BC, known as Archaic, developed a more diversified food supply but continued the seasonal migration of their ancestors. Permanent or semipermanent village settlement in Georgia came with the emergence of the Woodland culture in the period 1000 BC to AD 900. Small, widely dispersed, permanently occupied villages were inhabited by the Woodland agriculturalists, who supplemented their harvests with a variety of wild foods. Georgia Woodland Indians left their most lasting mark in the form of large mounds built of thousands of basketfuls of clay and earth. Some mounds contained human burials and elaborately worked jewelry, pottery, and figurines. Other mounds did not contain burials but were built in the shape of animals. The best-known of these is the Rock Eagle in what is now Putnam county, a large complex of quartz rocks laid out in the shape of a bird.

The Mound Builders

Spanish exploration. By the time Hernando de Soto led the first European expedition into the area about 1540, the Mississippian culture, so called after the river valley in which it flourished, had established its influence across the Southeast, with the Creek and Cherokee groups predominating in what is now Georgia. De Soto found a population of master farmers whose large permanent villages were built around enormous earthen temple mounds resembling the flat-topped pyramids found in Central America. Their reliable and productive system of agriculture, based on corn, beans, squash, pumpkin, and tobacco, provided surpluses in most years. Directly or indirectly, the Spanish expedition was disastrous for the Indians. In addition to the hundreds they killed or enslaved, the explorers were ultimately responsible—through the diseases they introduced, such as measles, smallpox, and whooping cough—for the deaths of thousands and the final decline of Mississippian culture in Georgia.

In 1565 the Spanish, responding to a French attempt to settle on the southeastern coast, began their occupation of Florida. From the stronghold at St. Augustine, Spain began to exert an increasing influence on the Indians of Georgia. A line of Roman Catholic missions and associated military posts were established on the barrier islands along the Georgia coast. The lives and settlement patterns of the coastal Indians were profoundly changed as they were converted to Christianity and persuaded to adopt a sedentary life-style in compact villages. Known to the Spanish as Guale, the Georgia coastal zone remained under the mission-presidio system for a century. In the second half of the 17th century increasing pressures from the British in South Carolina eventually led to the withdrawal of the Spanish missions from Guale. As Spanish power waned and British power grew, the area of modern Georgia came to be known as the Debatable Land. The South Carolinians began to build an Indian trade monopoly in the area but were slow to attempt permanent settlement south of the Savannah River.

English settlement. A trust for establishing the colony of Georgia was granted a charter by George II (after whom the colony was named) in 1732, long after the large English migrations of the 17th century. The prime mover in obtaining the charter was the English soldier and philanthropist James Edward Oglethorpe, who sought to found a colony where the poor of England could get a new start. He and other trustees encouraged the settlers to produce wines, silks, and spices, and thus relieve England of a dependency on foreign sources. The colony also would serve as a bulwark against the Spanish and French to the south and west.

First English settlement

The first English settlement in Georgia was made at Savannah in 1733. Some colonists paid their way; the colony's trustees paid the expenses of others. Oglethorpe directed the affairs of the colony, primarily its military operations. Essential to the trustees' utopian plan was a tightly structured settlement system designed to create a population of yeoman farmers living in compact villages and towns and cultivating outlying garden and small farm tracts. Slavery was prohibited in order to avoid the growth of large plantations. Like most such schemes, the colony failed to live up to the trustees' vision. Their most notable success was the planning and construction of Savannah. Faced with unrest and emigration, the trustees surrendered all power in the colony to the British government in 1752, a year before their charter was to expire.

Revolution and growth. In a thrust of pre-Revolutionary inland migration, substantial settlement of Georgia began as a belt extending along the Savannah River that reached the lower Piedmont. Georgia's response to the revolutionary tensions was complex and resulted in veritable civil warfare between loyalists and patriots and a time of chaos for most Georgians. After the Revolution, settlement expanded rapidly, especially westward from Augusta into the future "cotton counties" of middle Georgia. Speculations in public lands acquired through the removal of the Creek and Cherokee Indians from the state paved the way for the development of a largely commercial agriculture, which soon became overwhelmingly dominated by cotton. The Indian removal provided a traumatic chapter in the state's history. The Creeks and Cherokee, two of the Five Civilized Tribes of the Southeast, were removed under force of arms by way of the Trail of Tears to federally owned lands in what is now Oklahoma.

The slavery controversy and the Civil War. By the mid-19th century, most white Georgians, like most Southerners, had come to view slavery as economically indispensable to their society. When the Civil War came, most white Southerners, slave owner or not, joined in the defense of their common heritage and culture. Georgia, with the greatest number of true plantations, had in many respects come to epitomize this culture.

The war involved Georgians at every level. Coastal attacks and sporadic raids into the state were a prelude to the attack on Atlanta in the late summer of 1864, when General William T. Sherman launched his March to the Sea. In mid-November Sherman initiated a plan to cut a 50-mile-wide swath across Georgia. Starting from Atlanta, the left wing moved along the route of the Georgia Railroad to Madison and Milledgeville, while the right wing went overland to the southeast, leaving a broad belt of almost total destruction.

Sherman's March to the Sea

The aftermath of the Civil War has been seen as a return to essentially frontier conditions in Georgia. Georgians no longer enjoyed mastery over their environment, and new modes of social and economic organization emerged in efforts to regain such mastery. Agriculture still appeared to hold Georgia's most promising future, but the relationship between land and labour was changed dramatically.

After some experimentation with various contractual arrangements for farm labour following emancipation, the system of sharecropping, or paying the owner for use of the land with some portion of the crop, became a generally accepted institution in Georgia and throughout the South. The system encouraged both the landowner and the sharecropper to strive for large harvests and thus often led to the land being "mined" of its fertility. Almost invariably, land and capital remained in white hands, while

labour remained largely, though not entirely, black. This entrenched pattern was not broken until the scourge of the boll weevil in the early 1920s ended the long reign of cotton.

The 20th century. The Great Depression began a decade early in Georgia. The collapse of the cotton industry led to a New Deal program of agricultural diversification encouraged by both state and federal agencies. Peaches and peanuts made important contributions to the Georgia farm economy, and tobacco, livestock, dairying, and poultry and egg production got under way on a commercial scale. As World War II began, thousands of Georgians became involved in the war effort, and federal funds were directed toward military training facilities and defense industries located in the state.

The postwar years saw economic prosperity increase rapidly in Georgia. After 1950 more Georgians were employed in manufacturing than in agriculture. At the same time, efforts to end racial segregation began. After some initial opposition, integration of the state's colleges and universities was undertaken in the 1960s, and integration of public elementary and secondary schools followed.

Impatient with the slow pace of change, many black Georgians turned to social activism, most notably Atlanta-born Martin Luther King, Jr., whose advocacy of nonviolent struggle for racial equality brought him the Nobel Peace Prize in 1964. Increasing numbers of black voters led to the election of blacks to state and local offices, including the election in 1973 of Maynard Jackson as the first black mayor of Atlanta.

Georgia's advances in the spheres of politics, civil rights, and economic growth continued. The state received national recognition with the 1976 election of President Jimmy Carter, the first Georgian ever elected to that office. (G. He./L.De V.)

Kentucky

The Commonwealth of Kentucky is a south central state of the United States. Long the home of various Indian tribes, Kentucky was settled by Daniel Boone and other frontiersmen in 1769. Its name probably derives from the Iroquois word for "prairie." By 1792, when it was admitted as the 15th state of the Union—the first west of the Appalachian Mountains—Kentucky had drawn nearly 75,000 settlers.

Location and size

Rivers define Kentucky's boundaries except on the south, where it shares a border with Tennessee along a nearly straight line of about 425 miles (685 kilometres), and on its mountainous southeastern border with Virginia. The Tug and Big Sandy rivers separate it from West Virginia on the east and northeast. From the point where the Big Sandy empties into the Ohio River, the crested northern boundary cuts an irregular line across the country, following the Ohio and meeting the states of Ohio, Indiana, and Illinois to the north. Where the Ohio flows into the Mississippi, the short western edge of the state is separated by the Mississippi from Missouri. These boundaries encompass the state's area of 40,410 square miles (104,660 square kilometres). The capital, Frankfort, lies between the two major cities, Louisville, which lies on the Ohio River, and Lexington.

Kentucky brings to mind images of coal mines, of the bourbon whiskey named for the county where it was developed and is still made, of white-suited colonels and their ladies sipping mint juleps on summertime verandas, of mountaineers and moonshiners, of horse breeding and the Kentucky Derby. Actually, Kentucky encompasses a curious mixture of poverty and wealth, ugliness and beauty, North and South. Several hundred lives have been lost in Kentucky's coal mines, and strip-mining has left countless hillsides to erode. Yet the seemingly endless landscape of white-railed horse pens and paddocks, characteristic of the rolling Bluegrass region around Lexington, symbolizes an unhurried and genteel way of life that looks more to Kentucky's ties with the pre-Civil War South than to its position in the industrial frenzy of the nation. By further contrast, northernmost Kentucky, with its predominantly German heritage and suburban pattern of development, belongs to metropolitan Cincinnati, Ohio. Kentucky has always existed in the middle: as a state looking back and ahead, as a crossroads for westward expansion, and as a split personality during the Civil War. It was the birthplace both of Abraham Lincoln, 16th president of the United States, and of Jefferson Davis, president of the Confederate States during that strife.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Parts of Kentucky lie within three major physiographic regions of the United States—Appalachian Highlands, the Interior Lowlands, and the Coastal Plain. Within the state, six smaller regions may be identified, based on the underlying rock structure: Mountain, Knobs, Bluegrass, Pennyrile (or Pennyroyal), Western Coalfield, and Purchase.

More than 10,000 square miles of the easternmost part of Kentucky lie in the Mountain region, a sloping plateau of the Cumberland and Pine mountain ranges. It is a scenic land of narrow valleys, steep pinnacles, and transverse ridges. The state reaches its highest point at Big Black Mountain, 4,145 feet (1,264 metres) in altitude. An area of deep gorges, natural rock arches, and small valley farms, eastern Kentucky is drained by three major rivers and their tributaries: the Big Sandy, Cumberland, and Kentucky rivers. Natural passages through these mazes of mountains are sometimes provided by winding gaps, such as historic Cumberland Gap, or water gaps, which include the picturesque Breaks of Sandy. The Cumberland River descends from the plateau in the 68-foot Cumberland Falls, renowned for its moonbow, the only known occurrence of this phenomenon in North America. The great eastern coalfields of Kentucky lie in the mountains and, though the region has been a major coal-producing area throughout the 20th century, there are billions of tons of coal still buried in the eastern hills.

Physical diversity

A long, narrow region shaped like an irregular horseshoe with both ends touching the Ohio River, the Knobs embraces the Bluegrass country on its inner side, the Mountain area on the east, and the Pennyrile on the west. Its landscape is one of cone-shaped or rounded hills and ancient escarpments. The weathered shale soil is not rich and is easily eroded, making it better adapted to forest growth than to cultivation. Canebrakes grew along some of the lower ground before European settlement and attracted large herds of buffalo and deer. A major portion of the Daniel Boone National Forest lies in the eastern Knobs.

There is a folk saying that when east Kentuckians die they want to go to Lexington, the capital of the Bluegrass. The Bluegrass lies at Kentucky's geographic and legendary heart. Its 8,000 square miles are encircled by the Knobs and the Ohio River. The region was named for the long-stemmed grass that flourishes there. The underlying limestones are rich in phosphates and have created pasturage for some of the world's most famous horse farms.

The 7,800-square-mile Pennyrile adjoins every other region except the Bluegrass. On the east it joins the mountains; to the north its irregular boundaries are the Knobs, the Ohio River, and the Western Coalfield; in the west it joins the Purchase; and on the south it is bounded by Tennessee. Its name derives from the local pronunciation of pennyroyal, a plant of the mint family that is abundant in the area. The Pennyrile encompasses wooded rocky hillsides, small stock farms, cliffs, and an area once known as the Barrens—a condition caused by the Indians' continuous burning off of forest cover to make grasslands for buffalo. Most notably, it is a region of caves. Abundant waters, both surface and underground, and the limestones deposited during the Early Carboniferous epoch of geologic history (more than 300,000,000 years ago), have combined to create the area known as the Land of Ten Thousand Sinks and such famous subterranean passages as Mammoth Cave. The vast underworld cavern includes three rivers and three lakes, and it covers more than 300 miles on five distinct levels. Its temperature remains constant at 54° F (12° C) throughout the year. Many other caves underlie the Pennyrile. Along its western edge, Kentucky Lake and Lake Barkley, impoundments of the Tennessee and Cumberland rivers, respectively, have iso-

Cave region

lated a wooded peninsula known as the Land Between the Lakes, which is managed as an educational and recreational park by the Tennessee Valley Authority (TVA).

Surrounded by the Pennyrile and the Ohio River and crossed by the Green River, the Western Coalfield's 4,680 square miles comprise less than half the area of the eastern coal beds and only a little more than half of that of the Bluegrass. The region has a number of coal deposits throughout its extent, however, and it is fertile on some of its rolling uplands and its bottomlands. Hence it is both a mining and farming area.

The Purchase, also called Jackson Purchase, encompasses only 2,569 square miles in the southwestern corner of the state. It is bounded on the west by the Mississippi River, on the north by the Ohio River, and on the east by the impounded Tennessee River. The region's southern border is the "sunken" westernmost section of the long boundary with Tennessee. Geologically, the Purchase is the northernmost extent of the Gulf Coastal Plain. Its name refers to its purchase in 1818 by virtue of a treaty with the Chickasaw, of which Andrew Jackson, later the seventh president of the United States, was one of the signers. The Purchase is the lowest topographic area of Kentucky, but it is not uniformly flat. Wide floodplains are broken by low hills that may have been sandbars in ancient oceans. Bluffs, swamps, and lagoons form part of the terrain, and soft rocks of the region erode rapidly, altering the landscape. The area is one of the most fertile sections of Kentucky and is widely known both for crops and for its fine stands of poplar, hickory, and oak.

Reelfoot Lake (18,000 acres), on the Kentucky–Tennessee border, was formed by a series of earthquakes that began in December 1811 and lasted until March 1912. They were the most powerful tremors in U.S. history, estimated to have been between 8.4 and 8.8 on the Richter scale. There were some 1,800 shocks and aftershocks, the strongest of which were felt as far away as Washington, D.C., and New York City. It was reported that the shocks were so strong that the Mississippi River flowed backward for a few hours, as the land beneath buckled and surged upward; huge islands disappeared, and the town of New Madrid, Mo., dropped 12 feet and was buried underwater. As the river overflowed, Reelfoot, a lake 14 miles long, was formed. The shocks were caused by an ancient 40-mile-wide rift in the Earth that runs for roughly 200 miles along the Mississippi River valley from Memphis into Missouri and Illinois, bordering on Kentucky much of the distance. The rift is known as the New Madrid Fault, and scientists have predicted that additional quakes in the area are a certainty.

Rivers of the state

Drainage. Apart from the Mississippi, the Ohio, and the Big Sandy, which form parts of Kentucky's boundaries, seven main rivers flow through the state: the Licking, Kentucky, Salt, Green, Tradewater, Cumberland, and Tennessee. The rivers, which drain into the Ohio, have been dammed to form large reservoirs; Kentucky Lake and Lake Barkley are connected by a canal. The Green River, the longest stream that lies entirely within the state, rises in the eastern Pennyrile and flows some 370 miles before emptying into the Ohio.

Soils. The landscape of Kentucky is as diverse as the life it supports, extending from the wrinkled outcroppings of early Paleozoic rocks to the 2,300-mile shoreline of Kentucky Lake. Rich alluvial deposits lie along the rivers, while the rest of the state's soil derives from the long and gradual breakdown and decay of underlying rock.

Climate. Kentucky enjoys a temperate climate, plentiful rainfall, and distinctive soils, which combine to create variety in vegetation, animal life, and landscape. The state's mean annual temperature is between 55° and 60° F (13° and 16° C). The growing season lasts from 176 to 197 days a year. Mean annual rainfall for the entire state is about 45 inches (1,140 millimetres), evenly distributed throughout the year. The greatest differences occur between the southern areas, which average as much as 48 inches annually, and the northeast, which may receive only 40 inches. Prevailing winds are from the south and southwest, although winter's chill frequently arrives on north and northwest winds.

Plant and animal life. Kentucky was part of the hardwood forest region covering the nation from the Allegheny Mountains to the western prairies. Three-fourths of the state was once covered with stands of yellow poplar, oak, chestnut, sycamore, and walnut. By the close of the 19th century, however, all but a fraction of these virgin forests had been felled. Trees, shrubs, and plants of many kinds still flourish in all parts of the state, ranging from the native hardwoods and pines on the eastern slopes to the picturesque bald cypresses in the western river marshes, and the maples, beeches, and magnolia found throughout the state. Rhododendron, laurel, dogwood, redbud, and trillium are prominent among the dozens of flowering plants that can be found in the Kentucky mountains.

Birds and mammals of Kentucky include those native to the Deep South as well as those of southern Canada. Of the numerous hoofed animals that once roamed Kentucky—bison, elk, moose, and deer—only deer remain. Wolves and panthers have likewise disappeared. Among the many small animals found in the state are rabbits, squirrels, foxes, raccoons, opossums, woodchucks, and—in the numerous caves—bats and rodents. The northwestern corner of Kentucky, where the Green River flows into the Ohio, lies along one of the world's great migratory bird routes. More than 200 species of birds frequent this area, while close to 300 species have been found in the state as a whole. The marshes of the southwestern Kentucky–Tennessee border provide breeding places for such waterfowl as the American egret, great blue heron, and double-crested cormorant. A few wild turkeys remain as a reminder of pioneer days. The swift mountain streams, wide rivers, and man-made lakes of Kentucky provide habitats for more than 100 species of fish. The muskellunge, the largest member of the pike family and commonly considered a Great Lakes fish, is found in the Barren and Green rivers.

Birdlife

Settlement patterns. From the beginning, Kentucky has been a strongly rural state of small towns and crossroads. Only Louisville and Lexington have large populations. Part of the urban population lives in small cities such as Lexington, Covington, Owensboro, Paducah, and Frankfort. Following World War II many of the younger people left rural counties for cities both within and outside the state, creating severe economic, educational, and cultural problems.

The pace of living for most Kentuckians remains more leisurely than in many other areas of the country. The popular image of the Bluegrass Kentuckian includes leisurely colonels and fast horses, cold mint juleps and a hot game-meat stew called burgoo, while the mountaineer Kentuckian might be idealized as a dulcimer-strumming weaver of old English rhymes and homespun wisdom. Each is rooted, to some degree, in fact, and, though Kentuckians are losing many of their distinctive traits, part of the colour and flavour of individualism remains.

The people. Early settlers of Kentucky, who were predominantly English and Scotch-Irish, came from North Carolina, Virginia, Maryland, and Pennsylvania. The migrations of Daniel Boone reflected those of many of his fellow countrymen. He moved from Pennsylvania, where he was born, down the Great Valley of Virginia into North Carolina, where he lived until he led new settlers through Cumberland Gap to Kentucky. The best agricultural land was in the Bluegrass region, and this was the first area to be settled. The eastern mountains were settled last. Despite the horrors of backwoods warfare during and following the Revolution, the migration into the Bluegrass country continued. In addition to the Cumberland Gap route, the Mississippi brought early French émigrés from New Orleans, particularly to the Louisville area, while during the mid-19th century the Ohio River carried many German settlers and other migrants, via Pittsburgh, from New England and the Middle Atlantic states. There was also a large black population in Kentucky, though the proportion decreased after 1833. Just prior to the Civil War, the Underground Railroad flourished in Kentucky to help transport escaped slaves to free soil, and there was considerable black emigration during and after the war. The state's black population, constituting some 7 percent

of the total, is concentrated in the larger urban areas and in the southwestern part of the Pennyrile.

Diversified economy

The economy. Kentucky's economy comprises a balance among manufacturing, agriculture, mining, tourism, services, and trade. All regions of the state do not share equally, however. The Bluegrass is an affluent region. The Pennyrile is likewise diversified and prosperous, but economic conditions in the two coal-producing regions fluctuate with the demand for coal. The Purchase relies extensively on agriculture, and periods of drought or depressed crop prices cause the region to suffer. Manufacturing is the greatest income producer for the state.

Resources. Vast reserves of bituminous coal have made Kentucky the nation's number-one coal producer. Reserves are expected to last more than 200 years at present production rates. Kentucky's numerous streams and reservoirs provide more than 1,300 miles of navigable waterways and ample fresh water for manufacturing and recreation. Soils of the Bluegrass, Pennyrile, Western Coalfield, and the Purchase are excellent for agriculture. Only in eastern Kentucky is there a lack of land suitable for farming. Eastern Kentucky and the eastern Pennyrile have large timber reserves, primarily hardwoods. The two coalfields and the Pennyrile have oil and natural gas deposits, although not in large quantities. Several small deposits of vein minerals are found along with a variety of clays and an abundance of limestone.

Agriculture, forestry, and fishing. Prior to 1950 Kentucky was considered an agricultural state. Since that time the number of farms and the acreage devoted to agriculture have declined, although average farm size has increased. Principal crops are tobacco, corn (maize), soybeans, and hay. The Bluegrass region, with the richest soil, is highly specialized in horses, cattle, and tobacco. The Pennyrile produces a variety of crops and livestock. The Western Coalfield and the Purchase specialize in corn, soybeans, and tobacco, although some livestock and smaller acreages of other crops are found. Forestry is important in eastern Kentucky and in the eastern part of the Pennyrile; the trees cut are mostly hardwoods. Kentucky has little commercial fishing, but its streams and reservoirs provide excellent opportunities for sport fishing.

Mining and industry. Coal is by far Kentucky's most important mineral. It is found throughout the two coalfield regions. Eastern Kentucky coal is of coking quality. That of the Western Coalfield is higher in sulfur content and is used primarily for steam generation of electricity and domestic use. Modern mining methods produce vast quantities of coal with few workers, and unemployment rates in the coalfields are high. Several large steam-generating plants in the Western Coalfield—the largest at Paradise—consume thousands of tons of coal daily.

Coal mining

Manufacturing, although widely dispersed, is concentrated in the urban areas. The Louisville area accounts for 25 percent of the state's plants and nearly one-third of all manufacturing employment. The metals-related industries have dominated the state's industrial growth in recent years. Textile plants, often located in smaller communities, account for approximately 10 percent of manufacturing employment. Calvert City, near the mouth of the Tennessee River, has a large concentration of chemical industries.

For many county seats in Kentucky, the trade and service industries are of primary importance, although nearly all have manufacturing plants. Kentucky has excellent tourist facilities, especially near the lakes and in the larger cities.

A strong labour union tradition exists in the Ohio River towns, and the United Mine Workers of America (UMWA) is influential in the coal region. Early struggles between the UMWA and coal operators in eastern Kentucky gave rise to tragic violence. The name of Bloody Harlan commemorates that county's labour wars during the 1920s and '30s, highlighting working and living conditions that became popularly identified with those of the state as a whole. Numerous ballads recount the history of conflict and death surrounding work in the coal mines.

Transportation. Interstate highways cross Kentucky from north to south and east to west. They are supplemented by a system of parkways, U.S. highways, and state highways that make travel by automobile or truck relatively easy almost everywhere in the state. Rail lines connect all major cities for movement of freight. Bulky freight is often shipped by river barge over Kentucky's many miles of navigable waterways. Three major airports—in Louisville, Lexington, and northern Kentucky—serve the central Kentucky area. The Greater Cincinnati airport in northern Kentucky provides international service. Several medium-sized cities have connector lines.

Administration and social conditions. *Government.* Under the constitution adopted in 1891, the state government comprises the executive, legislative, and judicial branches. The governor is elected for a four-year term and may not succeed himself. The General Assembly, which meets in even-numbered years, is bicameral, with a Senate and House of Representatives. The Senate has 38 members who serve for four years, and the House has 100 members who serve for two years. Tax bills must originate in the House. There are several levels in the state court system, ranging from local police courts to the seven-member Supreme Court. The judges of the Supreme Court, the Court of Appeals, and the circuit courts are elected for eight-year terms. District court judges are elected for four years.

Kentucky has 120 counties, each headed by a county judge who has substantial appointive powers and is responsible for preparing the budget and estimating receipts. The fiscal court serves as the administrative and policy-making body of each county. County officials are elected for four-year terms. Kentucky has no townships but has a system of magisterial districts. Municipalities are divided into six classes according to population. There are three forms of city government: the mayor–council plan, the commission plan, and the city-manager plan. The mayor–council plan, which provides for separation of executive and legislative powers, is most favoured.

Education. Kentucky's first school was founded at Fort Harrod in 1775. Education is free and compulsory between the ages of seven and 16. The average educational level is below the national average but has risen in recent years. State taxation for the support of education was first levied in 1904. Elementary, middle, and high schools have in most cases been consolidated at the county level and occupy modern facilities. Most of Kentucky's private schools are church-supported.

Kentucky has a number of state-supported universities and private two- and four-year colleges, as well as vocational schools and state-supported community colleges. Transylvania University in Lexington, chartered in 1780, is the oldest institution of higher learning west of the Allegheny Mountains. The University of Kentucky, also in Lexington, is the state's largest university and has responsibility for the community colleges. The University of Louisville, founded by the city council in 1798, is the oldest public university in the state. It became part of the state university system in 1970. Berea College, founded in 1855 to serve needy students from Appalachia, has become a regional centre for traditional arts and crafts.

Higher education

Health and welfare. The Department for Human Resources is the largest department in state government. It has numerous divisions, including preventive medicine, medical inspection and licensing, medical care for the needy, chronic disease control, sanitation in water supply and sewage, public assistance, and child-welfare programs. Lexington and Louisville each have general and specialized hospitals and a university with medical and dental schools. Most county seats have hospitals, but there is a shortage of medical and dental personnel in the more remote areas. There is a need for improvement in health care facilities in smaller towns and rural areas where doctors and dentists must usually practice corrective rather than preventive health measures. A unique feature of health care in Kentucky is the Frontier Nursing Service, founded in 1925, which provides general nursing and obstetric service in the isolated mountain area of eastern Kentucky. A variety of programs throughout the state provide care for the elderly and the handicapped.

Cultural life. Kentucky life-styles are a little more Southern than those of states north of the Ohio River, but

differences are minor and in most cases are the result of smaller populations and a more rural outlook rather than ethnic differences. The larger cities of the Bluegrass are centres for the arts and have a variety of museums, theatres, galleries, and musical groups. Several of Kentucky's universities and colleges are in the Bluegrass. Their orchestras, theatre groups, concerts, and lecture series add to the cultural opportunities. Lexington is the centre of the world of horse breeding, and horse shows and horse racing are well-known Bluegrass traditions. Northern Kentucky, although part of the Bluegrass, reflects the German heritage of metropolitan Cincinnati in its churches, restaurants, family names, and an annual Oktoberfest.

Folk arts

Kentucky continues to make a special contribution to the national culture with its folk arts, especially in the rural areas. Haunting ballads from Elizabethan days and mournful songs relating recent tragedies or desertions combine to create a distinctive musical life among mountain people. Crafts handed down through generations still produce handsome homespun cloth, hand-carved furniture, patchwork quilts, and sturdy pottery. Surrounded by a mechanized, standardized world, Kentucky folk songs and handicrafts preserve a link with earlier days. The annual Big Singing, held at Benton in western Kentucky each May for more than 100 years, celebrates the heritage of shape-note, or "fa-so-la," singing. Among the nationally recognized writers identified with Kentucky are Robert Penn Warren, Wendell Berry, and Bobbie Ann Mason.

Kentucky has one of the finest state park systems in the United States. Several of the parks in the system are resort parks with lodges, cottages, campgrounds, and a variety of recreational facilities. The state also has three national parks: Mammoth Cave, Cumberland Gap, and the Abraham Lincoln Birthplace National Historic Site. A state fair is held in Louisville in August of each year, and Kentucky's counties have annual fairs, festivals, and horse shows. Thoroughbred and harness racing is found at several locations, and the horse farms in the Bluegrass attract many visitors. The Kentucky Derby, held at Louisville's Churchill Downs on the first Saturday of each May since 1875, is the first leg of the Triple Crown of Thoroughbred racing. The Daniel Boone National Forest in eastern Kentucky and the Land Between the Lakes national recreation area in western Kentucky are popular attractions. Kentucky's climate is favourable for outdoor recreation during most of the year.

SuperStock

Grazing paddock on a horse farm in the Bluegrass region, near Lexington, Ky.

HISTORY

Exploration and settlement. Before the arrival of Europeans, the Kentucky region was a hunting ground and battlefield for such Indian tribes as the Shawnee from the north and the Cherokee from the south. Even earlier agricultural and hunting peoples left burial mounds and other traces. French and Spanish explorers must have seen Kentucky from the rivers of the Mississippi basin, and traders entered the region from the eastern colonies during the early 18th century. During the 1750s and '60s Indian resistance and rough terrain hindered successful exploration of the region. In 1769, however, Daniel Boone penetrated to the central plateau region, or Bluegrass country.

Indian raids and immigration

Settlement was rapid during the 1770s, though the prophecies of an angry Cherokee chieftain, Dragging-Canoe—that Boone and other whites would find Kentucky "a dark and bloody land"—were in large part fulfilled. British officers spurred the Indians during the Revolution, notably in raids on Boonesboro in 1777 and 1778 and at a bloody ambush at Blue Licks in 1782, and settlers encountered numerous other sieges, scalpings, and skirmishes. Following the war immigrants poured down the rivers and traveled the Wilderness Road from Cumberland Gap. The settlers founded towns and before long began to call for separation of the judicial district of Kentucky from Virginia. Statehood conventions at Danville in the 1780s were somewhat ruffled by the "Spanish Conspiracy" of James Wilkinson and others to ally the region with Spain, but they led ultimately to admission into the Union on June 1, 1792, and to the organization of state government, which took place in a Lexington tavern.

Statehood and crises. Events leading to a second state constitution in 1800 revealed an internal division that has continued to characterize Kentucky. Farmers, who floated their grain, hides, and other products on flatboats down the Mississippi to Spanish-held New Orleans, allied themselves with other antislavery forces to oppose slaveholders and businessmen. The federal Alien and Sedition Acts of 1798, passed in an attempt to control criticism of the government, were vigorously opposed. One of the leading spokesmen for the opposition was the young politician Henry Clay, who was to stamp his personality on the state and national scenes as the "great compromiser."

Kentucky took a lead in the War of 1812, much of which was fought in the adjacent Northwest Territory against combined British and Indian forces. Following the war a land boom, with attendant speculation and inflation, and the chartering of 40 independent banks that flooded the state with paper money led to financial disaster during the national economic panic of 1819. Fierce controversy over relief to debtors split Clay's Whigs and Andrew Jackson's Democrats. Signs of progress from 1820 to 1850, however, included the building of a canal at Louisville, the chartering of railroads, and increased manufacturing. The slavery question was uppermost, however, until the Civil War. The few large slaveholders were located mainly in the plantation agriculture of the Bluegrass and Pennyrile sections, but by 1833, when the legislature forbade importation of slaves for resale, the state was already one-quarter black. Until the Civil War, proslavery forces maintained an iron control of government and prevented any constitutional change that endangered their property.

Civil War and its aftermath. During the war Kentucky was a state divided. Officially, it had sought to avoid war by continuing Clay's tradition of compromise. Some 90,000 soldiers fought for the Union armies and about 40,000 for the Confederacy, though after the war popular sentiment became strongly pro-South. Following the defeat of the Confederate general Braxton Bragg at Perryville, in 1862, the only action in the state consisted of widespread guerrilla warfare.

The opening of new rail lines into the eastern coal country and the introduction of a tobacco economy stirred the state in the last decades of the 19th century. The Ku Klux Klan evoked fears and hatred, but the freed blacks were given the vote and settled as tenant farmers or urban workers. Blacks, however, were not to become first-class citizens. Lexington and the Ohio River cities—Louisville, Owensboro, Paducah, and Covington—grew rapidly, and

the state was intensely involved in the populist agrarian politics of the period. Warfare between tobacco growers and tobacco trusts brought on an era of barn burning and similar attempts to keep tobacco prices up.

Labour strife

Continued diversification marked the 20th century, though the Great Depression and strikes by the UMWA brought serious problems and open strife in many sectors. The state followed the national trends toward the loss of rural population to industrial centres, both inside and outside the state. In 1966 the devastations of strip-mining were partially ameliorated by a law requiring restoration and reforestation of the landscape.

In the early 1970s a nationwide energy shortage created a demand for more coal, and Kentucky's coalfields prospered for nearly a decade. As oil prices stabilized, however, the demand for coal diminished. Layoffs in the automotive industry reduced the demand for steel, which in turn reduced the demand for coking-quality coal. Environmental concerns added to costs of coal production and use. Coal operators, in attempts to reduce production costs, introduced modern machinery that reduced the need for manpower. Unemployment in the coalfields became a major concern. In the interior of eastern Kentucky, where there is little agriculture and manufacturing, the incomes of many families dropped below the poverty level.

Kentucky's farms, which had numbered 279,000 in 1935, numbered less than 100,000 by the late 1980s as a result of falling prices for agricultural products, increased mechanization, and periods of drought. Tobacco, Kentucky's greatest income-producing crop, was declared to be a health hazard, making the future for this crop uncertain. Expansion of industry and educational reform were priorities of Kentucky administrations in the 1980s, including that of the state's first woman governor, Martha Layne Collins, elected in 1984. (W.D./W.A.Bl.)

Louisiana

With parts of its land lying farther south than any portion of the continental United States except for southern Texas and the Florida peninsula, and with New Orleans, its largest city, lying on roughly the same parallel as Cairo, New Delhi, and Shanghai, Louisiana owes much of its complex personality to its geographic position.

Admitted to the Union in 1812 as the 18th state, Louisiana commands a once strategically vital region where the waters of the great Mississippi–Missouri river system, draining the continental interior of North America, flow out into the warm, northward-curving crescent of the Gulf of Mexico. It is not surprising that seven flags have flown over its territories since 1682, when the explorer Robert Cavelier, Lord de La Salle, placed a wooden cross in the ground and claimed the territory in the name of France's Louis XIV. The consequent varieties of cultural heritage run like bright threads through many of the aspects—social, political, and artistic—of life in the state.

The subtropical climate of the state has provided the magnificent, brooding scenery of the coastal bayous, and the lush, dank vegetation of its shores conceals a wealth of oil. The fertile soil covering much of the terrain made Louisiana a rich agricultural area by 1860, with sugarcane and cotton plantations flourishing. A lumbering boom occurred at the turn of the 20th century, and Louisiana underwent rapid industrialization after World War II. Mineral output is great, and the state ranks among the nation's leaders in petroleum production.

Turbulent heritage

But progress has not been without its tragic and turbulent aspects: bitter territorial disputes and violent internal struggles for political power impeded the social and economic development of the state and crippled many of its political institutions. The wealth of the plantations was accumulated through the extensive use of slaves, whose descendants comprise almost one-third of Louisiana's population and whose culture has contributed much to the social fabric of the state. Racism and racial conflict have marred the development of the state from the Civil War period, through Reconstruction and the ensuing reaction, marked by the activities of the Ku Klux Klan, down to the civil rights conflicts of the 1960s and beyond.

The state is delineated from its neighbours, Arkansas, Mississippi, and Texas, by four natural and three man-made boundaries. The 47,752 square miles (123,678 square kilometres) of Louisiana include more than 3,000 square miles of inland waters. The capital is Baton Rouge.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Louisiana shares the general physiographic characteristics common to the Gulf Coast states of the southern United States, with the vital exception of the Mississippi River, which flows through the state and extends its delta far into the Gulf of Mexico. The changing course of this great North American river has created the huge Atchafalaya basin and has dumped tons of sediment along the coast. Despite this, it has been estimated that the beachless coastline of Louisiana is eroding at a rate of about 16 square miles per year because the system of levees, or embankments, constructed by the federal government to keep the Mississippi in a central channel, has left side channels open to erosion.

Three types of regions are found in Louisiana: lowlands, terraces, and hills. The lowlands consist of the coastal marshes and the Mississippi floodplain with its natural levees and moderate relief. The Red River valley has a low elevation relief but with many raft lakes, built up by impounding water from a number of log jams, and red soils in association with its alluvial plain. The terraces include much of the so-called Florida Parishes above and to the northeast of the Mississippi and the prairies of southwestern Louisiana. Upland hills are on either side of the Red River valley and in the northern portion of the Florida Parishes; the state's highest elevation, in northwestern Louisiana, is 535 feet (163 metres) above sea level.

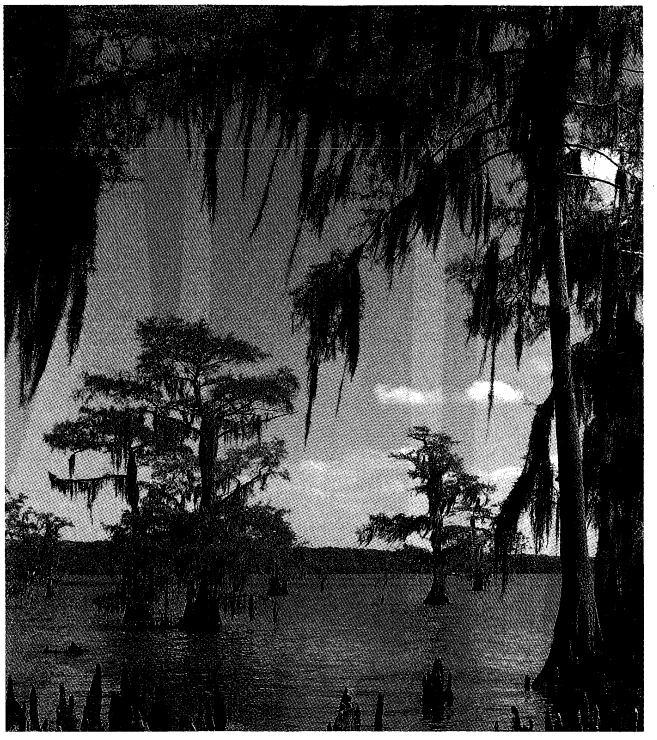
Soils. The soils of Louisiana have been one of the state's priceless resources; nearly one-third of the total land area is covered by the rich alluvium deposited by the overflowing of its rivers and bayous. Muck and peat soils are found within the coastal marshes, while the bottoms hold rich alluvial soils: the lighter and coarser bottom soils of the Mississippi and Red river valleys and older alluvium and loessial, or windblown, soils. Within the uplands, or hills, there are more mature soils that are less fertile.

Climate. Louisiana's climate is subtropical, a natural result of its location on the Gulf of Mexico. As it also lies at the mouth of the vast Mississippi–Missouri river valley, halfway between the Atlantic and the Pacific oceans, the state is also affected by continental weather patterns. Hot, humid summers, tempered by frequent afternoon thunder showers, alternate with mild winters. Louisiana is subject to tropical storms, and the hurricane season extends for six months, from June through November. Average annual temperatures range from 64° F (18° C) in the extreme north of the state to 71° F (21° C) at the mouth of the Mississippi River. The highest monthly average is 82° F (28° C) in July, and the lowest is 50° F (10° C) in January. Summer averages do not extend above the low 80s F, and it is the humidity, rather than the heat, that is one of the more marked characteristics of the state's subtropical climate. The frost season begins between November 1 in northern Louisiana and December 14 in the extreme southeast. The average growing season ranges from 220 to 320 days and the average precipitation from almost 45 inches (1,143 millimetres) at Shreveport to more than 56 inches at New Orleans.

Subtropical characteristics

Plant and animal life. Natural vegetation in Louisiana is found in three major divisions: the first consists of forest, upland pines and hardwoods, bottomland hardwoods, and bald cypress; the second of prairie, or dry grassland; and the third of marshland, or wet grassland. In the southern half of the state, along a zone running westward from Baton Rouge, the live oak with its characteristic drapings of Spanish moss predominates. The magnolia, whose blossom is the state flower, grows throughout the state.

Muskrats and other fur-bearing rodents, together with alligators, have been trapped in the marshes of southern Louisiana. There is a great variety of birds, native and migrant, but the once-frequent brown pelican (the state bird), the ivory-billed woodpecker, and the wild turkey are nearly extinct. The gray squirrel, deer, and dove are plen-



Spanish moss hanging from bald cypress in Lake Palourde, in the Cajun region of southern Louisiana.
© Raymond G. Barnes from TSW—CLICK/Chicago

tiful. Fish, shrimp, crayfish, crab, and oyster are a source of food and of income in the coastal and swamp areas.

Settlement patterns. Northern Louisiana forms a natural region including the northeastern Louisiana Delta, the Red River valley, and the northern Louisiana hills. Southern Louisiana, composed of the parish of Avoyelles and all the parishes that lie beneath the 31st parallel, has three major subregions: (1) the Florida Parishes, (2) southwestern Louisiana, which contains many Anglo-Saxon Protestants but also has an important French minority, and (3) in between, a region variously known as the Cajun, or the river and bayou country, or the sugar bowl.

The earliest settlements in the river and bayou parishes were "line" villages, where farmsteads were built at the riverfront of a long and narrow lot, with the stream serving as a highway. The line village pattern contrasted with the irregular pattern stemming from the ancient land division system of metes and bounds used by the Anglo-Saxons of the Florida Parishes. Where the natural levee was wide enough, plantations were established. Before the Civil War, people came to the uplands of northern Louisiana from the Eastern states and settled in isolated farmsteads among the pine woods. Southwestern Louisiana was developed after 1880, its prairies converted into rice fields. Settlement there resembled a grid system of land division found throughout the interior of the United States.

Variety of land divisions

The people. If a diversity of landscapes and forms of settlement characterizes the state, its peoples and its cultures also represent many Louisianas. The earliest European settlers were French or Spanish, and only later were the Florida Parishes and the northern part of the state settled by *les Américains.* Each area of settlement preserved a cultural heritage strongly marked by adherence to either the Roman Catholic or Protestant faith. The Louisiana French, particularly the descendants of the Acadians, came to dominate much of southern Louisiana; many of those who have arrived to live among them have been assimilated to the Cajun way of life. French is spoken in many parishes, and throughout southern Louisiana one may hear English spoken with a French accent. In addition, there are several cultural islands in both regions of the state. These are made up of Italian, Spanish (Isleños), Hungarian, German, and Dalmatian-Slavonian communities. There are also several racially mixed settlements.

The peoples of Louisiana exhibit a greater variety than those in other Deep South states not only because of the patterns of historical settlement but also because of the migration to and through New Orleans from Europe, Latin America, and Cuba. The number of foreign-born residents of Louisiana, however, has declined; now four-fifths of the population is composed of native-born Louisianians. The vast majority of foreign-born residents are found within the urbanized parishes of the state, especially New Orleans.

From the earliest days of the state, blacks have played an important role; and in the late 20th century they still constituted almost a third of the population. Historically, the black population was concentrated in areas containing the plantations sustained by their labour. During the early 20th century a large out-migration occurred, supplemented after World War II by a black migration to the state's urban areas. Although Louisiana's black population has been denied many of the traditional avenues leading to social and economic power, their culture has nevertheless contributed to the life and character of the state and of New Orleans, Louisiana's major city.

The black contribution

Southern Louisiana contains about three-fourths of the state's population, and almost one-third of the people in Louisiana live in the three most populated parishes. A predominantly urban population was achieved for the first time in 1950.

The economy. Louisiana has shared the economic underdevelopment afflicting Southern states. Prior to 1941 nearly one-third of the labour force was employed in the primary industries, based on raw materials, while only one-fifth worked in the secondary sectors of mining, building, and manufacturing. Such an imbalance depressed personal income per capita, which, until 1930, was but slightly more than half of the national figure.

World War II hastened the industrial growth of Louisiana to the extent that the numbers of the labour force engaged in manufacturing increased considerably. The most important development has been the establishment of a chemical industry, based on the oil, sulfur, salt, and water that are found in the area. An investment boom occurred from 1947 to 1957 when the first big move to offshore petroleum production was made.

Resources. Petroleum resources are found in all areas of the state, but the main oil fields have been developed between Shreveport and Monroe and throughout most of southern Louisiana. Drilling has been moved out into the Gulf. Natural gas resources have also been utilized. Including offshore drilling in federal waters, Louisiana ranks third nationally in the production of crude petroleum and second in the production of natural gas. Oil in Louisiana is often found in association with the more than 100 known salt domes, blisterlike intrusions in the bedrock, and sulfur lies in the caprock overlying the salt. Louisiana leads the nation in the production of salt and is one of the major sulfur-producing states. The natural gas resources of the state have been an important source of industrial power, and the abundant mineral fuels that are available have been used to develop electrical power. Conservation efforts have capitalized on the resources of soil and climate, and there is, as a result, extensive "tree farming." Louisiana also has a prime asset in its water resources.

Pipelines are used to carry crude oil to refineries or natural gas to provide energy for homes and industries in state and distant markets. Many miles of electric power lines crisscross the state.

Industry and agriculture. Chemical products, petroleum, and transportation equipment are the leading manufactured items. Cotton is no longer king in the agricultural domain: it was first in cash farm receipts in 1960, but within 20 years soybeans had become the leading agricultural product, with beef cattle ranked second. Other farm products include rice, sugarcane, dairy products, and poultry and eggs. The state has become much less dependent upon farming.

The decline of cotton

Transportation. Louisiana's waterways have always been an important means of transportation. The state's 4,800–7,500 miles (7,680–12,000 kilometres) of navigable waterways include the Intracoastal Waterway. It is Louisiana's only east–west waterway and canal system

and runs some 310 miles from Mississippi Sound to the Sabine River. It is part of a larger waterway extending from the Caloosahatchee River in Florida to Brownsville, Texas. The port of New Orleans is ranked second in the nation in volume of seaborne freight, while Baton Rouge, farther up the Mississippi River at the head of deep-channel navigation, is important for shipping of petroleum and chemical products, including aluminum, and grain.

Railroads became common after the 1830s, initially as feeders to the steamboat traffic, with the Clinton and Port Hudson line being the first railroad in the state. Railroading reached its peak in the early 20th century in connection with a feverish lumbering boom, and there are nearly 3,000 miles of track still in use in Louisiana. The state also has several thousand miles of highway. Louisiana has some 180 airports, and New Orleans International Airport, a leading continental link, is a major point of connection with Latin America.

Administration and social conditions. *Government.* A state constitution that was adopted in 1921 was Louisiana's 10th in 108 years. It remained in force for more than 50 years, a period during which the state underwent more fundamental change than had occurred in all the preceding 109 years of statehood.

The earliest document (1812) secured the political power of the planters and business classes and gave great appointive powers to the governor; the antebellum documents of 1845 and 1852 extended the suffrage and made every government office elective. Representation, based on population, continued legislative domination by the planter masters. The constitution of 1861, which substituted the phrase "Confederate States" for "United States," and its successors of 1864 and 1868, the latter extending the suffrage to all males, black as well as white, may be called the Civil War documents. The constitution of 1879, marking the end of the Reconstruction period, restricted the action of the legislature and granted executive powers rivaling those of 1812, while the constitution of 1898 effectively disfranchised the black citizens of the state. The constitutions of 1913 and 1921 were written by delegates of conventions called to grapple with the problems of the 20th century: in many respects they failed, primarily because of the shadows of the past that hung over them. A new state constitution was established in 1974.

The governor of Louisiana remains the state's most powerful official, not only from the weight of tradition (and personal performance) but also because of the extent of patronage among the many executive agencies, boards, commissions, and offices filled by gubernatorial appointment. The governor is elected for a term of four years and is permitted to serve no more than two consecutive terms. Several additional state officials are elected; others are appointed. The Louisiana legislature has two houses: the Senate, with 39 members, and the House of Representatives, with 105 members. Legislators are elected for four-year terms. The 1974 constitution retained a 1956 amendment that requires a two-thirds vote by both houses on taxation measures in order to curb spending money by the governor.

The parish system

Local self-government in Louisiana followed the Virginia system of county government. The parish (county), the municipality, and the special district are the units of local government. There are 64 parishes, with land areas that vary from the 199 square miles in Orleans parish to the 1,441 square miles in Cameron parish. The name of the elected parish governing board, "the police jury," is not found anywhere else.

There are about 300 incorporated municipalities in Louisiana, described as state units, which exercise narrowly construed powers. The charter of incorporation detailed by law outlines three classes of municipalities based upon population: city (5,000); town (1,000–4,999); and village (150–999). Special districts established by the legislature provide for the administration of new or expanding functions of local government.

A 1975 statute mandated an open primary system for the election of state officers. Under this system, all candidates appear on a single ballot. The candidate who receives more than 50 percent of the votes in the primary is elected to office. If no candidate receives a majority of the vote, a runoff election is held between the two candidates who received the most votes. Previously, Democratic nomination was tantamount to victory. In the open primary, however, the division of the Democratic vote by multiple candidates practically assured a runoff position for a lone Republican candidate. In consequence, Louisiana in 1979 elected its first Republican governor since Reconstruction.

The highest court in the state is the Supreme Court. Its seven justices are popularly elected for 10-year terms. Judges in the Court of Appeals, the district courts, and most of the lesser courts are also popularly elected.

Louisiana's legal system differs from that of the other 49 states in that it is based not on common law but on civil law, which is code, or written, law. The state draws upon its colonial inheritance, whereby the adopted code was based upon the Code Napoléon of France and further influenced by Spanish laws, both of which had a common source in Roman law. The civil law consists of broad principles drafted by authorities in various fields of law. In Louisiana the law is enacted in the constitution, which vests in the legislature the authority to make law, whereas the functions of the courts are limited to the application of the law to given sets of facts. Courts are not bound by previous decisions. The law governs all personal and property rights and has been extended to civil and criminal procedures.

Code Napoléon

Education. The constitution of 1974 revamped administration of education by establishing five boards: (1) a state board of regents to oversee higher education, (2) three separate management boards for Louisiana State University, Southern University, and the other state colleges and universities, and (3) a policy board for elementary and secondary education. Public elementary and secondary education were, beginning in 1988, administered by an appointed superintendent.

Louisiana State University, which is both the land-grant and the arts and sciences university, is the keystone of higher education in the state, with campuses in Baton Rouge, Eunice, Shreveport, and Alexandria. Its former campus in New Orleans has become the University of New Orleans. Louisiana Tech University, in Ruston, and the University of Southwestern Louisiana, in Lafayette, are among the state's other universities, many of which grew from junior colleges in the 1930s and '40s. Southern University, in Baton Rouge, with campuses is Shreveport and New Orleans, is a large public black university, as is Grambling State University, in Grambling. Tulane University, in New Orleans, has long excelled as a private institution. The Roman Catholic schools Loyola and Xavier and the private black school Dillard are all in New Orleans.

Health and welfare. The legislature has established programs to provide a system of economic security and social welfare for various categories of citizens, including persons 65 years of age and over. The state gives aid and welfare to mothers and children and provides aid to the disabled and to the needy blind. Various state departments provide some aspects of welfare aid, but by far the most important is the Department of Health and Human Resources. The department provides services to Louisiana citizens through a central office in New Orleans and local units in parishes. Penal and correctional institutions operated by the state are administered under the general authority of the Department of Public Safety and Corrections. The penal system has often suffered from an excess of political interference.

The so-called Charity Hospital system, supported and administered by the state, is fairly unusual among the 50 states. The system maintains several general hospitals, as well as hospitals for the mentally ill. The Charity Hospital of Louisiana in New Orleans received some public support after its founding in 1811 but has long since been funded by the state.

Cultural life. The extensive power of the Roman Catholic church in southern Louisiana and the domination of Baptists in northern Louisiana and among the black population remain important influences on social and cultural life. New Orleans and many smaller communities have been able to support the arts and philanthropic

institutions. The Creoles (descendants of French or Spanish settlers) developed a distinctive architecture, art, and cuisine centred on New Orleans. Planters emulated the Creoles, and, thus, the people of the alluvial parishes of northern Louisiana are more cosmopolitan in outlook than are the people of the uplands.

In folk culture and the arts Louisiana more than holds its own. This is especially evident in the realm of music, whether it be in black folk songs, including the celebrated rural blues; the Cajun bands at fais-dodos, country dances held in southern Louisiana; the community hymn singings of northern Louisiana; the jazz that New Orleans migrants took to Chicago and elsewhere; or the renaissance in Dixieland music played by bands at New Orleans' Preservation Hall.

Musical traditions

New Orleans supports an opera company and a symphony orchestra. It was a major cultural centre during the 19th century. Its French Quarter attracted such artists as John J. Audubon, the great wildlife painter, and George Catlin, noted for his portrayals of the American West, together with writers such as Walt Whitman, Sherwood Anderson, and William Faulkner. Since the 1930s other cities, notably Shreveport, Monroe, Baton Rouge, and Lafayette, have evolved their own museums and galleries, orchestras, choruses, and little theatres.

Tourism has developed as an important industry using the appeal of the antebellum past and the attraction of Creole cuisine, a blend of French, Spanish, black, and Indian dishes. A series of parades and balls culminating in Mardi Gras (Shrove Tuesday) has become a national attraction in New Orleans. There are many public parks and gardens, and the state is advertised as a sportsman's paradise for hunting and fishing.

The literate Creole culture provided the state with a long press tradition, the first newspaper, *Le Moniteur de la Louisiane,* appearing in 1794. Eight others were published in New Orleans at the turn of the 19th century, and the rural parishes likewise published their own papers. The leading newspapers are concentrated in the urban parishes.

HISTORY

Early settlement. At least 16,000 years before European exploration, Indians occupied the region that was to become Louisiana. At least seven archaeological sites have been excavated, notably the so-called Poverty Point sites (approximately 700 BC) and the Marksville site (AD 100 to 550). Most Louisiana Indians lived in hunting and gathering camps in the uplands and coastal prairies, though there were farming villages in the rich, low-lying areas known as bottoms. It is estimated that there were 15,000 Indians in the area when settlement by Europeans began during the 1700s. By 1980 only about one-fifth as many Indians remained. Their heritage is present in many place-names that lend colour to the state's map.

Initial colonization

While the Spanish were the first Europeans to discover the area, it was the French who colonized it. Serious colonization by France began in 1702, when Pierre Le Moyne d'Iberville and his brother Jean-Baptiste Le Moyne de Bienville explored the area and struggled to found permanent colonies. The city of New Orleans was established by Bienville in 1718. Royal charters covering the area had been granted, first to Antoine Crozat, in 1712, and then, in 1717, to the Scottish businessman John Law, whose Company of the West failed in 1720. When Louisiana became a French crown colony in 1731, its population had grown from fewer than 1,000 to nearly 8,000, including slaves. In addition to the French settlers, many thousands of Germans arrived, settling on the river just above New Orleans on what became known as the German Coast. Colonization was increased during the 1760s with the arrival of the French-speaking Acadians, who had been expelled from Nova Scotia by the British.

In 1762 Louisiana and New Orleans were ceded to Spain by a secret treaty that was to establish nearly four decades of Spanish rule and influence in the area. In 1779 the Spanish wrested Baton Rouge from the British and took all West Florida, which then extended from the peninsula westward across the Gulf Coast to the Mississippi River. In 1800 the Spanish re-ceded Louisiana to France, and in 1803 the United States concluded the Louisiana Purchase.

The 19th century. Louisiana was subsequently divided into the Territory of Orleans, which consisted essentially of the state within its present boundaries, and the Territory of Louisiana, which included all the vast area drained by the Mississippi and Missouri rivers. In 1810 the Territory of Orleans consisted of 77,000 people, and statehood proposals were beginning to be heard. When, in 1812, the territory petitioned to enter the Union, the eastern region, now called the Florida Parishes—where the people had rebelled against the Spanish and established the Republic of West Florida—was included.

There was an economic boom during the 1830s, generated by slave labour toiling on the flourishing sugarcane and cotton plantations, and sets of natural cleavages emerged in the political affairs of the state as French–American, and later planter–farmer, interests clashed in the political process. While the yeoman farmer held the suffrage, representation rested in the hands of a plantation aristocracy that overcame one-man, one-vote principles by counting slaves in the determination of district units. Under this circumstance, and with the breakdown of the two-party system during the 1850s, sentiment in the state was divided on the issue of secession from the Union. The pro-secession group prevailed in the convention of 1861, even though later research would make it appear that a majority of the citizens wanted to stay in the Union.

Separation was short-lived in southern Louisiana, for by May 1, 1862, New Orleans was occupied by Union forces. Following the end of the war, Louisiana was readmitted into the Union in 1868, and a severe Reconstruction period began. Political conflict occurred between the federal Republicans who were located in New Orleans and the former Confederates from the rural parishes. After 1876 the Democrats contested with the Republicans as the freed black citizen, whose vote represented the balance of power in the state, became the pawn in the electoral struggle. A number of clashes occurred between the factions, the most noted of which was Sept. 14, 1874, in New Orleans, when the White League briefly wrested control of the city from the Republican police. In 1876 the Democrats claimed that General Francis T. Nicholls was elected governor, but the Republicans claimed that S.B. Packard had won. Their claims were intertwined with the choice of presidential electors for that year in the famous Hayes–Tilden dispute. The Republicans manipulated the state returning board and sent two sets of election returns to Congress, and the Democrats sent their returns. The electoral commission accepted the Republican electors, just as it did those in dispute from South Carolina, Florida, and Oregon. Both Nicholls and Packard took the oath as governor in January 1877 and set up rival governments, which continued until President Rutherford B. Hayes, elected as a part of a bargain, ordered the withdrawal of federal troops from the capital on April 20, 1877, and the white Democratic Party was left in control.

The Civil War and its aftermath

The plantation economy continued as the farmer class, white and black alike, was squeezed from farm ownership and forced into sharecropping or tenancy. Agrarian protests that emerged during the 1880s and '90s produced the People's (Populist) Party and what seemed at the time to be a chance to overthrow the state's planter–merchant–lawyer rule. By the early 20th century, however, Louisiana was under a restrictive rule, as the elite was able to defeat the reform movement of the farmers in the gubernatorial election of 1896 and to enact the constitution of 1898. As a result, nearly all blacks were legally denied the right to exercise the franchise, while many of Louisiana's whites lost the will to do so.

The 20th century. Extensive lumbering attracted large corporations to Louisiana for three decades following 1890, and the discovery of oil and gas reserves helped to increase industrial development. While these trends may have laid the foundations for the eventual development of the Louisiana economy, the political leadership of the state was not ready to take advantage of these developments in terms of increased tax revenues and services to the people. Louisiana, therefore, continued to be a backward, segregated, and primarily agrarian society.

Part of the rise of the demagogic and populist Huey P. Long to the governorship during the late 1920s may well be attributed to the seriously arrested socioeconomic development of the state. With the support of the rural areas and the emerging working class, Long substituted a realism for the romance perpetuated by the conservative leadership. Under his administration welfare benefits and educational services were extended, and bridges, roads, and hospitals were constructed, not on the pay-as-you-go basis of the past but through the floating of bonded indebtedness. Since the rise of Longism and its perpetuation under Huey's brother, Earl K. Long (elected governor in 1948 and 1956), no political administration has seen fit to turn back the series of public benefits financed by increased taxation.

Policies of Huey P. Long

During and after World War II Louisiana experienced further economic development, which was heavily committed to the petrochemical industry and increasingly concerned with offshore oil and gas drilling in the Gulf of Mexico. Since the 1960s a "second Reconstruction" brought about by U.S. Supreme Court decisions on school desegregation and voting rights for blacks has had important consequences. Conflict, focused on race and religion, broke the bridge the Longs had built between northern and southern Louisiana.

Governor John J. McKeithen, who took office in 1964, stayed the furor over racial integration by building a firm consensus on "law and order." He was reelected in 1968 after successfully sponsoring a "second-term" constitutional amendment. Governor Edwin W. Edwards (elected 1972, reelected 1976) reaped benefit from an improved social atmosphere and expanding economic conditions in what was then called "oil-rich Louisiana." Unable to seek a third consecutive term in 1980, Edwards won an unprecedented third election in 1984, following the administration of Republican David Treen. A combination of southern Louisiana Cajun and black electoral power set the pattern.

Louisiana's progress in meeting the challenge of building a racially integrated society is evidenced by the fact that the ratio of blacks to whites among registered voters is nearly equivalent to that in the population as a whole. The potential to extend rights and opportunities is in place. The relative underdevelopment of the state, however, continued to be apparent in questions of political corruption and in the economic depression created by the world oil glut, as the state government had become heavily dependent on oil severance taxes. (P.H.H.)

Mississippi

Originally part of America's Old Southwest, Mississippi became the 20th state of the Union in 1817. Its name has long been identified with many of the characteristics attributed, correctly or incorrectly, to the Deep South. Since the 1960s the state has been engaged in efforts to alter the economic and social patterns of the past. Mississippi ranks 32nd among the states in area. Jackson is the state capital. The name is derived from an Indian word meaning "great waters" or "father of waters."

For decades an unusually large dependent population, a predominantly agricultural economy, and a prevailing resistance to change have kept Mississippi's per capita income low and created an inadequate standard of living for many families. At least half of all Mississippians live in rural areas—but not necessarily on farms—and the state continues to rank low in many economic indexes, including per capita income, which is well below the national average. In 1965 industrial income surpassed agricultural income for the first time in the state's history.

Throughout most of its 47,689 square miles (123,515 square kilometres)—from Tennessee on the north to Louisiana and the Gulf of Mexico on the south, from Alabama on the east to Louisiana and Arkansas on the west—much of Mississippi's soil is rich and deep, and its low-lying landscape is laced with many rivers. Almost inevitably it became an agricultural state. The long dominance of a rural, unhurried way of life has contributed much to the problems of present-day Mississippi, just as it earlier helped to enhance the state. This way of life has also left a sense of history among some Mississippians, whose ancestors created a culture of gentility that is still evident in the many historic mansions located in such old towns as Columbus, Biloxi, Natchez, and Holly Springs.

PHYSICAL AND HUMAN GEOGRAPHY

The land. Mississippi is a low-lying state, its highest point reaching only 806 feet (246 metres) above sea level. Except for its hilly northeast corner, Mississippi lies entirely within the eastern Gulf Coastal Plain physiographic region. It has generally low topographic elevations and extensive tracts of marshy land. Its major soil areas encompass hills, plains, prairies, river lowlands, and pine woods.

Relief. In the northwest the great fertile crescent called the Delta is the old floodplain of the Yazoo and Mississippi rivers, comprising 4,000,000 acres (1,600,000 hectares) of black alluvial soil many feet deep. Once subject to disastrous floods, the land is now protected by levee and reservoir systems. Though the Delta was only sparsely settled in antebellum days, it has become a region of highly mechanized farming.

The Delta

East of the Delta looms a high wall of loess bluffs, marking the beginning of the highlands, or hills. A brown loam belt of varying width extends from Tennessee to Louisiana. Most of southern Mississippi lies in the high and rolling Piney Woods. Though settled early, the area did not prosper until the early 20th century, when the great virgin pine forests were exploited and heavily reduced. It has since become a prosperous area based on diverse forest industries, cattle, and some specialty farming.

The coastal area, sometimes called the Coastal Meadows, or Terrace, borders the Gulf of Mexico. The soil is sandy and not well suited to crops, but its location, climate, and industrialization have made the region important.

Along the northern edge of the Piney Woods lies the narrow Central Prairie, separated from the Black Prairie by a section of hills and woods. The two prairies, with fertile black soil that is excellent for many types of agriculture, were once the site of cotton plantations. East of the Black Prairie, in the extreme northeast, are the Tennessee Hills. Arching between Tennessee and Alabama, these hills form the only area in Mississippi in which the terrain and people are reminiscent of the southeastern mountains.

West of the Black Prairie another highland area, the Pontotoc Ridge, ranges south from the Tennessee border. This ridge, averaging 400 to 600 feet (120 to 180 metres) above sea level, is one of the state's most distinctive features. Its fertile, sandy loam is excellent for orchards. A low-lying region called Flatwoods skirts the western edges of the Pontotoc Ridge and the Black Prairie. Its heavy clay soils drain poorly, and the area has never developed a prosperous economy. The North Central Hills range through northern and central Mississippi and eastward to Alabama. Their red clay soil supports small farms. Before scientific farming methods were widely adopted, erosion ruined thousands of acres in these hills.

Pontotoc Ridge

Drainage. Mississippi has many rivers, creeks, bayous, and other drainage. The state's principal river systems include the Tombigbee, now joined with the Tennessee to form the Tennessee–Tombigbee Waterway, in the northeastern section; the Pascagoula in the southeast; the Pearl in the south central section; and the Mississippi and its tributaries, notably the Yazoo and the Big Black, in the west. These streams all empty into the Gulf, either directly or through the Mississippi and other rivers.

Climate. Mississippi's location endows it with a favourable climatic range. The growing season is long (virtually year-round on the coast), rainfall is abundant, and extreme temperatures are unusual. Summers are warm, with temperatures sometimes exceeding 95° F (35° C). Autumn's bright, crisp days have the least rain and are usually the most agreeable of the year. January temperatures average from 42° F (6° C) to 50° F (10° C). Snowfall is rare but does occur. Supplies of water are abundant, and rainfall is usually adequate for replacement. The state's annual average is more than 50 inches (1,270 millimetres), varying by region. The coastal area is subject to hurricanes from June to October.

Plant and animal life. The mild climate, long growing season, and abundant rainfall provide Mississippi with a remarkable variety of plant and animal life. Live oaks and palms vary the landscape of the southern counties, and fruit trees and hardwoods thrive in the north. The magnolia and pecan trees are favourites throughout the state. Pine forest, often intermixed with oaks, is found extensively on the state's sandier soils. More than half the land area is in forests, and both natural and cultivated floral displays are diverse and abundant.

Opening land to farming and hunting reduced the once abundant wildlife to near extinction. The wolf and puma (cougar) are gone; the bobcat is rare and the bear even rarer. Yet deer are once more abundant, and wild turkeys have increased. The state has a variety of resident and migratory birds. Some game fish can be taken throughout the year, with catfish, bream, bass, and perch the leading freshwater species. The Gulf is rich in shrimp, oysters, and fish, the mainstays of extensive commercial fishery.

Settlement patterns. Almost every Mississippian knows and uses such regional designations as the Delta and the Hills, South Mississippi and the Coast, the Prairie or the Black Belt, and North Mississippi and the Northeast. The landscape of Mississippi is mainly one of forests, fields, and towns. Cities and factories take a lesser place. Soils unsuited to row crops may support tree farms, pastures, or orchards. The largest metropolitan areas are Jackson, Biloxi–Gulfport, and Pascagoula.

Regional distinctions

Geographers may speak of the Yazoo basin, but to the people of Mississippi it is the Delta. The white Deltans pay homage to aristocratic plantation traditions and tend to be politically conservative. The Hill people, however, do not defer to the Deltans, many of whose families originally came from the Hills.

What historians call the Piney Woods, covering most of the state south of Jackson, the people call South Mississippi. Its population and prosperity have grown in the 20th century, and change seems to come easier there than in most other regions. Farmers are rapidly being replaced by workers in commerce and industry.

The coastal area is atypical of the state as a whole. More Roman Catholic and Mediterranean than upstate, it blends French, Spanish, Latin-American, Dalmatian, and British heritages in the most heterogeneous of Mississippi's regions.

The Prairie region has some of the ways and style of both the Delta and the Hills. Its lands, however, are more productive than the Hills, and some traditions from the antebellum era remain. Northeast Mississippi developed as an area of small family farms and few plantations, and it has the lowest nonwhite population in the state.

The people. The white population of Mississippi is remarkably homogeneous. More than 98 percent native-born of native stock, whites are predominantly of British, Irish, and northern European ancestry. The black, Choctaw Indian, and Chinese segments of the population are also almost entirely native-born.

Until about 1940 blacks were in the majority, but by the late 20th century, largely because of a very high rate of out-migration, blacks made up only a third of the population. A few thousand Indians (mostly Choctaw) live in the east central section of the state. The small Chinese population found in the Delta is descended from farm labourers brought there from California in the 1870s. The Chinese did not adjust well to the Mississippi plantation system, however, and most of them became small merchants. The coastal fishing industry has attracted Southeast Asian refugees.

Various Protestant denominations dominate Mississippi's religious life, most notably Baptists and United Methodists. The Roman Catholic population is mainly concentrated in the urban centres and the southernmost areas, especially the coastal counties. The Jewish community is almost entirely urban.

Mississippi has no great extremes of population concentration or extensive uninhabited areas. Since 1950 there has been a slow but steady loss of farm population and a decline of smaller towns, but most of the centres of more than 10,000 inhabitants have had significant growth.

Patterns of migration

Mississippians, who inherited the frontier tradition of "moving on," have become as mobile as other Americans. Frequent movement by sharecroppers and tenant farmers from one farmstead to another was commonplace before 1920, at about which time the economic focus (and the locus of emigration) shifted to the cities and towns. About three-fourths of the white emigrants moved to other Southern states, whereas the same proportion of black emigrants left the South entirely. The net loss by emigration has largely offset Mississippi's high rate of natural increase. Significant population growth in Mississippi would require that the state retain more of its young people.

The economy. In many economic indexes Mississippi ranks low in its region and in the nation. There has been some improvement in employment, wages, and personal income; but the proportionate national and regional growth has been even greater, and the relative economic disadvantage continues. Nearly 95 percent of personal income is derived from eight sources—manufacturing, the federal government, property, farms, state and local governments, wholesale and retail trade, operation of nonfarm commercial enterprises, and personal and business services.

State agencies administer regulatory functions in the area of utilities, transportation, oil, gas, insurance, and pollution. The Department of Economic Development seeks new businesses and aids in expanding existing ones by such means as loans for training and recruiting workers. It is aided by the Mississippi Research and Development Center through research in economic development. The several economic-development districts promote activities in their constituent counties. The private sector also advertises the state's advantages.

Labour union membership

Labour union membership is relatively small, although widely dispersed. Most large employers have a union membership, though Mississippi has a right-to-work law that prohibits compulsory union membership.

Resources. Petroleum and natural gas account for more than nine-tenths of the value of all minerals produced. Important nonmetallic minerals include sand and gravel, fuller's earth, and other clays. Iron has been mined intermittently since 1887. Aluminum ores are low in quality, and they have been little exploited.

About half of the land area is in commercial forests that produce lumber, paper pulp, naval stores, and other forest products. Seafoods from Gulf waters are processed in coastal plants.

Electrical power produced within the state is from steam-generating plants and a nuclear power station near Port Gibson. Hydroelectric power is brought into Mississippi from Tennessee Valley Authority dams and through interconnections with power companies in other states. A few private companies, numerous rural cooperatives, and several municipal generating systems are in operation. Several large transmission facilities bring natural gas into and through the state.

Agriculture. Since World War II Mississippi's economy has become less dependent on agriculture, and both the number of farms and farm acreage have declined. Cotton is no longer king in Mississippi. The state's principal agricultural income derives from livestock, catfish farms, and poultry and from various crops, with soybeans and cotton being the most prevalent. Approximately 70 percent of the state's farms produce livestock and dairy products, while only about 10 percent of the farms produce cotton.

Industry. In 1936 Mississippi began a program to promote the expansion of manufacturing, and in 1965 the state's industrial employment exceeded the number of agricultural workers for the first time. The state's major manufactured goods include forest and chemical products. The Ingalls Shipyard, at Pascagoula, is the state's largest employer.

Transportation. The declining fortunes of rail transportation and the existence of obsolete segments of the state's highway system have created problems in transportation. The heaviest volume of traffic is along the Gulf Coast, where it merges into the flow from the numerous upstate north–south and east–west patterns. The 200-mile Natchez Trace Parkway, which extends from Natchez to

Nashville, Tenn., and is part of the National Park Service, is protected from commercialization and truck traffic. Peaceful and sylvan, it preserves the natural surroundings of the old Indian trace, or trail, and encompasses many sites of historic interest.

Commercial transportation, in addition to highway trucking, is diversified. Of the railways in the state, half are entirely intrastate. Most large cities have commercial airports with scheduled air service. Interstate bus lines serve almost all major cities and towns. Gulfport and Pascagoula can accommodate oceangoing ships, and low-draft oceangoing vessels can travel up the Mississippi to Natchez, Vicksburg, and Greenville. Barge traffic moves on the Mississippi, Pearl, and Yazoo rivers. The Gulf Intracoastal Waterway passes just off the coast across the Mississippi Sound.

Administration and social conditions. *Government.* The state government has executive, legislative, and judicial branches, but it differs from those of some states in that most heads of executive departments are elected rather than appointed. The bicameral legislature, which meets in annual sessions, comprises a 122-member House of Representatives and a 52-member Senate. Members of both houses are elected to four-year terms.

The executive branch of state government includes the governor, lieutenant governor, and eleven other officials, all of whom are elected to four-year terms. Executive officials can succeed themselves.

The lowest court in the state's judicial system is the justice court, which has original jurisdiction in misdemeanors for which fines, sentences, and judgments do not exceed prescribed limits. Some large counties also maintain a county court. Chancery courts have jurisdiction over matters of equity, probate, juvenile delinquency (where county courts do not exist), divorce, and mental competence. Circuit courts are the main trial courts for major suits, criminal cases, and appeals from justice and county courts. The Supreme Court is the court of appeal; its nine justices are elected, from three judicial districts, for staggered terms of eight years. All other judges are elected to four-year terms.

Each of Mississippi's 82 counties is governed by a board of supervisors. Municipalities may be incorporated as villages, towns, or cities. Governments of these units are of the mayor–council type with aldermen, the commission type with three commissioners (including the mayor), or the city-manager type, in which the manager is appointed by the council. State and county officials are elected in a November general election. Party primaries are held in August.

Race and politics

From the end of Reconstruction until the late 1940s, the Democratic Party was essentially the only party in Mississippi. As in many Southern states, literacy tests, poll taxes, and other restrictive measures kept blacks at a proportionally low level of political involvement. Disaffection among Mississippi's white voters with the national Democratic Party broke its domination in 1948, and thereafter, with few exceptions, states' rights or conservative Republican presidential candidates received the state's electoral ballots. Increasing participation by blacks, made possible by the federal Civil Rights Act of 1964 and the Voting Rights Act a year later, further redefined party politics in Mississippi. For a time, two Democratic parties existed—a predominantly black Loyalist faction allied with the more liberal national party, and a conservative wing favouring the traditional state power structure. The party eventually unified, with the Republican opposition gaining support among conservative Mississippians. Since the 1960s blacks have won election to an increasing number of local and state offices throughout Mississippi.

Finances. Sales taxes are the state's major source of revenue, followed by personal and corporate income taxes and gasoline taxes. Local governments derive their greatest income from property taxes. Considerable amounts of federal monies are provided through numerous federal and state agencies.

Education. Mississippi's public school system has been ranked at the bottom of almost all measurable standards for many years. Following years of turmoil brought on by the Supreme Court's 1954 ruling against segregation in public schools, the state committed itself to a dramatic improvement of its schools. The Education Reform Act of 1982, the reinstatement of compulsory attendance for children between the ages of six and 14, competency testing of high school students, the creation of an appointed state school board, and the establishment of an early childhood education program are measures that the state has taken to enhance the quality of education. The success of these measures is evident in the diminishing school dropout rate and the steady improvement in educational test scores.

Higher education

Mississippi has a distinguished history of higher education, however. Although it closed in 1826 and again in 1863, Jefferson College, founded in 1802, was among the earliest public institutions of higher learning in the nation. Elizabeth Female Academy, which is sometimes considered to be the first women's college, was established at Washington in 1818. In 1878 Mississippi established Alcorn Agricultural and Mechanical College (now Alcorn State University), the first land-grant college for blacks in the United States, and in 1884 it established the Industrial Institute and College (now the Mississippi University for Women), the first state-supported institution granting diplomas to women. The University of Mississippi, in University, near Oxford, was chartered in 1844 and opened in 1848. In addition to the more than 40 junior colleges, colleges, and universities, other educational facilities include a medical centre, the Gulf Coast Marine Research Laboratory, and the Mississippi Sea Grant Consortium.

Health and welfare. Education, health, welfare, and other measurements of the quality of life in Mississippi necessarily must be considered in the perspective of the state's long history of segregation and racial discrimination. The high infant mortality and illiteracy rates, low educational achievement scores, and welfare dependency are linked to a century of poverty, injustice, discrimination, and a resistance to modernization that has impeded the development of the state's natural and human resources. Almost all counties have some form of relief or welfare programs that involve federal funds. Welfare services include aid to the blind and disabled, the elderly, and dependent children. Health programs are administered by several state agencies, including the Board of Health, which dates from 1877.

Historically, the state has been in the vanguard of public health services. The causes of pellagra were discovered in 1915 through experiments at the state penal farm. A model mosquito-control program eradicated yellow fever, and the state tuberculosis sanitarium became recognized nationally. Pioneer work has continued in Mississippi in the education of blind and deaf children.

Cultural life. Before the Civil War the "Planter Society" and those who identified with it had a highly developed sense of gentility. The life-style to which they aspired made patronage of the arts obligatory. They built Greek Revival mansions and furnished them with art objects and fine furniture, their children were tutored in the social graces and the arts, and hospitality became an art in itself. The rural gentry, however, was only a very small part of the total society. The small landowner and slave alike fashioned simple, sturdy furniture, made their own oxbows and spinning wheels, and patiently crafted musical instruments. When they rested, they heard folk songs, ballads, or African lullabies. Their literature was myth, legend, and tall tale, and their dances were traditional or improvised.

From these sources came the present-day cultural and artistic heritage of Mississippians. In the 20th century, technology, mobility, and mass communication have created an adherence to a national culture. A freshening sense of history is evident, however, in the efforts to preserve historic landmarks and in the intensity of the collectors of the artifacts and furnishings of the past—of folk songs, implements and utensils, furniture, and manuscripts.

Mississippi writers

Mississippi has been a vital part of the flowering of Southern literature during the 20th century. The mythical county of Yoknapatawpha and the generations of its people were created by William Faulkner in a celebrated series of novels. The works of this Nobel Prize-winning Mississippian are often ranked among the highest attainments in American literature. Other natives of international literary renown include novelists Eudora Welty and

Richard Wright, novelist-critic Stark Young, and playwright Tennessee Williams. Mississippi's "second generation" of writers includes Elizabeth Spencer, Walker Percy, Willie Morris, and Ellen Douglas; more recently, Barry Hannah, Richard Ford, and Pulitzer Prize-winning playwright Beth Henley have attained national prominence.

Newspapers

Mississippi has produced several famous newspaper editors, including Turner Catledge, of the *New York Times,* and the Pulitzer Prize winner Hodding Carter, of the Greenville *Delta Democrat-Times.* All of the state's large towns are served by local dailies, and the smaller towns and communities are served by one of the strongest systems of weeklies in the United States.

In music, both white and black folk traditions are found, in English and Scottish ballads and in blues, spirituals, and sacred harp singing. This rich heritage has given rise to such acclaimed performers as Elvis Presley, blues artist B.B. King, and lyric soprano Leontyne Price. The state has several symphony orchestras, an opera guild in Jackson, and extensive musical activities at several colleges. The theatrical tradition dates from 1800, when a Natchez audience saw the first dramatic production to be presented west of the Alleghenies. Today some 20 community theatres and 30 colleges and universities offer dramatic fare, in addition to a semiprofessional company in Jackson.

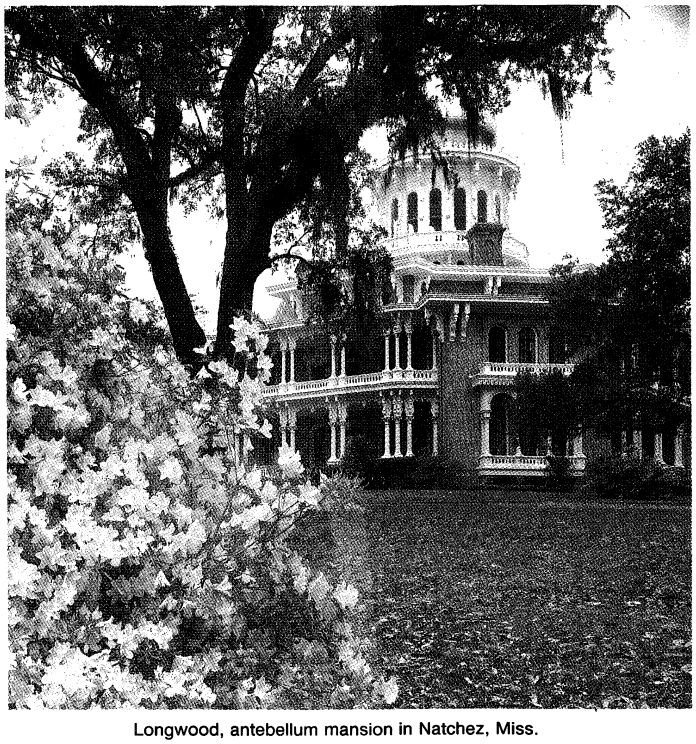
Mississippi's rural heritage continues to be a strong influence on the life-styles and recreational habits of the people. Hunting, fishing, boating, camping, and other outdoor activities are among the most popular forms of leisure in the state. Mississippi maintains a system of state parks, and the U.S. Department of the Interior maintains the Vicksburg National Military Park and the Natchez Trace Parkway. The Natchez Pilgrimage is the best known of several festivals featuring antebellum homes and gardens.

HISTORY

The Indians. The Indians of Mississippi lived in harmony with the environment of the Southern woodlands and took great care to maintain the ecological balance they found in nature. Early European travelers often spoke of the Indians' love for the land and their bravery in defending it.

The Choctaw tribe, which numbered approximately 20,000 and was located primarily in the southern and central part of the state, was the largest of the three major tribes that lived in present-day Mississippi. The other two tribes were the Natchez, which numbered about 4,500 and were centred in southwestern Mississippi, and the Chickasaw, which had a population of about 5,000 and ranged from their principal villages in the northeastern part of the state into Tennessee and Kentucky. The Natchez were virtually exterminated during a war with the French garrison at Fort Rosalie (the present city of Natchez) in 1729–31. The Choctaw and Chickasaw were eventually removed from Mississippi to the Oklahoma territory in the 1830s.

© Tom Till



Longwood, antebellum mansion in Natchez, Miss.

Exploration and settlement. In the winter of 1540 Hernando de Soto led a large expedition into Mississippi and wintered along the Pontotoc River. In the following spring he reached the Mississippi River, but, because he found no gold or silver in the region, Spanish explorers directed their efforts elsewhere.

Nearly 130 years later a small group of French Canadians sailed down the Mississippi River and immediately realized its commercial and strategic importance. In 1699 a French expedition led by Pierre le Moyne d'Iberville established France's claim to the lower Mississippi valley. French settlements were soon established at Fort Maurepas, Mobile, Biloxi, Fort Rosalie, and New Orleans.

Following the French and Indian War, which ended in 1763, France ceded its possessions in the lower Mississippi valley, except New Orleans, to Great Britain, which also gained possession of Spanish Florida and divided that territory into two colonies. One of those was West Florida, which included the area between the Apalachicola and Mississippi rivers. The original northern boundary of West Florida was the 31° parallel, but it was extended in 1764 to the 32°28′ parallel. Fort Rosalie was renamed Fort Panmure, and the Natchez District was established as a subdivision of West Florida. Natchez flourished during the early 1770s. After the outbreak of the U.S. War of Independence, Spain regained possession of Florida and occupied Natchez. The Treaty of Paris of 1783 fixed the 31° parallel as the boundary between Spanish Florida and the United States, but Spain continued to occupy Natchez until the dispute was settled in 1798.

The Natchez District

Statehood and Civil War. The original Mississippi Territory created by the U.S. Congress in 1798 was a strip of land extending about 100 miles north to south and from the Mississippi River to the Chattahoochee on the Georgia border. The territory was increased in 1804 and 1812 to reach from Tennessee to the Gulf. In 1817 the western part achieved statehood as Mississippi (the eastern part became the state of Alabama in 1819). Natchez, the first territorial capital, was replaced in 1802 by nearby Washington, which in turn was replaced by Jackson in 1822.

The 1820s and '30s were marked by the decline of the Jeffersonian Republicans, the ascendancy of the Jacksonian Democrats, and the removal of the Indians to Oklahoma. They were the days of steamboats, land speculation, and the growth of a plantation-based cotton economy, with its concomitant slave population. Slave owning, however, was not common among the small landowners, who became more numerous than the large planters but who had little influence on public affairs for many years.

Sectionalism in both North and South had been growing for some time. Its ill feelings gradually became dominated in both North and South by slavery. In January 1861, a convention adopted an ordinance of secession, and within a year the state was in the midst of war. The people suffered much privation, and the land underwent great devastation; by 1865 the state was in economic ruin.

The aftermath. For 25 years following the Civil War, Mississippi's former slaves and their former owners grappled with the political, social, and economic consequences of emancipation. The white minority could not or would not accept a biracial society based on equality of opportunity. In 1890 the ruling elite adopted a constitution that established a caste system of racial segregation and an economic order that kept blacks in a position of dependency.

Mississippians hoped to find economic salvation in the coming of industry and the railroads, but the hope was only partially realized. Emancipation had made the former slaves free to go where they wished, but most remained and eventually were absorbed into the tenant-farming system. The continued economic interdependence of the two races

kept intact many of the customs and social systems that had developed before the war. The constitution of 1890 effectively disfranchised most of the black population.

World War I hastened the end of Mississippi's physical and psychological isolation, and most of the bitterness remaining from the Civil War was lost in a surge of patriotism. Between the wars the state was affected by the agricultural depression of the 1920s, the devastating 1927 flood, the Great Depression of the 1930s, the coming of farm-production controls, and the beginnings of new industrialization. After World War II government farm programs and mechanization on a broad scale created another agricultural revolution.

The Civil Rights movement. After World War II a series of events developed that may be characterized as a revolution in race relations. Many long-accepted practices, customary throughout much of the South, received their first major jolt in 1954 with the U.S. Supreme Court decision declaring racially segregated schools to be unconstitutional. The decision was followed by years of increasing protest against other aspects of segregation and by large-scale registration of black voters. Whites reacted to the black protest against segregation with increasing violence during the early 1960s. In 1962 state officials refused to abide by a U.S. Supreme Court ruling that ordered the admission of a black student, James Meredith, to the University of Mississippi. Following a night of rioting during which two people were killed, Meredith was finally admitted and the colour barrier was broken in Mississippi.

Tensions between tradition and change

The most serious violence occurred in the summer of 1964, when three young civil rights workers were murdered by the Ku Klux Klan. Those murders convinced most white Mississippians that continued resistance and violence was a greater danger to their safety and welfare than was peaceful acceptance of change.

In 1969, under a federal court order, the state's dual segregated school system was unified. Although the great majority of white Mississippians opposed school integration, they adjusted to that change with only minor isolated incidents of violence. Over the next decade Mississippi's long-standing racial traditions and customs were dramatically altered. A succession of enlightened and progressive governors, including Paul B. Johnson, Jr., William L. Waller, William F. Winter, and Ray Mabus, led Mississippi out of its troubled history as a rigidly segregated closed society into a new era of racial cooperation.

After accommodating themselves to such change, Mississippians could at last turn their attention and energy to the development of the state's human and natural resources. Economic development during the 1960s and '70s, though not spectacular, was steady. Out-migration of whites has virtually ceased, and among blacks it has declined significantly. In 1960 Mississippi's income per capita amounted to only 55 percent of the national average, but by 1985 it had risen to 69 percent. This increase was a result of urbanization and industrialization and the decline in agricultural employment.

A development that both paralleled and promoted economic and social progress was the growth of the two-party system. After World War II the Democratic Party lost its monopoly on the state's political process, and the Republican Party now challenges the once-dominant party at every level. This development has shifted the focus of the political debate in Mississippi from a defense of old traditions to a discussion of new alternatives. Although the state's limited natural resources and its long years of agricultural dependency and racial discrimination have left their mark, Mississippi is now largely free of the attitudes and attributes that kept it for so long in the 19th century.

(J.N.B./D.G.S.)

North Carolina

Twelfth of the 13 original states of the United States, North Carolina lies on the Atlantic coast midway between New York and Florida. It is the leading industrial state of the Southern Atlantic states. Approximately one-half of the state's inhabitants live outside urban communities, giving it one of the largest rural populations in the nation.

North Carolina's beginnings are tied closely to the earliest attempts at English colonization of the New World. Roanoke Island in the northeast, a part of the heavily indented and island-fringed coast, was the site of the famous Lost Colony that vanished sometime after the original landing in 1587. This eastern region retains much of the flavour of colonial life, while the higher Piedmont region centred around the capital, Raleigh, has become the state's hub of industry and population. The mountains of the west remain the focus of a lively folk culture and the home of a group of North American Indians.

Bounded on the north by Virginia, on the east by the Atlantic Ocean, on the south by South Carolina and Georgia, and on the west by Tennessee, North Carolina has an area of 52,669 square miles (136,413 square kilometres). Its 3,826 square miles of inland water, the fifth largest such area of any state, are concentrated both in the extensive marshlands of the coastal tidewater and in the lakes and reservoirs of the Piedmont and Appalachian regions. These three physical regions are related to major diversities in life-styles among the people of the state, creating three distinct cultures within the state's boundaries.

PHYSICAL AND HUMAN GEOGRAPHY

The land. North Carolina extends across three major physiographic regions of the United States—the Coastal Plain (or Tidewater), the Piedmont, and the Appalachian Mountains. In addition to giving the state a spectacular landscape, this variation has influenced the character of its climate, soils, plant life, and human geography.

Three major regions

Relief. As the land reaches westward from sea level, it rises gradually to the fall line, a zone some 30 miles (48 kilometres) in width that separates the Coastal Plain from the Piedmont. In the latter, the topography becomes irregular and rises about five feet (1.5 metres) a mile to the base of the Appalachians, a distance of about 140 miles. The mountains, many over 6,000 feet, have a worn, rounded appearance, reflecting a geologic origin older than the rugged peaks of the American West. Mount Mitchell, rising to 6,684 feet (2,037 metres), is the highest peak east of the Mississippi.

Soils. Soils in North Carolina are commonly grouped according to regional variations. Coastal soils are rich and humus-laden, while farther to the west the sandhills consist mostly of sand and have almost no organic materials. The Piedmont region is predominately clayey, and mountain soils are a combination of clay, sand, and silt, commonly called loam. All of North Carolina's soils are affected by excessive leaching, which causes high mineral loss, and successful agriculture depends on large additions of lime and fertilizers.

Climate. North Carolina's climate ranges from medium continental conditions in the mountain region, though summers are cooler and rainfall heavier, to the subtropical conditions of the state's southeastern corner. The growing season ranges from 275 days along the coast to 175 days in the mountains. Average annual temperatures range from 66° F (19° C) in the eastern region, to 60° F (16° C) in the central, and 55° F (13° C) in the mountains. July and August are the wettest months, and October and November are the driest. Annual rainfall varies from 46 to 54 inches (1,170 to 1,370 millimetres) on the coast, 44 to 50 inches in the Piedmont, and 40 to 80 inches in the mountains. Severe storms are rare and heavy snow infrequent. Hurricanes occasionally occur along the coast, and there have been tornadoes inland.

Plant and animal life. Vegetation varies greatly throughout the state primarily because of the geographic and climatic differences of the three main regions. However, changes effected by human habitation are perhaps becoming equally significant. For example, trees that once covered the landscape as dense forests have been cut and burned and now cover only slightly more than 50 percent of the state. Not only has the loss of trees affected animal life by changing important habitats, but it has also contributed to soil erosion and leaching.

A greater variety of plant life is found in North Carolina than in any other state in eastern North America. There are many species of hardwood trees. Red spruce and bal-

sam fir are found in the mountains, and the subtropical palmetto and the carnivorous Venus's-flytrap grow in the southern coastal area.

The common fauna of North America, including rabbits, squirrels, raccoons, opossums, deer, and also bears and wildcats, are found within the state. The commonest birds are the cardinal, wren, mockingbird, chickadee, and many varieties of woodpecker and warbler. Inland-water fish such as bluegills, crappies, bass, and sunfish are common. Brook and rainbow trout are found in the mountains.

Coastal Plain

Settlement patterns. Comprising some 45 percent of the state, the Coastal Plain consists of a gently rolling, well-drained interior and a swampy tidewater area close to the coastline. The latter region was the first to be explored and settled. A long chain of islands, the Outer Banks, extends from Virginia to South Carolina, generally covered with sand dunes from a few feet to more than 100 feet in height. Three capes—Cape Hatteras, Cape Lookout, and Cape Fear—jut into the ocean in an area known as the Graveyard of the Atlantic, a reference to the many ships that have gone down in the dangerous waters. The entire area averages less than 20 feet above sea level. Only small-craft navigation is possible because of silting and shallow sounds and estuaries. The Intracoastal Waterway threads its way between the Outer Banks and the mainland on its way from New Jersey to the Gulf of Mexico. The inner Coastal Plain extends from 120 to 140 miles westward to the Piedmont.

Eastern North Carolina has been the citadel of the state's history since Raleigh's dream of colonization came to so mysterious an end. Close to Roanoke Island are the sand dunes of Kitty Hawk, where in 1903 Wilbur and Orville Wright ushered in the age of powered flight. Legends tell of pirate treasure buried beneath the dunes of the Outer Banks. Rusting smokestacks, masts, and boilers protrude from offshore waters, testimony to the more than 2,000 ships that have gone down. Nearby Nags Head got its name, according to tradition, because unscrupulous settlers tied lanterns to their horses' necks and drove them along the coast to lure unsuspecting seamen to the reefs. On Ocracoke Island visitors are astonished at the Elizabethan-sounding speech of the residents, for whom "high tide" is "hoigh toide."

In New Bern, the state's second oldest town, named by its Swiss settlers, is Tryon Palace, a restored palace and garden that has been called the most beautiful building in the colonial Americas. Along the southern coast, fishermen set out to battle large deepwater fish of the Gulf Stream, and in Edenton memories survive of the colonial ladies who held one of the first tea parties to protest duties imposed by the British. Morehead City and Wilmington are the state's two deepwater ports, both significant in world trade, while major military installations in the area add to the state's economic life.

Piedmont Plateau

The North Carolina Piedmont is a region of rolling, forested hills. The prominent ridges and hills of the eastern Piedmont may be the remains of an ancient mountain chain that paralleled the Appalachians, from which spurs extend into the western Piedmont. The area is well drained by rivers flowing into the Coastal Plain or South Carolina. Dams on the Catawba and Yadkin rivers are important sources of hydroelectric power.

This region is a prime symbol of the New South, in which modern industry has largely replaced the traditional agriculture. A concentration of industry occurs in a sweeping crescent westward and southward from Raleigh to below Charlotte, the state's largest city. Such cities as Durham, Greensboro, and Winston-Salem have made North Carolina the capital of the nation's tobacco industry and significant in textiles and furniture. The colleges and universities that have been so influential in the state's history are centred in this region.

In spite of industry the many antebellum homes in these cities maintain an aura of serenity, and farmlands are still found close to the city limits. The lakes and the upper reaches of the rivers provide havens for fishing and camping, and in many small towns general stores still serve the rural populations. Under the streets of Charlotte—described by Lord Cornwallis, the English general of Revolutionary fame, as "a piddlin little place"—are traces of early mines, which once produced many tons of gold.

Flue-cured tobacco ready for auction in Oxford, N.C.
Larry Lefever from Grant Heilman

Mountain region

The mountain region comprises a highly desiccated intermontane plateau bounded by two ranges of the southern Appalachians. On the east are the Blue Ridge Mountains, which rise steeply from the Piedmont to peaks of 3,000 to 4,000 feet, with several to 6,000 feet or more. In the far west the Unaka Mountains contain the Great Smoky Mountains that roll westward into Tennessee. This region is divided into several cross ridges and a number of smaller plateaus and basins. One of the chief ridges is made up of the Black Mountain group. A total of 43 peaks rise above 6,000 feet and 82 above 5,000 feet in western North Carolina.

In North Carolina's mountains, ways of life change slowly. Many communities, relatively isolated since the early history of the state, remained self-sufficient until recent times. Wood carving, basketry, needlework, rug and quilt making, and ceramics are among the many cottage industries whose crafts have been passed down through the generations. The isolation has been broken, however, both in the mountains and in the resort centre of Asheville. Winter and summer sports have become popular on the slopes, and the Pisgah National Forest and Great Smoky Mountains National Park are among the areas that attract a growing number of tourists and campers.

The people. Archaeologists have found traces of human habitation in the state that date back some 16,000 years. It is estimated that when the first European explorers arrived there were between 35,000 and 50,000 Indians in the region. In the late 1830s most of the largest remaining group of Indians, the Cherokee, were forcibly removed to lands west of the Mississippi, an exodus recorded in history as the Trail of Tears (1838–39). In the late 20th century about 65,000 Indians lived in the state, making them the largest group east of the Mississippi and the fifth largest in the nation.

Backgrounds of early settlers

Permanent white settlers came into North Carolina in the 1650s from the English colony at Jamestown, Va. Others came down on the great wagon road from Pennsylvania through the Shenandoah valley into the Piedmont. Some came by ship from Europe, all yearning for a plot of land and for freedom from rigid class and religious restrictions. The early North Carolinians were a heterogeneous group, representing a variety of religious faiths, nationalities, and economic and social classes. The Anglican church was

established by law in the early 18th century, but there were also Presbyterians, Quakers, Moravians, Lutherans, Reformed, Baptists, Methodists, and a small number of Jews. Nationalities represented included English, Scottish, Irish, Welsh, Swiss, French, and German.

Blacks were an important part of the early North Carolina population; the labour-intensive crops of rice, indigo, tobacco, and cotton accounted for the spread of slavery in the state, especially after the perfection of the cotton gin. Today blacks account for approximately one-quarter of the population. Despite continuing disparities between the living conditions of whites and blacks, blacks have made impressive gains during the 20th century in education, the arts, sports, business, and politics.

North Carolina's rural and urban populations are approximately equal, despite the large industrial employment. Many industrial plants are located in small towns, and workers tend to commute long distances and live in rural areas. Rapid urbanization and the persistence of extremely rural areas accentuate the demographic contrasts in the state. Traditional patterns of subsistence farming and small farms are giving way to consolidated farms; there also has been a marked decrease in the number of tenants and sharecroppers.

The economy. North Carolina's economy depends largely on industry and agriculture, but tourism is gaining in importance.

Industry and agriculture. Through the first half of the 20th century nearly half of the state's nonfarm work force was employed in manufacturing, primarily of textiles, furniture, and cigarettes. Today manufacturing makes up about one-third of nonfarm jobs, and the industrial base is more diversified. Strong growth has occurred in computers, electronic communications equipment, chemicals, machinery, and a host of other industries. Research Triangle Park, located near Durham, Raleigh, and Chapel Hill, and University Research Park, in Charlotte, have become major focal points for industrial research and development. Economic growth has most dramatically affected the larger cities, especially Charlotte and Raleigh. By the end of the 1980s Charlotte had become one of the largest banking centres in the nation.

Agriculture remains an important industry in the state, although the number of people it employs continues to decline. North Carolina leads the nation in the production of tobacco, sweet potatoes, and turkeys. Other principal agricultural products include peanuts (groundnuts), corn (maize), and eggs. Farm income tends to be greatest in the central and southern counties of the Coastal Plain. Forest products are used for furniture and as a source of pulp for paper. An active reforestation program has resulted in a growth of forest reserves.

Tourism has a diversified base, including the attractions of both ocean and mountains as well as the memorials to the state's past.

Transportation. Geographically, the state is one day's trucking time to New York City or to the rapidly expanding Florida market. Within the state, the highway system accounts for more than 80 percent of freight transportation; most of the remainder is conducted by rail. Raleigh–Durham and Charlotte airports serve as regional hubs for national airlines, thereby providing direct service to many U.S. cities as well as to some international destinations. North Carolina has two Atlantic gateways to world markets. Modern ports are found at Wilmington and Morehead City, both of which are equipped to handle any type of cargo.

Marine traffic

Administration and social conditions. *Government.* The structure of the government of North Carolina is based on constitutions of 1776, 1868, and 1971. Administration of the state is supervised by elected executives, including the governor and lieutenant governor (each limited to two four-year terms, not necessarily consecutive), and by the heads of state agencies, some of whom are elected for four-year terms and some of whom are appointed. The governor has great appointive powers but, as of the late 1980s, no veto over legislation—the only governor in the nation lacking this power. The General Assembly consists of the 50-member Senate and the 120-member House of Representatives. Both senators and representatives are elected for two-year terms.

The state is divided into more than 30 judicial districts. District courts deal primarily with less serious civil and criminal matters. Each district elects, for four-year terms, its district court judges and a district attorney who represents the state in all criminal matters. The superior courts handle the more serious criminal and civil cases. Superior court judges are elected for each district in statewide elections for terms of eight years. Above the superior courts are the Court of Appeals and the Supreme Court. The latter is the highest state court; it has seven justices elected for eight-year terms. The Court of Appeals was established by a constitutional amendment in 1965 to help relieve the state's Supreme Court. It has 12 judges, all elected for eight-year terms.

North Carolina is divided into 100 counties. County governments act for the state in providing education, health, and welfare services. Locally elected officials include county commissioners, the sheriff, the register of deeds, the clerk of the superior court, and the school board. Compared to those of many other states, North Carolina's government is uncomplicated. In general, counties provide services that apply to all citizens of the state, while municipalities provide the additional services appropriate for urban areas. As urban development has continued, counties have been authorized to offer services that are similar to those provided by municipalities, such as water supply and garbage collection. Because North Carolina's constitution discourages the incorporation of municipalities near existing ones, North Carolina is relatively free from the proliferation of municipal governments in urban areas that is found in many other states.

Throughout most of the 20th century, North Carolina has been dominated by a single party—the Democrats. The vast majority of state and local officeholders are Democrats, but candidates from the rival Republican Party have made major gains. Since 1970 the state has usually voted Republican in presidential elections, has elected Republicans to the U.S. Senate and House of Representatives, and has twice elected a Republican governor.

Education. The public school system, supported by the state since 1933, has improved steadily, though it is still below national levels. Other problems include a relatively low salary scale for teachers and an expenditure per pupil that is below the national average.

In higher education, however, North Carolina has a number of institutions of national standing. The University of North Carolina opened its doors to students at Chapel Hill in 1795 as the first state university in the United States. Since 1972 all 16 senior public institutions have been part of the University of North Carolina, and all are governed by a single board elected by the General Assembly. In addition to Chapel Hill, its campuses include North Carolina State University at Raleigh and the North Carolina School of the Arts at Winston–Salem, the first state-supported residential school for the performing arts. The state's community college system, which comprises some 58 institutions, is one of the largest systems in the United States. Among the dozens of private institutions around the state, most of them established by various Protestant denominations, Duke University in Durham is noted for its undergraduate and graduate programs.

Institutions of higher education

Health and welfare. State-funded hospitals cover a number of specialized areas such as children's orthopedics, alcoholism, retardation and mental illness, cerebral palsy, and tuberculosis. An effective public health program has been in operation since 1877, and each county has a local health department. State aid is provided also to the aged or disabled, to families with dependent children, and to various counseling and other social service programs. The state's social expenditures, however, remain far below the national average.

Cultural life. An arts council, established in 1964 and now a part of the state's Department of Cultural Resources, assists in bringing the highest obtainable quality in the arts to the greatest number of people in the state and also in expanding the role of the arts. The council makes grants of public funds to sponsor numerous proj-

ects. North Carolina was the first state in the country to set aside public funds for the purchase of an art collection. Housed at the North Carolina Museum of Art in Raleigh, the collection spans some 5,000 years, from the art of ancient Egypt to contemporary works.

The North Carolina Symphony has the distinction of being the first state symphony in the country. The orchestra tours the state from September through May. Many of the performances are free matinees for children.

It is in the field of the folk arts and of historical pageantry that North Carolina excels. The many cottage industries of the western mountains combine with those of the coastal communities to offer some of the richest folk culture in the United States. Outdoor epic dramas are held all summer long in Manteo on Roanoke Island, where Paul Green's drama *The Lost Colony* revives the colonizing efforts of Sir Walter Raleigh in the court of Elizabeth I and on the soil of Roanoke itself; in Boone, where *The Horn in the West* recreates such characters as Daniel Boone; and in Cherokee, where *Unto These Hills* is played by the descendants of the Cherokee Indians upon whose history the saga is based.

Historical dramas

North Carolina shares with Tennessee the Great Smoky Mountains National Park, which contains approximately 461,000 acres of mountain forestland and includes museums, nature trails, and campgrounds. The Appalachian National Scenic Trail passes through the park, and the Blue Ridge National Parkway begins in the park and extends to the Shenandoah National Park in Virginia. National seashores are located at Cape Hatteras and Cape Lookout. Other National Park Service sites mark the first English settlement on Roanoke Island (Fort Raleigh National Historic Site), the Wright brothers' first flight at Kill Devil Hills (Wright Brothers' National Memorial), the home of Carl Sandburg in Henderson county (Carl Sandburg Home National Historic Site), and the battles at Guilford Court House and Moores Creek.

HISTORY

The proprietary and royal colony. Following the attempts by Raleigh and others to colonize the coastal regions in the 1580s under patents from Queen Elizabeth I, the region remained Indian territory for decades. A grant by King Charles I in 1629 for the lands south of Virginia brought the term Carolina into being, but no permanent settlement was made until farmers and traders from Virginia moved into the Albemarle Sound area in the 1650s. This resulted in a grant from Charles II in 1663 that created Carolina, but for years the settlers resisted the ineffective government imposed by the proprietors in England. Between 1712 and 1729 the separate province of North Carolina was ruled by a deputy dispatched from Charleston, which had become the centre of proprietary government. Boundaries between North and South Carolina were agreed upon in 1735 but not completely surveyed until 1821.

North Carolina's growth was hampered by restrictions on shipping imposed by Virginia on its tobacco crop, by economic and religious quarrels with absentee proprietors that led to several uprisings, by war with the Tuscarora Indians (1711–13), and by coastal piracy involving Edward Teach (Blackbeard) and others. Unlike other colonies, which had grown up around coastal towns that represented the first settlements, North Carolina had no town until Bath was incorporated in 1705. By 1729, when the colony came under royal rule, several other towns had been chartered.

Effects of royal rule

The decades of royal rule saw a turnabout in the colony's fortunes. The population rose rapidly, settlement spread across the Piedmont, and the wealth and quality of life expanded. A large slave population maintained an agricultural economy based on tobacco and rice and on naval stores from the region's extensive pine forests. Prior to the U.S. War of Independence, the beginnings of an intense east–west hostility had grown into several insurrections, but joint antipathy to British rule united North Carolinians and forced the flight of the royal governor in 1775.

Statehood. The war in North Carolina comprised not only a miniature civil war involving the many Tories in the new state but also the suppression of Cherokee uprisings in the west. Much of the state's energy went to resolving the conflicting interests of the eastern counties and those of the west until constitutional reforms in 1835 broke the dominance of the east. A period of great economic and social progress, first under the Whigs and after 1850 under the Democrats, was slowed by the furor over slavery and was ended by the American Civil War.

Civil War and after. Unlike South Carolina, whose strident proslavery voices led the South into secession, North Carolina left the Union reluctantly, seeking compromise until the last moment. Once committed, however, the state fought with the Confederacy and experienced the ignominy of defeat and the years of corruption and instability that characterized the postwar Reconstruction throughout the South. The "Bourbon Democrats" who controlled the state after readmission to the Union in 1868 were oriented to the railroad and industrial interests, ignoring the small farmer. Constitutional amendments in 1900 virtually disfranchised the state's black population.

The 20th century. North Carolina in the 20th century has been a part of the national experience of changing economic cycles. A decade of significant economic and social developments followed World War I, but the Great Depression of the 1930s brought widespread hardship and severe curtailment of education and other public services. In the view of some historians, the federal New Deal programs were responsible for more lasting changes in the state than any other force in its history.

In the 1940s the national defense program and World War II affected North Carolina. Some of the country's largest military installations were located in the state, notably Fort Bragg. North Carolina was a major supplier of manufactured war materials and delivered more textile goods to the Army than did any other state.

Impact of World War II

After World War II the state began a time of rapid change. New highways were built, and cities grew as new industry and new people moved to the state. Interest in politics revived, and by the 1970s the state again had a viable two-party system. The painful struggle to eliminate racial segregation, beginning in the public schools in the 1950s and at the lunch counters in Greensboro in 1961, absorbed the state's energies for several decades. While most racial segregation had ended by the late 1980s, the state continued to be burdened by the remnants of earlier discriminatory practices and prejudiced attitudes.

North Carolina faces the enormous challenges of extending the benefits of education and economic prosperity to all its citizens, of preserving its environment from the pressures of growth, and of eliminating remaining racial discrimination. The state's economic prosperity, its widely respected system of higher education, and a growing confidence in its role as a leader of the New South give it the resources to face those challenges with optimism.

(Pe.S.G./D.G.M.)

South Carolina

Settled by the English in 1670, South Carolina was one of the 13 original colonies. Its wealthy, aristocratic, and influential colonial society was based on a plantation agriculture that relied on a labour force of black slaves. By 1730, blacks had come to represent two-thirds of the colony's total population. The plantation system spread into the Piedmont in the early 19th century, and the new state became part of the Cotton Belt that stretched across the South. The Civil War shattered South Carolina's economy and influence, and for a century thereafter the state suffered economic, social, and political turmoil. The 1960s brought a major change as South Carolina's economy industrialized, its metropolitan areas grew, and the Civil Rights movement swept across the state.

Location

South Carolina lies on the southern Eastern Seaboard of the United States. Shaped like an inverted triangle with an east–west base of 285 miles (459 kilometres) and a north–south extent of about 225 miles, the state is bounded on the north by North Carolina, on the southeast by the Atlantic Ocean, and on the southwest by Georgia. It is ranked 40th among the 50 states in size and has a geographic area of 31,113 square miles (80,583 square

kilometres). Columbia, located in the centre of the state, is the capital and largest city.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* South Carolina can be divided geologically into three separate provinces. A portion of the Blue Ridge Mountains dominates the northwestern corner of the state, covering about 2 percent of the state's area; the highest point in South Carolina, Mount Sassafras, 3,554 feet (1,083 metres), lies on a crest in this region. The Piedmont, with an elevation ranging from about 300 to 1,200 feet, stretches over about a third of the state from the mountains to the midlands around Columbia. The Coastal Plain comprises the southern and eastern thirds of the state, and its elevation varies from about 300 feet to sea level; the 187-mile coastline has sandy beaches and salt marshes. Two major boundaries separate these provinces: the fall line, where rivers form major rapids, divides the sedimentary rocks of the Coastal Plain from the metamorphic rocks of the Piedmont; and the Brevard zone is a fault line separating the Blue Ridge and the Piedmont.

Blue Ridge Mountains

The rugged Blue Ridge Mountains formed some 250,000,000 years ago from the collision of the early North American continent with Africa. These mountains and their outliers, such as Table Rock and Pinnacle Mountain, rise 1,000 to 2,000 feet above the Piedmont and constitute an area devoted to wilderness preservation and recreation. Their forests include many northern species, such as white pine and hemlock, relicts of the Pleistocene epoch. The poorly developed Blue Ridge soils lack clay accumulation beneath the surface and are not well suited for farming.

The Piedmont has a rolling relief but is much worn down. The reddish Piedmont soils are clayey, with noticeable deposits of calcium, magnesium, and potassium. Planted to grow cotton and corn (maize) for 150 years, the region suffered severe soil erosion. Mid-20th-century abandonment of row crops has allowed the return of forests but not the oak and hickory that typified the 18th century. The most common tree in the Piedmont today is the loblolly pine, vast acreages of which have been planted. At the edge of the Piedmont lie the Sandhills, which run diagonally across the state from Marlboro county to Aiken county. These sandy ridges were formed when the Atlantic Ocean covered the present-day Coastal Plain about 55,000,000 years ago.

The Coastal Plain is slightly rolling near the midlands and flat toward the coast. The state's agricultural belt dominates the inner Coastal Plain, while the outer Coastal Plain is largely forested. South Carolina's coast was abandoned as an agricultural area around the turn of the 20th century, but good management has allowed truck farming to prosper. Since the mid-20th century it has been the focus of major tourism and recreation development and of intense land-use competition. The Grand Strand is an unbroken beach that extends from the North Carolina line almost to Winyah Bay, but southward the coast is fringed by the Sea Islands that extend into Georgia. The islands comprise much tidal and freshwater marsh. Huge pines, gums, live oaks, cypresses, and magnolias are draped with Spanish moss. Many blacks on these islands still speak Gullah, a patois that derives from English and several West African languages and dates to the plantation era.

The Sea Islands

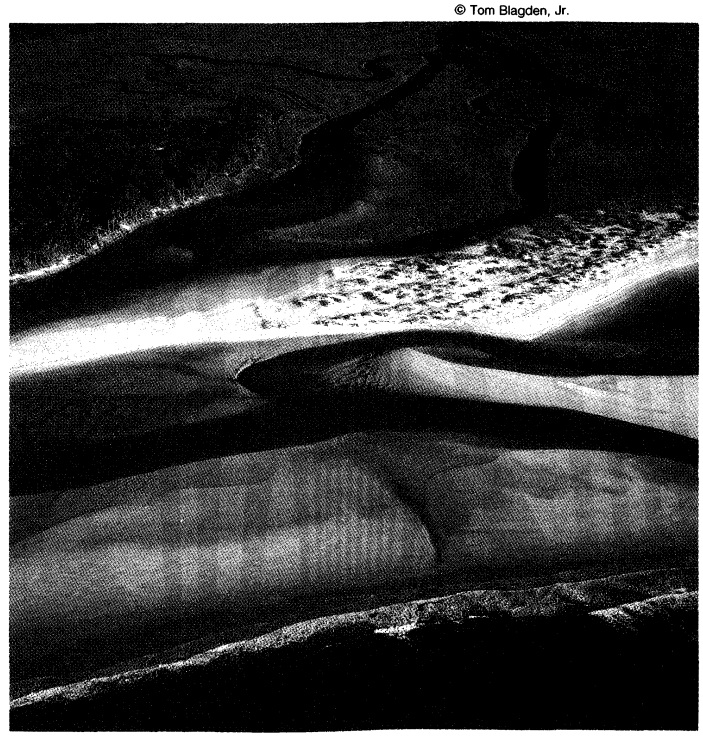
Drainage. South Carolina's rivers flow generally from northwest to southeast. Three major systems, the Pee Dee in the east, the Santee, whose tributaries drain much of the Piedmont, and the Savannah on the western boundary, cover about 80 percent of the state's area. The Ashley–Combahee–Edisto system comprises the short rivers that form near the Sandhills and flow across the Coastal Plain. Carrying little sediment, their waters are blackened by tannic acid from the swamps along their courses. South Carolina has no large natural lakes; those on the Savannah River and Santee tributaries resulted from hydroelectric development in the 20th century. On the Coastal Plain are hundreds of elliptically shaped depressions of varying sizes typified by swamp vegetation and standing water in the centre. The formation of these so-called Carolina bays is a mystery; some geographers have attributed them to the impact of a comet or meteor.

Climate. South Carolina's climate is subtropical, with hot, humid summers and generally mild winters. Average July temperatures range from 71° F (22° C) in the highland northwest to 81° F (27° C) in the midlands and along the coast. Average winter temperatures vary from 38° F (3° C) in the mountains to 45° F (7° C) in the midlands to 48°–50° F (9°–10° C) on the coast, which is warmed by the Gulf Stream. The growing season ranges from less than 200 days in the northwest part of the state to about 290 days on the Sea Islands. Most of the state receives about 49 inches (1,245 millimetres) of rain annually, but 70 to 80 inches are recorded in the mountains in the northwest. Summer rainfall, typified by afternoon thunderstorms, normally exceeds that of any other season. The state experiences an average of 10 tornadoes a year, usually occurring during the spring. Hurricanes are less frequent, but they do in some years cause damage to South Carolina's coast.

Animal life. The white-tailed deer is the best known of the mammals found in South Carolina. Its population in the Piedmont had declined severely by the 1940s, but restocking from the Coastal Plain and restoration of forests have now ensured its statewide distribution. Other species, such as the beaver and wild turkey, have also made comebacks. Many species, including the bison, wapiti (elk), cougar, and wolf, disappeared by the 1800s, and the black bear is rare today. The red fox and wild pig, both introduced by Europeans, are widespread. Species such as the woodchuck and red squirrel are found in the Blue Ridge, while the Coastal Plain harbours alligators and varieties of all four poisonous snake groups found in the United States. South Carolina has recorded some 360 species of birds. Most faunal species occur statewide, but some are limited to either the Piedmont or the Coastal Plain since the Sandhills create a formidable barrier, especially to reptiles and amphibians.

Settlement patterns. The state's population distribution has changed since World War II. Although South Carolina remains more rural than the nation as a whole, its metropolitan areas have grown and account for more than half the total population. Areas of high density are found in the upper Piedmont, focusing on Greenville–Spartanburg and Anderson; the midlands around Columbia and Florence; and along the coast, including Charleston, Beaufort, and Myrtle Beach. The large metropolitan areas that

© Tom Blagden, Jr.



Part of the outer Coastal Plain, in the Cape Romain National Wildlife Refuge, near the Santee delta, South Carolina.

have developed around small cities during the automobile era now have overwhelmingly suburban populations.

The people. Native American Indians inhabited what is now South Carolina for thousands of years before white settlement. Though their population declined rapidly after European contact, a few thousand still reside in the state. The largest group are the Catawba, who have a small reservation in Lancaster county.

South Carolina's colonial population was a mixture of European nationalities. Settled initially from England and Barbados, the colony by the 1680s was receiving Scots and a number of Huguenots escaping France after the revocation of the Edict of Nantes. Some of the French later settled in the western part of the state near Abbeville.

The ethnic mix

Germans moved into the midlands in the 1730s, locating along the Saluda River, and Welsh Baptists settled the Welsh Tract on the Pee Dee River. In the 1750s and '60s, Scotch-Irish settlers from Pennsylvania and Virginia spread into the South Carolina Piedmont. In the 19th century, however, few of the millions of European immigrants that flooded the United States settled in South Carolina. Today, nearly a quarter of the state's small number of foreign-born residents are recent arrivals from Asia.

Five blacks were among the 148 original colonists in 1670. In 1880 the black population represented almost 61 percent of the state's total population, but heavy out-migration to largely urbanized Northern states during the 20th century saw this decline to about 31 percent in 1970 before stabilizing.

Throughout the first half of the 20th century, South Carolina's population grew much more slowly than that of the nation as a result of out-migration of both blacks and whites. By the 1970s this began to change, and South Carolina's growth rate has since exceeded the national average. Almost three-quarters of South Carolina's more than 3,000,000 residents were born in the state, but metropolitan areas, especially in the midlands and along the coast, show smaller percentages of South Carolina natives, while rural counties show much higher percentages.

The economy. During the first half of the 20th century, agriculture was the key to the state's economy, but by 1923 the value of manufactured goods exceeded that of agricultural products. Agriculture continues to be important, but its role has declined as manufacturing and service employment have increased. South Carolina's industry is typified by low-wage, nondurable-goods manufacturing.

Resources. South Carolina's major resources include its climate and forests. Known for sunny summer beaches, the state seeks a year-round tourism based on golf and other recreational activities. Forests cover about 65 percent of the state and contribute sawtimber for lumber and pulpwood for paper production; tourism and forestry are the state's second and third leading industries, respectively. Mineral resources include clay, sand, gravel, and crushed stone (granite and limestone); and a 19th-century industry, gold mining, has been revived to a limited degree.

Agriculture and fishing. Since World War II the number of farms has declined nearly 85 percent, and land in farms has fallen from more than one-half to about a quarter of the state's land area. Another major change has been the virtual disappearance of cotton. Once planted almost across the state, it is limited now to only a few counties in the inner Coastal Plain. The major crop in cash receipts since the 1950s has been tobacco, grown principally in the Pee Dee region. By far the largest acreage is devoted to soybeans, a crop introduced successfully into South Carolina only in the 1940s. The state has long been famous for its peaches, grown in the upper Piedmont and in the midlands. Previously of little value, livestock and poultry have come to play an increasingly important role in the agricultural economy, especially in the Piedmont. The important coastal commercial seafood industry includes shrimp, crabs, and oysters.

Decline of cotton

Industry. At the beginning of the 20th century many textile mills began operation in the upper Piedmont and midlands, and by 1910 almost 150 mills employed 45,000 workers. Although today suffering severe competition and a steady decline, textiles remain the state's leading industry, accounting for more than a fourth of the manufacturing employment. Other major industries include wearing apparel, paper, machinery, and chemicals. This industrialization has resulted from the opening of many branch plants by companies from the northern United States. On a per capita basis the state also is one of the leading recipients of foreign capital investment.

Transportation. South Carolina is crisscrossed by interstate highways that link it with every part of the country, but railway mileage has declined. The two major railroads continue to abandon branch lines serving smaller towns, although a few of these are now operated as independent short lines. Major air carriers serve the metropolitan centres of Greenville–Spartanburg, Columbia, and Charleston, as well as Myrtle Beach, while commuter airlines connect smaller cities with regional hubs. The State Ports Authority has developed Charleston into the major container port on the South Atlantic coast and also operates port facilities in Georgetown and Port Royal.

Administration and social conditions. *Government.* State government operates under the 1895 constitution, which has, however, since 1968 been rewritten article by article by the General Assembly, each amendment being submitted for voter approval. The governor, the state's chief executive, is elected to a four-year term, and an amendment in 1980 allowed reelection to a second consecutive term. The lieutenant governor serves as ex officio president of the Senate and succeeds the governor in the event of an uncompleted term. Other constitutional officers, all elected for four years with no limitation on consecutive terms, are secretary of state, treasurer, attorney general, comptroller general, adjutant general, superintendent of education, and commissioner of agriculture. The governor, although able to veto bills and specific items in the budget, has much less authority over state government than does the legislative branch. The General Assembly comprises two houses: the Senate and the House of Representatives. Originally, each county had one senator and at least one representative, but the U.S. Supreme Court declared that apportionment unconstitutional. In 1974 the House was divided into 124 single-member districts of equal population, and in 1984 the Senate was divided into 46 districts. Senators serve four-year terms, and representatives serve two-year terms.

Revised constitution

Judicial authority is vested in the Supreme Court, and all courts are unified under the administration of the chief justice. The Supreme Court comprises the chief justice and four associate justices. All are elected by the General Assembly for 10-year terms, which are staggered so that one justice is elected every two years. The Court of Appeals, created in 1983, has a chief judge and five associate judges that sit in three-judge panels. It automatically hears all appeals from lower courts, except for cases involving death penalties, elections, constitutionality of laws, actions of state regulatory agencies, and bonded indebtedness, which go directly to the Supreme Court. The state is divided into 16 judicial circuits served by circuit courts; these courts have the widest jurisdiction and hear civil (Court of Common Pleas) and criminal cases (Court of General Sessions).

The 1895 constitution had no provision for local government, and the county legislative delegation, with approval of the General Assembly, decided most local matters. The Local Government Act of 1975 transferred some authority to local administration. The act identified the forms of government that could be adopted by the 46 counties and almost 300 municipalities and defined their responsibilities, powers, and taxing authority.

For more than a century, South Carolina politics was dominated by an all-white Democratic Party. The 1960s brought two developments: the rise of the Republican Party, whose strength lay in the suburban metropolitan counties, and the Voting Rights Act, which allowed black participation in the electoral process and paved the way for election of blacks to local offices and the General Assembly. Although Democrats continue to hold sway at the state level, Republicans have made inroads in congressional representation, the governor's office, the General Assembly, and county councils in metropolitan areas.

Education. Public education was guaranteed under the

1868 Reconstruction constitution, which created the Superintendent of Public Instruction. The State Board of Education today certifies teachers, sets standards, and establishes courses of study for the public schools, which are grouped into approximately 90 local districts. Educational attainments have improved at all levels, but South Carolina has long scored near the bottom in such rankings as percentage of high school graduates and public school teacher salaries. To improve the situation, the General Assembly has equalized funding for education across school districts and established minimum standards of educational achievement. The Education Improvement Act, passed in 1984, increased state funding for education to improve instruction and raise student performance.

Colleges and universities

The largest institution of higher education in the state is the University of South Carolina in Columbia, chartered as South Carolina College in 1801 and opened in 1805. Clemson University, a land-grant institution established in 1889, has a major research and teaching focus on agriculture. Winthrop College (1886), located at Rock Hill and long known as South Carolina College for Women, is now coeducational. The Citadel, one of the few state-supported military colleges, is located in Charleston, as is the Medical University of South Carolina. South Carolina State College in Orangeburg was chartered originally as a black college in 1896; Francis Marion College in Florence was established in 1970; and the College of Charleston, the oldest publicly supported institution in the United States, founded in 1770 and chartered in 1785, officially became a part of the state college system in 1970. Throughout South Carolina are many private institutions supported by Protestant denominations.

Health and welfare. County health departments provide basic health services from prenatal care to immunization, but South Carolina still faces some severe problems. It has one of the highest infant mortality rates and lowest life expectancies of the 50 states and ranks low in the number of physicians per 100,000 population. The metropolitan areas have excellent hospital facilities and some of the most advanced medical technologies, enhanced in Charleston and Columbia by the two medical colleges. The rural counties, however, struggle to maintain local hospitals.

Underlying the health problem is poverty, and South Carolina, despite great economic strides, remains among the states with the highest percentage of persons below the poverty level. Within the state, this has a specific geographic pattern: more than a quarter of the residents of the rural counties across the Coastal Plain are identified as below the poverty line.

Cultural life. Charleston was the cultural centre of the South during the colonial period and early 19th century. The Charleston Library Society, founded in 1748, provided the literary focus; the St. Cecilia Society, formed in 1762, regularly held public concerts; the Dock Street Theater, opened in 1736, was one of two theatres in Charleston and perhaps the site of the first play produced in the United States; and scientific interests underlay the foundation of the Charleston Museum in 1773. All of these institutions continue today.

Literature

William Gilmore Simms was the most successful and prolific writer of the antebellum South. Julia Peterkin (1880–1961), one of the first to describe the plantation from the perspective of blacks, won the 1928 Pulitzer Prize for Literature. Archibald Rutledge (1883–1973) served for 39 years as the state's first poet laureate. DuBose Heyward achieved his greatest success in 1925 with the novel *Porgy,* which provided the basis for George Gershwin's opera *Porgy and Bess.* The humorous short stories about local life by William Price Fox have received critical acclaim, and James Dickey, a longtime resident of South Carolina, is among the most widely read American poets.

Portraiture was the dominant form of art in colonial and early 19th-century South Carolina and was exemplified by the work of Henrietta Johnston (d. *c.* 1729) and Charles Fraser (1782–1860), a well-known miniaturist. Washington Allston is considered the first important American painter of the Romantic movement. The black artist William Henry Johnson (1901–70) has received notice for his paintings in the so-called primitive style. The sculptural work of Anna Vaughn Hyatt Huntington (1876–1973) forms the core of the internationally known Brookgreen Gardens near Georgetown. The paintings of Jasper Johns, who was born and raised in South Carolina, have been exhibited at major museums throughout the world.

Among South Carolina's notable art collections are those of the Gibbes Art Gallery in Charleston, emphasizing 18th- and 19th-century portraiture, and the South Carolina State Museum in Columbia, which houses the State Arts Commission's collection of contemporary South Carolina artists. The McKissick Museum of the University of South Carolina has developed collections and exhibits of indigenous folk art, including Edgefield and Catawba pottery and Afro-American basketry.

The Spoleto Festival

The Spoleto Festival in Charleston was founded in 1977 by the Italian opera composer Gian Carlo Menotti as the New World branch of his Festival of Two Worlds in Spoleto, Italy. The annual event features hundreds of actors, singers, dancers, musicians, and other artists in more than 100 performances. Of a more recreational character are the many harvest-based and local festivals held by small towns across the state.

Charleston is noted for its splendid, well-preserved 18th- and 19th-century houses and public buildings. Beaufort and Georgetown also have well-maintained historic districts, and many other South Carolina cities have preserved buildings, restored the architectural integrity of downtowns, and designated historic areas.

HISTORY

Earliest settlement. The first inhabitants of present-day South Carolina arrived about 11,000–12,000 years ago. Hunting and gathering typified their first 10 millennia, but they developed agriculture about 1000 BC. The Mississippian peoples, the most advanced in the pre-Columbian Southeast, arrived about AD 1100 with their complex society, villages, and earthen mound-building, but they disappeared soon after European contact. Perhaps 15,000–20,000 Indians resided in South Carolina in 1600, representing three major language groupings: Siouan (Catawba), Iroquoian (Cherokee), and Muskogean (related to Creek). Disease, conflict, and continued European expansion contributed to the virtual disappearance of the Indian population by the time of the U.S. War of Independence.

Colonization. The first Europeans to visit South Carolina, in 1521, were a party of Spaniards from Santo Domingo (Hispaniola). In 1526 Lucas Vásquez de Ayllón led a colony of Spaniards in the earliest settlement of what is thought to have been South Carolina, but it failed within a few months. French Protestants under Jean Ribaut made an unsuccessful attempt to occupy the area of Port Royal in 1562. A few years later, in 1566, the Spaniards returned and established Santa Elena on Parris Island. It was an important Spanish base until 1587.

Colonial charter

In 1665 Edward Hyde, 1st Earl of Clarendon, and seven other lords proprietor received a charter from King Charles II to establish the colony of Carolina (named for the king) in a vast territory between latitudes 29° and 36°30′ N and from sea to sea. Under it, the English made the first permanent settlement on the west bank of the Ashley River at Albemarle Point in 1670. A decade later, the government and most inhabitants moved to a more favourable location on the peninsula formed by the Ashley and Cooper rivers, the site of Charleston today. The colony grew slowly and by 1720 had a population of about 19,000, settled almost exclusively along the coast. Trade with the Indians and export of deerskins constituted the major sources of income, complemented by naval stores after 1710. Conflicts with the lords proprietor over economic support, Indian trade, and authority of the Commons House resulted in the overthrow of proprietary rule in 1719.

In 1729 the colony was divided into two provinces, North and South; Georgia was carved out of the southern part of the original grant in 1731. Under crown rule, South Carolina prospered, and exports of rice and indigo contributed to its growing wealth. Based on this successful trade, Charleston entered its golden age, and its much-envied refinement and cultural attainment made it a leading city in the colonies. The flood of Scotch-

Irish settlers overland from Pennsylvania saw a population explosion in the interior after 1760 and demands for political representation, resulting in a conflict between the Low Country (coast) and Up Country that continued into the 19th century. British troops occupied Charleston in 1780, but much of the Revolution was fought as a civil war between patriot and loyalist South Carolinians. Two major American victories were the battles at Cowpens and Kings Mountain.

Statehood, Civil War, and aftermath. In 1788 South Carolina became the eighth state to ratify the U.S. Constitution. The relocation of the state capital in 1786 from Charleston to the newly created city of Columbia in the interior was intended to reduce the regional conflict, but the 1790 constitution continued Low Country dominance of the government. After the proliferation of the cotton gin by 1800, the cotton plantation and slavery moved into the Piedmont and created common interests between the two regions. The Up Country also benefited from internal improvements that included a canal-building program.

The Denmark Vesey slave revolt in 1822 contributed to a climate of anxiety in South Carolina over the slavery issue, and the high tariffs of 1828 precipitated talk of disunion. South Carolina proposed a convention in 1832 to nullify tariff laws, but no other Southern state supported it. Senator John C. Calhoun, the architect of nullification, was the major spokesman for the South until his death in 1850. Radicals such as Robert Barnwell Rhett led South Carolina to secede from the Union in December 1860, and the firing on Fort Sumter in April 1861 began the Civil War. Four years later, General William T. Sherman's troops had burned their way through the state and the Confederacy had surrendered. Some 60,000 South Carolinians had gone to war; nearly a quarter never returned.

Nullification and secession

Reconstruction (1867–77) was a bitter era, marked by military occupation, disfranchisement, and corruption, but the 1868 constitution did establish basic political equality and committed the state to public education. The election of Wade Hampton as governor in 1876 was facilitated by fraud and intimidation of blacks. Hampton was not allowed to take office until 1877, however, when Reconstruction ended and the era of "Bourbon rule" by the "old guard" planters and merchants began.

The conflict between Up Country and Low Country became a struggle between the poor and the propertied. The former voted Benjamin R. Tillman into the State House in 1890 to end the Bourbon era. A leader of the farmers' movement and a blatant racist, "Pitchfork Ben" held office until 1894 and served in the U.S. Senate from 1895 to 1918. The farmers'—or, more broadly, "reform"—movement was marked by the establishment of Clemson Agricultural College (later Clemson University) and Winthrop Training School for Teachers (later Winthrop College), the dispensary system of state liquor monopoly (later abandoned because of corruption), and the 1895 constitutional convention that disfranchised blacks as much as possible. "Tillmanism" remained a major political force into the 20th century.

The modern era. After World War I, cotton prices collapsed, and the boll weevil destroyed up to half the cotton crop in 1922. This disaster, which caused a wave of out-migration, was followed after 1929 by the Great Depression. Many New Deal programs benefited the state, including construction of the Santee–Cooper hydroelectric complex funded by the Works Progress Administration (WPA).

South Carolina underwent economic, demographic, social, and political revolutions after World War II. Between 1950 and 1980, nonagricultural employment grew three times faster than the overall population as the State Development Board, created in 1945, actively recruited industry. Income per capita increased dramatically and reached 77 percent of the national average by the 1980s. The economic revolution was paralleled by a demographic revolution as urban population grew 17 times faster than the rural population. After 1975 new economic opportunities reversed migration patterns as more people, both black and white, moved into South Carolina than left it. The social revolution that ended racial segregation included some tragic events, such as the Orangeburg Massacre, in which three blacks died in a confrontation with state police on the South Carolina State College campus. Moderate governors, such as Ernest F. Hollings (1959–63), Donald S. Russell (1963–65), Robert E. McNair (1965–71), and John C. West (1971–75), led South Carolina through this difficult but generally peaceful era. The political revolution involved both the rise of the Republican Party and the increased participation of blacks, who by 1970 accounted for about a quarter of the state's registered voters. Strom Thurmond, first elected to public office in 1946, successfully accommodated the changing electorate. He was the presidential nominee of the segregationist States' Rights Democratic Party ("Dixiecrats") in 1948 and became a Goldwater Republican in 1964. Thurmond served in the U.S. Senate from 1954 with wide support, becoming a nationally recognized spokesman for Southern conservatives.

Demographic revolution

In the era of the so-called New South, South Carolina's achievements have brought new challenges. Sustainable economic growth calls for further diversification of the industrial base, coupled with increased protection of the environment. Improved health services, educational programs, and employment opportunities, especially in rural counties, are needed to help distribute the benefits of economic growth to all South Carolinians. (J.J.W./D.O.B.)

Tennessee

Located in the upper South of the eastern United States, Tennessee became the 16th state of the Union in 1796. It embodies so many diverse elements of the nation that it constitutes a virtual microcosm of the country. Early pioneers in East and Middle Tennessee, far removed from established authority, set precedents of self-government that are identified with the frontier tradition. From the beginning, Tennesseans, despite loyalty to the larger society, found that they had to rely on themselves for security. This nurtured a spirit of independence, which persists today. Furthermore, the geographic diversity has created a variety of economic, social, and cultural patterns within the state. While the East is well known for its mountain tradition, the Bluegrass area of Middle Tennessee is a balanced agricultural and commercial region, and the West has an economy based largely on cotton and the Mississippi River, with closer ties to the Deep South. Even though Tennessean Andrew Jackson long symbolized the Democratic Party of the common people, the rival Whig Party carried the state for many years. Strongly divided by the Civil War and its own version of Reconstruction, Tennessee became a part of the solid Democratic South, lagging in wealth and prestige behind the rest of the nation. The dreams of the industrialists of the late 19th century were not realized until the second part of the 20th. In the 1970s, with the return of the Republicans to prominence, Tennessee became once again a two-party state. As the South has become a part of the Sunbelt, the state has shared in this emerging prosperity.

The geography of Tennessee, an area of 42,144 square miles (109,153 square kilometres) is unique. Its extreme breadth of 432 miles (695 kilometres) stretches from the Appalachian Mountain boundary with North Carolina in the east to the Mississippi River borders with Missouri and Arkansas in the west. Its narrow width, only 112 miles, separates its northern neighbours, Kentucky and Virginia, from Georgia, Alabama, and Mississippi on the south. Nashville is the capital and Memphis the largest city.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The state is divided into six natural regions. In the extreme eastern part of the state lie the Unaka Mountains—a section of which is popularly known as the Great Smoky Mountains—with 16 peaks that rise above 6,000 feet (1,830 metres); the tallest of them, Clingmans Dome, rises to 6,643 feet (2,025 metres). The Great Valley of East Tennessee, varying from 30 to 60 miles in width, includes a series of low ridges that rise above the intervening valleys. The Cumberland Plateau has a generally flat, slightly undulating surface cut by deep and sometimes wide river valleys. The Interior Low Plateau

Peaks, plateaus, and plains

in Middle Tennessee is dominated by the Nashville, or Central, Basin and the Highland Rim. About 60 miles wide, and running roughly north to south across the state, the basin floor is a slightly rolling terrain punctuated by small hills known as knobs. To the west, the eastern Gulf Coastal Plain undulates only slightly and is laced with meandering, low-banked streams. In the extreme west, the plain ends in the Mississippi alluvial plain, a narrow strip of swamp and floodplain alongside the river.

Drainage and soils. The land drains directly into three major rivers. The Tennessee River, which flows southward in the east and northward in the west, drains the east, the southern part of the middle region, and a major part of the west. The Cumberland River, dipping into the state from the north, drains the upper middle region, while the Mississippi River directly drains a small portion of the west. The damming of the Tennessee and, to a lesser extent, of the Cumberland has created an impressive chain of slack-water lakes, sometimes known as the Great Lakes of the South, many of which lie in Tennessee. The valleys and upland basins of Tennessee have moderately fertile soil of limestone origins, and the streams have created rich alluvial lands along their beds. The soils of the ridges and the plateau, however, are thin, stony, and moderately acid, while the coastal plain has a sandy, thin soil that does not support agriculture. Approximately one-third of the soils of the state are unfit for any kind of cultivation, but four-fifths of the total land area is used for crops, livestock, timber, or other agricultural products.

Types of soils

Climate. Tennessee has a moderate climate featuring cool, but not cold, winters and warm summers. The drop in elevation causes temperatures to rise significantly from east to west. The average high temperature is 85° F (29° C); the average low temperature is 30° F (−1° C). The growing season ranges from 130 days in the mountainous east to nearly 240 days at Memphis. Most of Tennessee is within the range of 160 to 220 days. The state receives ample precipitation, about 52 inches (1,321 millimetres) a year, rather evenly distributed over the seasons and regions.

Plant and animal life. Because of the state's central position in the eastern half of the United States and its diverse elevations, many plants, animals, and fish identified more with the extreme northern and southern parts of the nation are found in the state. About one-half of Tennessee is forested, and there are more than 200 species of trees, of which some 60 are commercially valuable. Such trees as locust, poplar, maple, oak, elm, beech, pine, spruce, walnut, hickory, and sycamore are found throughout the state.

© Tom Till

Forested summits of the Great Smoky Mountains, a section of the Unaka Mountains, in eastern Tennessee.

Settlement patterns. The first white settlers found their way to river valleys in upper East Tennessee and then to Middle Tennessee north of the Cumberland River. In the early part of the 19th century, pioneers moved south in these regions. West Tennessee was purchased from the Chickasaw in 1818, but the Cherokee retained a small portion of lower East Tennessee until their removal in 1838.

Unlike the Deep South, in which huge plantations were the symbol of the cotton empire, Tennessee developed largely as an agglomeration of smaller farms. Small farms and small towns still characterize much of the landscape of the state. Six of the state's major cities and their surrounding counties have attained metropolitan size, reflecting Tennessee's growth in population, industry, and urbanization at a generally more rapid pace than that of its neighbours. Knoxville, the site of the oldest and largest campus of the University of Tennessee, and Chattanooga, where in 1863 Confederate and Union soldiers clashed on Lookout Mountain, are the major centres of East Tennessee. Nashville, the state capital and the cultural centre of Middle Tennessee, is perhaps best known as the national capital of country music. Clarksville, 50 miles northwest of Nashville, is the centre for dark-fired tobacco and the home of the 101st Airborne, which is based at nearby Fort Campbell, Ky. Jackson, an old frontier town and transportation centre, is in the heart of West Tennessee, a region acquired from the Chickasaw in 1818 as the "Jackson Purchase." Memphis is the hub of West Tennessee, and its history includes its major role in the Mississippi River steamboat traffic and its prominent and colourful position in the development of jazz. Three smaller cities in northeastern Tennessee are considered together as the state's seventh metropolitan area.

The people. The original white settlers were of predominantly Scotch-Irish and English stock, although Germans were well represented. Black slaves were present from the beginning of white settlement, enduring with their masters the privations of frontier life. Initially few in number, slaves and free blacks had a measure of liberty. Antislavery societies and journals flourished in the eastern part of the state until the opening of plantations in Middle and West Tennessee brought increasing numbers of slaves to the state. During the Civil War blacks defended the Union as civilian workers and as soldiers. They served in the state legislature during Reconstruction but were given few positions of leadership by the Republicans. Losing political power after Reconstruction, blacks played only a limited role in politics until the 1960s.

The Cherokee heritage

The Cherokee, the only Indian people who actually lived in Tennessee during most of the 18th century, were originally allies of the English traders, but they eventually began to resist the flood of settlers that poured into the area. Ceding bits and pieces of land in a series of treaties, the Cherokee adopted ways of the whites in the hope that they could live in peace. One leader, Sequoyah, created an alphabet for his people, enabling them to write and read their language for the first time. They became farmers, and, in some cases, slave owners, prospering in ways that aroused the envy of their white neighbours. Thus, the pressure for Indian land increased. A minority faction ceded their remaining land in Tennessee in a treaty that was enforced by the U.S. government. The result was the forcible removal of most of the remaining Cherokee to Oklahoma in the tragic Trail of Tears in 1838–39.

Women have played a prominent role in the life of the state from early pioneer times. Long active in religious and cultural pursuits, women began to organize in the latter part of the 19th century to gain the vote. These efforts accelerated in the 20th century and were crowned with success in 1920, when Tennessee's approval of the Nineteenth Amendment gave the vote to all American women. In recent years women in the state have become increasingly prominent in business, politics, education, literature, entertainment, and the arts.

People from other regions and nations have come into

Tennessee in recent years because of increased economic opportunities. They have generally found a warm welcome, and have, in turn, enriched the state through their energies and varied cultural patterns.

Religion has been important in Tennessee from the days of early pioneers. Presbyterian ministers reinforced a belief in divine Providence as settlers faced the hazards of nature, Indians, and internal divisions. Baptists soon outnumbered Presbyterians in Tennessee, and the Methodist circuit rider became a significant figure in the area. The Great Revival of 1800, spawned in nearby Kentucky, intensified the importance of religion in the state, where two new denominations emerged: the Cumberland Presbyterians and the Disciples of Christ, or Christian church. As immigration increased, so did religious diversity. Episcopalians, Lutherans, Roman Catholics, and Jews were all strong in the state before the Civil War. North–South division of denominations preceded the division of the Union. In the late 19th century the concern over the prohibition crusade caused denominations to join forces to combat the evils of drink. While this issue was still important in the early decades of the 20th century, churches also turned their attention to missions, urban problems, and church organization. Still, the popularity of itinerant evangelists signified the persistence of the revivalist tradition in Tennessee. In addition, churches concerned themselves with social needs during the Great Depression and in World War II, with the Civil Rights movement of the 1950s and '60s, and, later, with the homeless.

Revivalist tradition

In recent years the most rapid growth in the state has been in East and Middle Tennessee and the West Tennessee county of Tipton. The highest increases in population have been in fringe areas around the largest cities. While Tennessee exchanges population with other states, more people come from Florida, Kentucky, and Virginia than go from Tennessee to these states. Tennessee's percentage of native-born citizens is higher than the national average, although a significant number have lived outside the state for a time. In Tennessee, as in the nation as a whole, the number of households has increased, particularly single-person housholds. While the median age has increased, the distribution of population by race has remained extremely stable since 1970.

The economy. Manufacturing dominates the state's economy, accounting for nearly a third of the total state product.

Industry. The major products manufactured in the state are chemicals, foods, aluminum, rubber products, nylon, and whiskey. Because of the dependence on manufacturing, recent administrations have sought to attract industry to Tennessee. As a result, many firms, some from Japan, have located plants in the state. Tourism is an important industry because of the scenery, the facilities in parks, the abundance of historic sites, and entertainment facilities. Tennessee is also a centre for insurance, printing, and the recording of music. There has been an increase in service jobs in the state.

Agriculture, mining, and energy. The main crops are soybeans, tobacco, cotton, corn, and small grains, but livestock nearly equals crops in terms of cash receipts. The leading products in forestry are from oak, hickory, poplar, elm, pine, and cedar. Coal is the major mineral resource, followed by stone, zinc, phosphate rock, copper, and marble. Long known for Tennessee Valley Authority (TVA) dams and nuclear research at Oak Ridge, Tennessee has made important contributions in energy production. The TVA was established in 1933 to develop the resources of the Tennessee River valley. The nation's largest electric-power generating system, it has also improved navigation and controlled flooding on the Tennessee River and has aided industrial development throughout the Tennessee River valley.

Tennessee Valley Authority

Transportation. Tennessee's river system is a vital component of the state's transportation complex. Railroads remain important, despite a general decline. The most notable recent addition to the highway system has been the network of interstate routes and urban beltways. Nashville and Memphis are important regional transportation centres, not only in highways but also in air travel.

Administration and social conditions. *Government.* The constitution of 1870, drawn up after Reconstruction, closely resembled the original document of 1796. Despite the recognized need for revision in the following decades, the amendment process itself remained difficult. Hence, it was not until 1953 that the document was actually amended, although the government had been reorganized by statute in the 1920s and '30s. The 1953 and subsequent amendments helped to modernize the state government.

Among the changes effected were clauses that simplified the complicated amendment procedure, increased the terms of the governor and state senators from two to four years, increased the pay of state legislators, provided for annual rather than biennial meetings of the state legislature, abolished the poll tax, and increased the power of cities to govern themselves independently of the state.

The structure of the executive, legislative, and judicial branches of government in Tennessee resembles that of many other states. The governor is the only executive official elected statewide. The speaker of the state Senate serves as lieutenant governor. Other executive officers are elected by the legislature, while the attorney general is appointed by the Supreme Court. Executives of major departments and important state commissions are appointed by the governor. The bicameral General Assembly comprises the 33-member Senate and the 99-member House of Representatives. Representatives are elected for two-year terms. The General Assembly can override a governor's veto by a simple majority.

The judicial system consists, on the local level, of general sessions judges who act as committing magistrates and enjoy limited civil jurisdiction. Above these is a complex of inferior courts—chancery, circuit, criminal, and probate—that try various types of cases. The Court of Appeals, the Court of Criminal Appeals, and the Supreme Court have their judges apportioned among the three "grand divisions" of the state (East, Middle, and West Tennessee). Judges on the local level are elected by the voters of their respective counties. On the next level, judges are elected from their respective chancery divisions and circuits. The members of the Court of Appeals and of the Court of Criminal Appeals are elected by the state at large under the Missouri Plan. The members of the Supreme Court are elected by the voters of the state at large. All terms are for eight years.

Local government follows the national pattern. A county commission consisting of commissioners from the civil districts into which the counties are divided constitutes the legislative authority of most of the state's 95 counties. School boards, either elected or appointed, administer the schools under the direction of county superintendents. Popularly elected sheriffs enforce the criminal laws, and elected officers collect the property taxes and record real estate transfers. The chief executive officers of the counties are the county executives. Types of city government include the council–manager, mayor–council, and commissioner systems. In 1962 Nashville and Davidson county merged into a single governmental unit, called Metropolitan Government, or Metro. This experiment has been considered a success in solving problems of overlapping or conflicting authority in city and county governments.

Experimentation in local government

Education. Almost half of every state tax dollar goes to public education. The State Board of Education administers elementary and secondary education in the state. One notable innovation has been the Better Schools program, which rewards teachers for upgrading their credentials and for their performance in the classroom. In the elementary schools there has been increased emphasis on art, music, and physical education, while on the secondary level requirements in math and science have been increased. The Tennessee Higher Education Commission coordinates the work of two boards, the University of Tennessee Board of Trustees and the Board of Regents of the State University and Community College System of Tennessee. The University of Tennessee system consists of campuses at Knoxville, Memphis (the medical school and other schools related to health services), Martin, and Chattanooga. There are six regional universities: Austin Peay (Clarksville), East Tennessee (Johnson City),

Memphis State, Middle Tennessee State (Murfreesboro), Tennessee State (Nashville), and Tennessee Technological (Cookeville). There are several community colleges and technical institutes. Tennessee has long been known for its private colleges, of which Vanderbilt and Fisk universities, both at Nashville, and the University of the South at Sewanee are perhaps the best known.

Health and welfare. The increased longevity of Tennesseans is in part attributable to improved health facilities, characterized by outstanding medical centres in the major cities. The Department of Human Services has expanded its program in an attempt to meet the growing needs of the underpriviledged in society. In addition to basic welfare services, the state concerns itself with foster care, adoptions, licensing of day-care centres, prevention of child abuse, and provision of services for the handicapped.

Cultural life. The geographic, economic, and social divisions of the state are reflected in a diversified culture. The self-reliance of the pioneer tradition helped to shape the music, crafts, and legends of East Tennessee, while the slave heritage of West Tennessee blacks gave rise to the blues. Memphis has long been a centre for the arts, with special emphasis on music and the theatre. Middle Tennesseans brought religious, educational, and other cultural institutions into their region to mitigate the crudeness of frontier life. In more recent times Nashville has become the cultural as well as the political capital of the state, weaving traditional patterns of the three "grand divisions" with contemporary trends in music and literature. Beginning in 1925, when the "Grand Ole Opry" radio program was first broadcast, Nashville has been the centre for country music, including recording and publishing. Modern literature also has important ties to Nashville. In the 1920s the so-called Fugitive poets, associated with Vanderbilt University, gained international attention. Four of them, John Crowe Ransom, Donald Davidson, Allen Tate, and Robert Penn Warren, joined with eight other writers in contributing essays for *I'll Take My Stand* (1930), a defense of traditional agrarian culture against the changes in values associated with industrialization. This theme persists in the work of a number of distinguished Tennessee writers, of whom Peter Taylor is probably the best known.

Country music

The completion of the Tennessee Performing Arts Center in Nashville has led to a flowering of cultural life in music, theatre, and dance. Exhibits of contemporary and earlier art as well as historical artifacts at the Tennessee State Museum, also in Nashville, demonstrate the cultural richness and vitality of the state.

Large numbers of vacationers visit the state each year. Opryland USA, a musical theme park featuring the Grand Ole Opry, is a major attraction. The natural beauty of the region draws visitors to Great Smoky Mountains National Park (shared with North Carolina), the largest national park in the eastern United States, and to Tennessee's 51 state parks, many of which encompass man-made lakes. In addition, there are many historic sites, of which the Hermitage, home of Andrew Jackson, near Nashville, and the Civil War battlefield of Shiloh are the most famous.

(S.McC.H/J.A.Ho.)

HISTORY

Prehistory and settlement. The first inhabitants of Tennessee are believed to have been Ice Age hunters descended from Asians who crossed the former Bering Strait land bridge more than 20,000 years ago. They were succeeded by various groups, who refined hunting methods and ultimately developed a life based on agriculture. The European explorers, beginning with the Spaniard Hernando de Soto in 1540, found several groups of Indian peoples in Tennessee, the most powerful of whom were the Cherokee, who succeeded in driving the other Indians out of the state by the early part of the 18th century. The name Tennessee derives from that of the Cherokee village Tanasi. The Cherokee developed warm relations with English traders from Virginia and South Carolina and were initially their allies in the French and Indian War of the 1750s. As English traders and hunters became land-hungry settlers, the Cherokee came to see them as a threat. Thus began a long period of intermittent conflict, which ended with the final removal of the Cherokee from the state in 1838–39.

As for the English settlers, a group in upper East Tennessee, learning that they were not under royal authority, set a precedent for self-government in the Watauga Association in 1772, the example of which was later followed by the signers of the Cumberland Compact on the site of Nashville. An important group of Tennesseans showed their support for independence during the American Revolution by contributing to the defeat of the Tories in the Battle of Kings Mountain in South Carolina in 1780. This was one of several encounters that encouraged British leaders to withdraw their forces.

The Watauga Association

At first a part of the new state of North Carolina, Tennessee made a bid for admission as "the state of Franklin." Because North Carolina had rescinded its original cession of western lands, however, the Continental Congress turned down this petition for statehood. Under the new federal Constitution, the region was organized as the Territory South of the River Ohio. In 1796 Tennessee became a state, the first admitted from territorial status.

The Jackson era. With the coming of the War of 1812, Tennesseans played a decisive role as volunteers under the leadership of General Andrew Jackson, whose victory at New Orleans discouraged the British from renewing hostilities. While the Democrat Jackson, elected president in 1828 and 1832, became the hero of the common man, he was opposed in Tennessee, as in other parts of the country, by the growing commercial faction. His champion in the U.S. House of Representatives, Tennessean James K. Polk, was elected president in 1844 although the majority of Tennesseans, sympathetic to the commercially oriented Whig Party, voted against him.

The Civil War and Reconstruction. As the Civil War approached, the majority of Tennesseans remained loyal to the Union, until Lincoln's call for volunteers indicated that he would hold the Union together by force. Then Tennessee, like other states in the upper South, voted for secession. Only Virginia was the scene of more fighting during the Civil War. Names such as Fort Donelson, Shiloh, Stones River, Chattanooga, Knoxville, Franklin, and Nashville evoke images of extreme sacrifice, both Union and Confederate.

While Middle and West Tennessee were sympathetic to the South, the majority of East Tennesseans remained loyal to the Union. This turmoil was reflected in the career of Andrew Johnson, popular Democratic governor and U.S. senator before the war. His loyalty to the Union during the war and his position as military governor of Tennessee subjected him to threats by many people in the state. He became a hero once again when he, as president, was impeached by a House of Representatives dominated by Radicals. (The Senate, however, fell one vote short of the majority needed to remove him from office.)

Andrew Johnson's career

During the Reconstruction era, Tennessee provided its own Radical governor, William G. Brownlow, who expressed his hatred for former rebels in vituperative rhetoric. Radical Reconstruction lost popularity in Tennessee as it subsequently did in the North. However, former Confederates regained their rights, and blacks lost the little political power they had gained.

The modern period. Factionalism within the ascendant Democratic Party and popular crusades such as prohibition and women's rights captured public attention in the early 20th century. The preoccupation with prohibition delayed effective reform of state government until the ascendancy of Governor Austin Peay in 1922. Plagued by the Great Depression and party factionalism in the early 1930s, Tennessee resumed its reform program under Gordon Browning, who was elected governor in 1936. Prentice Cooper was the frugal but efficient governor of the wartime years. After World War II, under the leadership of governors Frank G. Clement and Buford Ellington, the state gave increased attention to education, mental health, highways, and constitutional reform, and Tennessee became a testing ground for breaking the barriers of racial segregation in schools and in other public facilities. A small minority of white extremists used violence in opposing integration. In the 1970s and '80s, with Republicans twice winning the

gubernatorial chair, Tennessee once again became a two-party state. Tennesseans have been increasingly concerned with improving educational opportunities and attracting outside industry. (S.McC.H.)

Virginia

Virginia, nicknamed the Old Dominion for its loyalty to the exiled Charles II of England during the Puritan Commonwealth, was one of the 13 original colonies. It has one of the longest continuous histories among the American states, dating from the settlement of Jamestown in 1607. It was named for Elizabeth I, the Virgin Queen, and under its original charter was granted most of the unexplored lands west of the Atlantic seaboard settlements, to the Mississippi River and beyond. The contributions of such Virginians as George Washington, Thomas Jefferson, and James Madison were crucial in the formation of the American nation, and in the early decades of the republic the state was known as the Birthplace of Presidents.

The Jamestown settlement

Although Virginia gave its support—including the leadership of Robert E. Lee and other generals—to the Confederacy during the Civil War, it has developed in the 20th century into a bridge state between the North and the South. Its northern counties reflect the cosmopolitan character of the national capital, Washington, D.C., which lies across the Potomac River to the north. Other areas of the state retain the tinge of conservatism developed over centuries of agricultural life and through aristocratic traditions that made the term "a Virginia gentleman" synonymous with gentility and refinement.

History and nature make Virginia a leading tourist centre. Within its borders lie many important historical monuments. They include colonial restorations and reconstructions, such as those at Williamsburg; the homes of Washington, Jefferson, and other noted Virginians; and many of the battlefields of the War of Independence and the Civil War. Although it is becoming increasingly an industrialized and urbanized state, slightly more than three-fifths of Virginia's land remains under forest cover as it descends from the mountains and valleys in the west to the beaches of the Atlantic shore. Virginia has an area of 40,767 square miles (105,587 square kilometres). It is bordered by Maryland to the northeast, North Carolina and Tennessee to the south, Kentucky to the west, and West Virginia to the northwest. The state capital is Richmond.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Western Virginia comprises three physiographically defined mountain provinces. From west to east, they are the Appalachian Plateau, the Ridge and Valley province, and the Blue Ridge. The state's highest point, Mount Rogers, at 5,729 feet (1,746 metres) above sea level, lies in the Blue Ridge area. The Piedmont province of middle Virginia is a region of lower rolling hills, reaching from the Blue Ridge to the fall line, the place where rivers descend, often in rapids, from higher and geologically older regions onto the flatter coastal plains. The Coastal Plain province (also known as the Tidewater) lies low between the fall line and the Atlantic coast, deeply interlaced by its tidal rivers and dominated by the Northern Neck Peninsula, the Middle Peninsula, and The Peninsula (so called because of the many points of cultural interest thereon)—all west of Chesapeake Bay. East of the Chesapeake and separated from the rest of the state is the Eastern Shore, the southern tip of the Delmarva Peninsula, which Virginia shares with Delaware and with Maryland's Eastern Shore. The Coastal Plain also contains the area south of the James River, including the Norfolk region and the 750-square-mile Great Dismal Swamp, which extends south into North Carolina.

Mountain heights and seascapes

Drainage and soils. The eight major drainage systems that empty into the Atlantic include the Potomac, which receives the waters of the north-flowing Shenandoah River at Harpers Ferry, in West Virginia, and becomes the state's border with Maryland on its way to Chesapeake Bay. The Rappahannock, York, and James rivers indent the coast to form the main peninsulas. Two other systems pass into North Carolina, while in the extreme southwest corner of the state two major systems flow eventually into the Gulf of Mexico. The soils of Virginia are generally fertile. In the Coastal Plain, the tidal lowlands are usually covered with loam, a mixed soil rich in organic materials. To the west, sandy loams and clays predominate. In the Piedmont, clay and limestone soils dominate, and limestone soils are found in the valley areas west of the Blue Ridge.

Climate. The state's climate, generally mild and equable, varies according to elevation and proximity to Chesapeake Bay and the Atlantic. In southeastern Virginia and the Eastern Shore, January temperatures average about 40° F (4° C); July temperatures average about 78° F (26° C). These temperatures allow growing seasons of up to eight months, three months longer than those in far western Virginia. Elsewhere in the Coastal Plain and Piedmont, continental weather overcomes the eastern marine influence to produce colder winters. In the mountains, winter temperatures of 0° F (−18° C) may occur, but cool nights in summer follow daytime highs that usually stay below 90° F (32° C). Throughout the state, rainfall averages from about 32 to 44 inches (813 to 1,118 millimetres). Snowfall averages from a few inches in the southeast to about 30 inches in the mountains.

Plant and animal life. Forests of the Coastal Plain and Piedmont areas have mainly pine and some hardwood. Cover other than trees includes marsh grass in the Tidewater section, and broom sedge, crabgrass, wire grass, and cultivated crops elsewhere. The mountainous areas contain tracts of various coniferous species and hardwoods such as hickory and oak. Bluegrass and field crops generally cover nearby valleys. Wildflowers and berry bushes abound, depending on climate and soils.

At the time of European settlement of the Great Valley of Virginia (*c.* 1730), large herds of native bison abounded along the banks of the Shenandoah, but, as elsewhere, the bison populations were destroyed. Black bears are found in Virginia's mountains and in the Great Dismal Swamp. Common wild animals are rabbits, chipmunks, squirrels, opossums, muskrats, woodchucks, foxes, and deer. Less common are otters, beavers, mink, and wildcats. The main game birds are doves, quail, ducks, and geese; a few wild turkeys and woodcocks may be found. Scavengers include coastal sea gulls and the ubiquitous turkey buzzard. Predatory birds include a number of hawks, owls, and the scarce golden and bald eagles. There are numerous songbirds, including the cardinal, the state bird. Poisonous reptiles include rattlesnakes, copperheads, and water moccasins. Game fish and panfish abound in Virginia's inland waters and offshore. Chesapeake Bay is the world's richest marine-life estuary, noted for finfish, blue crabs, oysters, and clams. Yearly commercial and sport fishing catches indicate generally plentiful stocks of these fishes and shellfish.

Mammals and birds

The people. Nearly 80 percent of Virginia's people are of European descent. Most of the rest are black, serving as a reminder of the important role that African slaves played in the development of the state. Few of the state's residents are foreign-born, but various ethnic groups are represented, especially in the northern counties located around Washington, D.C.

Most of eastern Virginia was first settled by English settlers from the Midland and southern counties of England, especially from and around London. During the 1700s the Welsh and the French Huguenots were prominent among the immigrants, and a large number of people of Scotch-Irish and German descent moved from Pennsylvania into the Shenandoah valley. People of Scotch-Irish and English ancestry still predominate, notably in western and southwestern counties. Over the centuries, differences in speech developed as a result of both class structure and isolation. Folk speech was largely localized, the more "cultured" or mainstream patterns attaining a wider regional usage. Virginia's main speech patterns are Southern, but population mobility has diffused the local patterns and introduced others from different parts of the United States.

Patterns of immigration

When the English colonists founded Jamestown, England's first permanent settlement in the New World, American Indians lived all around them. From these Indians the colonists learned tobacco cultivation, and tobacco became the mainstay of Virginia's early agricultural econ-

omy. In larger perspective, what are now the boundaries of the Commonwealth of Virginia contained in 1607 three linguistic stocks: the Algonquian, the Iroquoian, and the Siouan. Estimates of the Algonquian population in 1607 range from 14,000 to 21,000. Today in Virginia there are only two American Indian reservations; one each for the Pamunkey tribe and the Mattaponi tribe, respectively situated along the Pamunkey River and the Mattaponi River near West Point, where these two rivers join to form the York River. Although some Native Americans live throughout the state, especially in the urban areas of Washington, D.C., Norfolk–Virginia Beach–Newport News, and Roanoke, the only other present concentration in Virginia is that of the Chickahominy Indians, clustering along the Chickahominy River through New Kent and Charles City counties to its confluence with the James River. The Pamunkey, Mattaponi, and Chickahominy are Algonquian tribes.

Blacks were first brought to Jamestown in 1619 as indentured servants. Legalized slavery was not introduced for several decades. Black slaves were the foundation of the plantation agriculture that began in the Tidewater and spread into the Piedmont. At the start of the Civil War, about half the state's population was black, a proportion that has dropped drastically since then, despite the doubling of the total number of blacks.

Urban growth

In the 20th century the greatest growth has occurred in the urban corridor, an area that stretches south from Washington, D.C., through Arlington and Alexandria to Richmond and bends southeast to the Hampton Roads area, comprising Newport News, Hampton, Norfolk, Virginia Beach, and Portsmouth. This corridor is often classified as an extension of the great population mass, or megalopolis, arcing across the northeastern United States from Boston to Washington, D.C. Other major metropolitan areas include those around Roanoke and Lynchburg, with a number of emerging metropolitan areas around Danville, Bristol, and Charlottesville.

The increase in agricultural mechanization and productivity, with an attendant decrease in acreage and number of farms, has sent both blacks and whites to the cities for their livelihood. Richmond is more than one-half black; rural Charles City county, lying in the urban corridor, is more than three-fourths black. The counties west of the Blue Ridge have mainly small, family-run farms, and the black population is small in comparison with the Piedmont and Tidewater regions.

Throughout Virginia, race relations have been less troubled than in many other parts of the nation. Despite some early concern with federal school-desegregation orders, a general atmosphere of mutual interest and moderation has developed over the years during which black and white citizens have worked together.

Bruton Parish church in Williamsburg, still active, was the established church in the colonial capital. The Anglican church, which was disestablished during the Revolution, became the Episcopal church, retaining only one-third of the population adhering to a denomination. Dissenters, primarily Presbyterians, Quakers, Baptists, and Methodists, made up the Protestant balance. Virginia continues its Protestant tradition, although there are many Roman Catholics. The largest denominations are the Southern Baptists, United Methodists, Roman Catholics, Episcopalians, and Presbyterians.

The economy. Virginia has developed a well-balanced economy far beyond its original agricultural base. Farming has dropped to approximately 1 percent of the total yearly value of the state's goods and services, surpassing only the economic contribution of commercial fishing. Virginia's economy has changed markedly toward a mixture of the old and new and the opportunity for increases in the standard of living from both private and public initiatives and cooperation.

Mixed economy

The role of the state government in the economy is considerable in terms of the state's revenues and expenditures, its services and promotion, and its controls and enforcement of regulations through a host of agencies. Most management of the economy, however, rests in the private sector, with due regard to governmental regulations, public needs, and stockholder pressures. The major private coalition for helping business is the strong Virginia Chamber of Commerce, with counterparts in local chambers. Also, powerful in Virginia's manufacturing sector is the Virginia Manufacturers Association. Trade associations abound in the state, from the Virginia Travel Association to the Virginia Turkey Association. Important assistance is maintained for broad business and economic development at the state level through the cooperative Department of Economic Development under the Governor's Office. Besides this department's main office in Richmond, it has regional offices in Abingdon, Staunton, and South Boston and foreign offices in Brussels and Tokyo.

There is good support for the state's economic growth in its proximity to Washington, D.C., and the consequent benefit of further research and development at many of the military and other federal installations in Virginia. The state's historical heritage and natural beauty also offer much in the way of recreation, and its excellent port facilities are among the busiest in the nation. Virginia's income per capita is the highest of the southeastern states and above the national average.

Beginning in the 1960s, Virginia's economic productivity experienced an average annual increase slightly higher than that of the United States as a whole. Manufacturing was the leading sector. The federal government remains a dominant economic presence in Virginia, being responsible for nearly one-seventh of the state's gross product. Virginia also ranks as one of the top five states in per capita distribution of federal funds. Other increasing segments have been services and trade, including tourism, and transportation and public utilities. Some improved productivity has also come from finance, insurance, and real estate services, as well as from local government and from mining. Construction has undergone a relative decline, as has farming, in its share of productivity. Commercial fishing, while very small, continues to contribute to Virginia's total productivity.

With regard to employment in Virginia, wholesale and retail trade provide the most jobs, followed by services, federal, state, and local government, and manufacturing. Others, in descending order, are construction, transportation and public utilities, financial services, farming, mining, and commercial fishing.

Following the pattern of Southern states, Virginia has a right-to-work law that forbids the all-union, or closed, shop. As a result, only a small percentage of Virginia's nonagricultural labour force is unionized. Unemployment in Virginia traditionally is comparatively low, compared with both the United States as a whole and the neighbouring states of the Southeast.

Government operations. State and local government provide about twice as many jobs in Virginia as the federal government provides directly. However, through contracts to private firms in the state, federal operations account for a significant further percentage of employment in private industry and business.

Military installations

The concentration of military facilities alone in Virginia covers nearly 450 square miles, including the Pentagon. Numerous army installations feature training, engineering, supply, and transportation services throughout the state; and all have considerable effect on local economic conditions and employment, both military and civilian. The transportation training centre for the army is located at Fort Eustis; and nearby is historic Fort Monroe, headquarters of the Army Training and Doctrine Command and ROTC Cadet Command. Naval activities are concentrated around the Norfolk naval base, the largest U.S. Navy installation in the world. The Marine Corps installation at Quantico is a major development and education base. The U.S. Air Force and the National Aeronautics and Space Administration (NASA) have major installations in the Hampton–Newport News area. The U.S. Coast Guard has a large facility near Yorktown, as do the U.S. Naval Weapons Station and the Naval Supply Center. Military intelligence training is conducted at Camp Peary Naval Reservation near Williamsburg.

Manufacturing. Tobacco and chemical products head the list of products made in Virginia. Other nondurable

goods include food, textiles, and apparel. The leading durable goods are transportation equipment, centred on Tenneco's experienced Newport News Shipbuilding and Dry Dock Company, which is the world's largest privately owned shipyard. Other important durables include furniture, electrical equipment, and wood products.

Farming, forestry, mining, and commercial fishing. Truck farms dot the Eastern Shore and Norfolk areas. Other farms are spread throughout the state. Products include dairy products, grains and feeds, and vegetables. Tobacco, however, is still featured in the southern Piedmont; and the state's apples and peaches are famous, especially those from the huge orchards around Winchester and in the horse-breeding country of northwestern Virginia. Rockingham county in the Shenandoah valley has one of the nation's major turkey-raising operations. Virginia's major forest product is pine timber. The main commercial minerals are coal from the southwest and stone, clay, sand, and gravel from many areas. Products from Chesapeake Bay include flounder, bass, and a number of other edible fish. Additionally, large amounts of schooling menhaden sighted by aircraft spotters are seined and processed for their oil and for protein-rich fish meal. Oysters are tonged, hard and soft clams are dug and dredged, and blue crabs are trapped and netted. Offshore, large quantities of sea clams and scallops are harvested. In addition, large ocean fish, such as swordfish and tuna, are caught on baited hooks strung out on lines relatively close to the surface of the Atlantic. Some lines are as long as 50 miles. Bans on fishing in the James River downstream from Hopewell in the early 1970s because of spills of pesticide were essentially removed in 1988. Virginia and Maryland have both passed new antipollution laws for Chesapeake Bay for purposes of conservation and recreation.

Marine products

Transportation. Although there are excellent pipeline and water transportation facilities and the port of Hampton Roads is usually the leading U.S. port in foreign tonnage, Virginia's major transportation facilities are roads, railroads, and airports. Most traffic is north–south, adding to Virginia's status as a "bridge" state. Among the many scenic routes is the Colonial National Historical Parkway, connecting Jamestown, Williamsburg, and Yorktown; planned in the 1930s, this road has only two lanes. The Skyline Drive and Blue Ridge Parkway, which join at Rockfish Gap to form a continuous road following the crest of the Blue Ridge Mountains of Virginia and North Carolina, offer spectacular views and park facilities. Also striking is the 23-mile (37-kilometre) Chesapeake Bay Bridge-Tunnel linking Cape Charles on the Eastern Shore with the vacation centre of Virginia Beach, east of Norfolk. Comprising a trestled roadway raised above the mouth of the Bay and two tunnels (under the main shipping channels), it is the nation's largest structure of its kind.

©Medford Taylor—FOCUS/VIRGINIA

Seining for menhaden in Chesapeake Bay, off the Virginia coast near Reedville.

A network of commercial airports, the largest being Dulles International, Washington National, Richmond International, Norfolk International, and Patrick Henry (Newport News) International, makes air transportation easily available. A number of large transportation corporations, including CSX and Norfolk Southern, two of the world's largest railroad-based transportation companies, have their headquarters in the state.

Financial services. As banking has crossed state lines under deregulation, several Virginia-based institutions have shown strong growth in the region. Also of note is the location in Richmond of the Fifth Federal Reserve District Headquarters.

Administration and social conditions. *Government.* The Commonwealth of Virginia's constitution, of 1776, was revised for the seventh time in 1971. In it the state retains the basic powers first delineated separately in the third constitution (1851), which enumerated the organization of the executive, legislative, and judicial branches. The only elected administrative officials are the governor, the lieutenant governor, and the attorney general; each serves a four-year term, but the governor is the only one who cannot succeed himself. The General Assembly includes a 40-member Senate and a 100-member House of Delegates, which meet annually in Richmond. State and local government is structurally simple and intended to be responsive to citizens' needs and complaints. The 95 counties are governed by elected boards of supervisors. Below them are about 190 towns within the counties and some 40 chartered cities separate from county administration and governed by elected councils employing city managers.

The Democratic Party thoroughly dominated state politics from its revival in 1883 until the Republicans elected their first governor in a century, in 1969, and six of the state's 10 U.S. representatives, in 1970. In the following decades, the two parties competed closely for domination of the state's congressional delegation. The General Assembly, however, has remained substantially Democratic, and the electorate continued to show moderately conservative tendencies.

Dominance of the Democratic Party

The Virginia judicial system is relatively uncomplicated. With four levels of courts, it avoids numerous special courts and levels at which cases may be originated or appeals made. The seven judges of the Virginia Supreme Court, the highest state judicial body, are elected to staggered 12-year terms by the General Assembly. The primary work of this court includes hearing criminal and domestic appeals from the Virginia Court of Appeals and civil appeals from the circuit courts; exercising original jurisdiction over cases of habeas corpus, mandamus, and matters filed by the Judicial Inquiry and Review Commission; and developing the body of Virginia common law. After a 15-year study conducted by the Judicial Council, the General Assembly created the Virginia Court of Appeals, which began hearing cases on Jan. 1, 1985. The primary purpose of this court is to expand judicial capacity to relieve the immense backlog of criminal and domestic cases pending before the Supreme Court. There are 10 judges on this court, elected by the General Assembly to eight-year terms. The 31 judicial circuits are the courts of general jurisdiction in Virginia. Judges of these courts are elected to eight-year terms by the General Assembly. Other courts include general district (municipal and county), juvenile, and domestic-relations courts. Additionally, all judicial circuits have magistrates who have the authority to issue warrants but lack trial jurisdiction. Counties and cities have a commonwealth's attorney, whose main job is criminal prosecution.

Law enforcement is largely the responsibility of local police departments and county sheriffs. The main job of the Department of State Police is to enforce the state's highway laws.

Education. The public schools, which date as an institution from 1846, have accelerated improvements in facilities and curricula since the 1960s, and average num-

ber of years of education completed by Virginians has risen. The state Board of Education supervises public primary and secondary education, and the State Council of Higher Education coordinates higher public education.

Higher education

Virginia has a strong public community-college system, with branches throughout the state. Its four-year public and private colleges and universities are numerous. The best known are the College of William and Mary in Williamsburg, a state university founded in 1693 and the second oldest college in the nation, and the University of Virginia in Charlottesville, founded in 1819 largely as the creation of Thomas Jefferson, both in its organization and in the design of its buildings and grounds. Virginia Polytechnic Institute and State University, in Blacksburg, is a large land-grant institution. James Madison University, in Harrisonburg, was previously a state teachers' college for women. Substantial urban state universities have developed starting in the 1960s; notably George Mason University, in Fairfax; Virginia Commonwealth University, in Richmond; and Old Dominion University, in Norfolk. Well-known private Washington and Lee University and state-supported Virginia Military Institute are both located in Lexington. Hampton University, private and predominantly black, has a national reputation. Norfolk State University is the state's largest predominantly black public university. A substantial percentage of the state's many private colleges were founded by religious denominations.

Health and welfare. Private and public facilities in Virginia share in providing health care. Public services involved directly or indirectly include water and sewage, immunization, and care for crippled children and the mentally ill. Welfare and public assistance are administered in conjunction with local welfare boards and superintendents.

Cultural life. Many of Virginia's settlers and leaders were immigrants from the educated, often aristocratic, classes in England. They set the tradition of a cultural life that appears somewhat contradictory when considered against the state's long agricultural background.

The arts are an active concern of the state government, as well as of private patrons. The Virginia Museum of Fine Arts in Richmond was the first state museum of the arts when it was established in 1934. The museum also sponsors an active program in the performing arts. Using local arts groups as affiliates, it exerts a statewide influence. In addition to the Virginia Museum there is a state Commission of the Arts and Humanities.

The Barter Theatre was founded by actor Robert Porterfield in 1933 in the tiny southwestern town of Abingdon; its original charge for admission was produce, handicrafts, or whatever the prospective viewer could afford. Dozens of art galleries are located throughout Virginia. There are several ballet companies, orchestras, civic choruses, and opera and theatre companies, as well as numerous festivals of the arts.

Historical sites

Millions of visitors annually are attracted to Virginia's historical sites. In the forefront is Colonial Williamsburg, its many restored or reconstructed 17th- and 18th-century buildings representing expert research. Striking examples of colonial architecture are found at such preserved homes as George Washington's Mount Vernon, near Washington, D.C., and Thomas Jefferson's Monticello, near Charlottesville. Civil War monuments in the state include the battlefield known to Southerners as Manassas and to Northerners as Bull Run, near Washington, D.C., and Appomattox Court House, the site of General Robert E. Lee's surrender to General Ulysses S. Grant in 1865. There are more than 100 historical societies and museums in the state and several scholarly and popular historical journals with national readership.

HISTORY

Settlement and colonial period. The purposes of the Virginia Company that landed at present-day Jamestown in May 1607 were not only to colonize but also to Christianize, to open new areas for trade, and to guard against further Spanish inroads. Hunger, poor shelter, Indian hostility, and rampant disease plagued the early years, but, while the colony tottered constantly on the brink of dissolution, a tobacco industry was begun by John Rolfe and a representative assembly was convened. In 1624 the company's charter was revoked and Britain's first royal colony established. In the following years new settlements were made and local administrative systems were devised.

The governorship of Sir William Berkeley—begun in 1642, interrupted by Puritan rule from 1652 to 1660, and ending in 1677—marked the solidification of the colony. The many anti-Puritan cavaliers who fled to Virginia after 1649 added an important element to the population, much of which consisted of indentured white and black servants. A rebellion in 1676, led by Nathaniel Bacon, though short-lived, led to Berkeley's recall and signaled a growing restlessness for more self-government among the colonists. This sentiment became strong during the century that followed, when England attempted to govern fairly but did not allow the inhabitants of its American colonies the full rights of Englishmen at home.

This was a period of expansion as well as of internal strengthening. Settlers from the Tidewater region spilled over into the Piedmont, across the Blue Ridge, and, by the 1740s, into the Ohio country beyond, there running afoul of French ambitions for that region. For decades the popularly elected House of Burgesses led the way in opposing royal prerogatives in the colony, and, following England's prohibition of westward expansion in 1763, a concerted drive to rationalize rebellion began. On the eve of the Revolution, Virginia had more than 120,000 residents, many of them persons of considerable sophistication and learning, and a stable—if narrowly based—economy.

Independence and statehood. Virginians were among the leaders of the American Revolution and of the events leading to it, including the calling of the first Continental Congress in 1774. Thomas Jefferson was the primary author of the Declaration of Independence, while George Washington assumed command of the armies. It was at Yorktown that the British armies were forced to surrender to combined American and French forces on Oct. 19, 1781, leading to acknowledgement of the colonies' independence in the Treaty of Paris of 1783. In 1788 Virginia became the 10th state to ratify the Constitution.

Leadership during the Revolution

The state continued its national leadership in the following decades, furnishing four of America's first five presidents and, especially through Jefferson and James Madison, much of the intellectual ferment out of which the basic political institutions of the young nation gradually were shaped. The state had abolished the African slave trade in 1778, but slavery itself continued as the basis for the state's agricultural economy. Nat Turner's slave insurrection in Southampton county (1831) raised tensions.

(C.L.Q.)

Civil War and Reconstruction. In 1861 Virginia seceded from the Union. Richmond became the capital of the Confederacy, and Virginia was the chief battleground throughout the war. In 1863 the state lost one-third of its territory to form West Virginia. In 1867 Congress placed the South under military rule, and Virginia was not readmitted as a state of the Union until 1870.

Virginia escaped much of the punishment of Reconstruction, but it had lost thousands of its young men and had been devastated by invading armies; its banks had been closed, its currency turned into worthless paper, its labour force demoralized, and its territory occupied by its former enemy. In April 1871 it also had a prewar debt, made for internal improvements, amounting to over $45,000,000, more than one-third of which was interest accrued during the war and Reconstruction.

Strife over the state debt was the prominent feature of political life during the 1870s and '80s. When the pre-Reconstruction leaders regained control, they provided for payment of the entire debt, designating one-third as West Virginia's share, for the payment of which Virginia assumed responsibility. The bankrupt state could not meet its obligations to its citizens and pay interest to its creditors, however, and the new system of public schools, organized in 1870, suffered. Then in 1882 a group known as the Readjusters, claiming that the debt and interest needed pruning and with aid of the Republicans, seized control of the government and "readjusted" the debt. Not until 1891–92 was a satisfactory compromise settlement

reached with the creditors, and later West Virginia paid its share. The Democratic Party was revived in 1883. Virginia adopted a new constitution in 1902.

The 20th century. In 1926 Harry F. Byrd became governor of Virginia, and he revolutionized the governmental machinery. During the first 60 days of his administration, the General Assembly revised the tax system; reformed the fee system; initiated constitutional amendments that shortened the ballot, concentrating authority in the governor's hands; and encouraged industries to settle in the state. After World War I the state's prosperity increased as agriculture diversified, manufacturing grew in importance, and tourism became a major enterprise.

Reforms of 1926

The Great Depression of the 1930s was less severe in Virginia than in many other states. In the period before U.S. entry into World War II, Virginia was the first to set up a state defense system. The war brought tens of thousands of soldiers into its military camps. The Hampton Roads area had a great boom with the expansion of the Norfolk naval base and the shipbuilding activities in Newport News. Employment continued at a high rate after the war, with continued growth in the nonagricultural sector, including government.

In presidential elections the state voted for the Republican candidate in 1928, for the first time since 1872. In 1948 the plurality of voters supported the Democratic candidate, Harry S. Truman, while the majority was split between the Republicans and Dixiecrats; and in subsequent elections the majorities were usually Republican.

In the state, political leadership continued to be exercised by former governor Byrd, who became U.S. senator by appointment in 1933 and by election in 1934, remaining in office until his retirement in 1965. Before Byrd's death in 1966 his organization was rapidly becoming more sensitive to the political, social, and economic needs and rights of all the people.

A revised constitution, effective July 1, 1971, did not change the fundamental structure of the state. It deleted outmoded and nonfundamental sections and added provisions responding to the requirements of modern conditions, such as an updated fiscal policy, annual legislative sessions, conservation and environment.

Following the U.S. Supreme Court's order in 1954 to desegregate public educational facilities, the schools of Prince Edward county gained nationwide attention by closing their doors from 1959 to 1964. Despite such events and a few sporadic outbreaks, Virginia was able to avoid much of the racial strife afflicting North and South alike and moved carefully to implement civil rights laws. The principal challenge of the late 20th century appeared to be the problems of increasing urbanization.

(R.L.Mo./C.L.Q.)

West Virginia

West Virginia justifies in every way its nickname, the Mountain State. With an average altitude of 1,500 feet (457 metres) above sea level, it is the highest of any U.S. state east of the Mississippi River. It is a region tied economically and socially to the mountain spines that span its length and breadth and to the rivers that enclose it on many sides. The settlers of northwestern Virginia who, in 1861, defied the state's secession convention, choosing instead to remain within the Union and to form a new state, acted much in the tradition suggested by the motto of West Virginia, "Montani semper liberi" ("Mountaineers always freemen").

West Virginia, admitted to the Union as the 35th state in 1863, is a relatively small state, with an area of 24,232 square miles (62,761 square kilometres). It is bordered by Pennsylvania to the north, Maryland and Virginia to the east, Kentucky to the southwest, and Ohio to the northwest. The state capital is Charleston. In comparison with the national standards and averages of the United States, West Virginia is poor in personal incomes and in overall economic development. For decades the rich coal beds underlying most of the 55 counties made West Virginia a leading producer of bituminous coal in North America. The gnarled terrain locked West Virginians into their small communities in the narrow valleys and posed both literal and symbolic obstacles to people from the outside world. From 1940 to 1970 and again in the 1980s large numbers of the state's population left West Virginia for places offering greater employment opportunities. The 1970s, however, marked a turning point in out-migration and the beginning of a stabilizing trend. During the 1980s population loss from the coalfields and heavy manufacturing was partially offset by an influx of urban professionals and retirees in the eastern panhandle. West Virginians have become determined to develop a more modern social and economic climate in their state.

The state's image

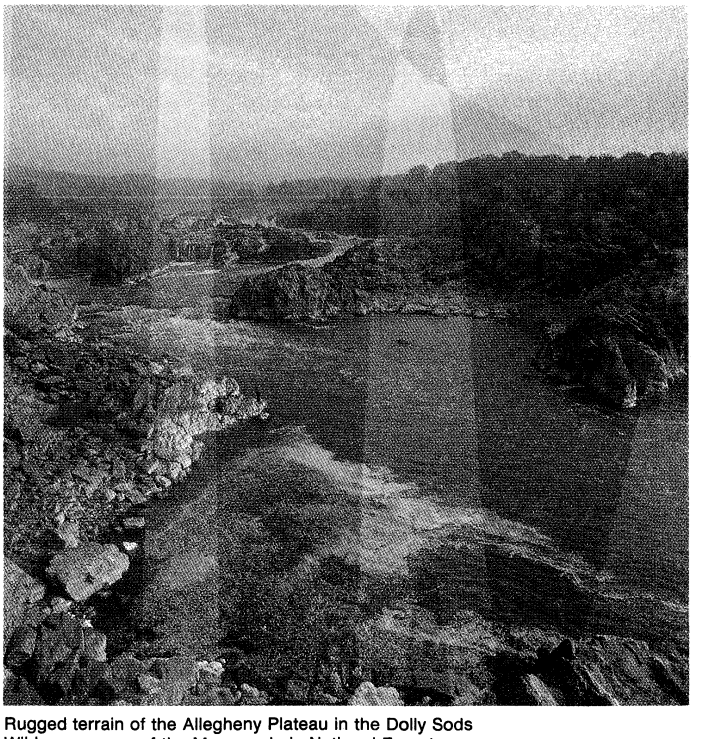
PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The maximum elevation in West Virginia is 4,861 feet (1,482 metres) at Spruce Knob in the east. The lowest land is 247 feet (75 metres) at Harpers Ferry, located on a steep tongue of land rising above the confluence of the Shenandoah and Potomac rivers. The land is rugged, ranging from hilly to mountainous, and there are no extensive expanses of level land. The state's two panhandles, one knifing northward between Pennsylvania and Ohio, the other eastward between Maryland and Virginia, provide West Virginia's alternate nickname, the Panhandle State.

All of West Virginia is a part of the Appalachian Mountain system. It is commonly subdivided, however, into two major physiographic regions: the geologically older Allegheny Plateau and the newer Appalachian, or Great Valley, region. These are separated by the Allegheny Front, dividing the waters that flow to the Atlantic Ocean from those flowing to the Gulf of Mexico. The Allegheny Plateau covers the western two-thirds of the state and coincides with the Ohio River drainage basin. It is a region severely dissected by streams into a maze of hills and valleys, and, in places, the original plateau surface shows as the uniform top levels of the remaining ranges. The northeastern portion of the Allegheny Plateau near the Pennsylvania border, with the highest mountains of the state, is sometimes referred to as the Allegheny Mountains. More than 40 of these mountains stand in excess of 4,000 feet, inducing heavy precipitation in the area and making it the wettest in the state and the source of many of its rivers. The eastern edge of the state and the eastern panhandle comprise the second region. This geologic province takes

Mountain regions

B. Thomas—H. Armstrong Roberts, Inc.



Rugged terrain of the Allegheny Plateau in the Dolly Sods Wilderness area of the Monongahela National Forest, West Virginia.

in most of the Potomac River basin and is famous for its northeast–southwest folded mountain alignment, part of the chain reaching from Canada to central Alabama. The northern end of the Blue Ridge Mountains forms a minor region in the easternmost end of the panhandle.

Drainage and soils. The pattern of drainage east of the Allegheny Front features a short, trellised drainage system flowing toward the northeast and ultimately draining into the Potomac. The western portions drain across an inclined plane by a longer system that flows generally northwest into the Ohio River. A very small area drains into the eastward-bound James River system of Virginia.

Drainage over West Virginia's rugged surface has created some of the most productive and level alluvial soils on the larger river floodplains. The weathered limestone soils of the east are suited for pasture and orchards. The acid, sandstone-derived soils of most of West Virginia foster a variety of agricultural activities. Some of the clay soils along the Ohio River are bases for the ceramics industry.

Climate. The state has distinct seasons of about equal length. It has a humid continental climate except for a marine modification in the lower panhandle. Mean annual temperatures reflecting latitude and altitude range from about 56° F (13° C) in the south to 52° F (11° C) in the north and 48° F (9° C) in the most mountainous regions. January is the coldest month, with a statewide average of 33° F (1° C), and July is warmest, with a 73° F (23° C) average. The growing season averages 160 days but ranges from 120 to 180 days. Mean annual precipitation varies from more than 60 inches (1,520 millimetres) in the mountainous areas to 35 inches in the eastern panhandle. Snowfall, which makes up some 8 percent of the total precipitation, varies from a seasonal average of less than 20 inches in the southwest to more than 64 inches in the northern mountains.

Plant and animal life. Forests cover about 75 percent of the state. Accessible virgin forests were harvested by the turn of the 20th century, but successional forests are now established except in agricultural and urban concentrations. The plateau forests consist of hardwoods of red and white oak, yellow poplar, sugar and red maple, hickory, beech, basswood, black cherry, and yellow birch. Softwoods of loblolly pine, shortleaf and white pine, spruce, and hemlock cover the mountain slopes, deep gorges, and other scattered areas. The eastern section is predominantly an oak and pine woodland. Other species such as the sycamore, locust, chestnut, elm, and dogwood are common.

Rabbits, squirrels, gray foxes, opossums, skunks, raccoon, and groundhogs are common in West Virginia. The larger white-tailed deer found in abundance by the settlers is numerically increasing in the remote and protected areas, and black bears are found in the isolated high country. Mountain streams feature trout, bass, pike, and muskellunge, while the improving water quality of the larger rivers accommodates increasing numbers of perch, bluegills, catfish, and other species.

Urban and rural patterns

Settlement patterns. Broad, level ridge tops and valley bottoms are commonly cleared for agriculture and living. The field patterns are usually linear, in conformity with the landscape. Rural dwellings are distributed as ribbons of settlement along the highways or near other transportation systems. Many rural dwellers commute to urban areas for employment because of the decline in agriculture and the mechanization of mining.

West Virginia has a number of cities with populations of more than 20,000. Of these, Huntington, Wheeling, Parkersburg, and Weirton are situated on the Ohio River, where water and rail transport and room for expansion have permitted growth. Most urban development and industrial growth extend along other streams, as in the Kanawha and Monongahela valleys. The larger cities, with their industrial concentrations, their political importance, and their colleges and universities, dominate the state's activities. County seats of government in the rural regions exert considerable influence over the areas they serve.

The people. The first pioneer settlement in what became West Virginia was by Germans along the Potomac River at Shepherdstown. Later German settlements were too scattered to establish patterns within the area. The 1790 census reported 55,873 inhabitants, of whom about 15,000 were of German descent. In the early 1800s many orders of the Virginia government relating to the frontier occupants were printed in both German and English. English descendants dominated the settlement of the Greenbrier, New, Kanawha, and Monongahela valleys, while Scotch-Irish tended to settle in the less accessible areas. Americans of African descent shared in this early heritage, although the number of slaves in western Virginia was limited since the rugged topography curtailed extensive agriculture. Only two counties, Jefferson and Kanawha, ever had more than 2,000 slaves, and in one-third of the counties slaves constituted less than 1 percent of the population. After the Civil War the development of railroads, mining, and industry attracted many blacks from the South as well as numerous labourers from southern Europe. The contrasting cultural influence of these more recent immigrants is apparent in the industrial northern panhandle and in the towns dominated by coal mining.

Later immigrants

The Quaker and Presbyterian denominations were established by the 1730s, with Baptists entering Berkeley county in the eastern panhandle in 1743. A Methodist circuit was established in Berkeley and Jefferson counties in 1778. Leading denominations in West Virginia are Methodist, Baptist, and Presbyterian. Roman Catholic and Lutheran adherents were prevalent in the early years of West Virginia, but they were limited primarily to the German settlements. Expansion of these faiths, particularly of Roman Catholicism, occurred with the immigration of Irish, German, Italian, Polish, and Hungarian labourers.

The economy. West Virginia has traditionally maintained a poor economic position among the states. A number of major factors caused the large migration out of the state during the 1950s and '60s. After 1950 mine mechanization and declining coal use contributed to a decreasing demand for labour. Rugged land and limited farm size hampered mechanization of agriculture, and the competitive advantage shifted to states with more level and expansive land. Foreign competition in the glass and ceramics industry also reduced economic opportunity. Some industries neglected to make progressive changes or could not create space for expansion, and the increasingly obsolete facilities hampered competition.

During the 1970s, resurgence of coal as a major energy resource and progressive government efforts to improve social and economic conditions in the region reversed the migration flow. The chemical, steel, and glass industries were modernized, and new high-technology industries were established in the Ohio and Kanawha valleys. These factors brought about migration within the state from poorer agricultural and mining regions to the urban areas, where better employment and educational opportunities exist.

In the 1980s, numerous jobs in the coal industry were lost, and the cyclical nature of mining created pockets of poverty in the state. Employment declined also in the manufacturing and chemical-processing industries. These job losses resulted in an out-migration among the young, with the oldest age groups tending to remain fixed. With this movement, the state began placing a greater emphasis on a more diverse economy, with a large service sector. The Governor's Office of Community and Industrial Development, established by the West Virginia Economic Development Act of 1985, has initiated successful action in three areas: strengthening the state's industrial base, directing marketing efforts to those concerns that will enjoy a comparative advantage by operating in the state, and creating an environment that encourages entrepreneurship and fosters the growth of small business.

Coal deposits

Resources. West Virginia has an abundance and variety of natural resources. The vital mineral is bituminous coal. Natural gas dominates the state's west central portion. Petroleum is extracted in the northern two-thirds of the gas-producing areas. Coal and extensive deposits of rock salt and brine have been supportive of a chemical industry. Abundant sand and clay distributions are basic to glass, tile, and brick production. Limestone is common in the eastern quarter.

The natural-resources development has influenced indus-

trial location and greatly affected the economy of certain areas. Industrial cities of the Ohio, Kanawha, and Monongahela valleys are dependent on local coal, limestone, clay, salt, glass sands, oil, and natural gas, as well as on the ready availability of water.

Agriculture and forestry. West Virginia has enormous stands of high-quality hardwood forests. Wisely used, this renewable resource is recognized as a safeguard against floods, slope erosion, and air pollution. It is significant to the state's tourist industry, to the thousands of workers in timber production, and to the manufacturing of finished wood products.

The rugged terrain of the state limits agriculture. Since the 1950s there has been a decline in acreage harvested and in the value of most crops per acre. But West Virginia ranks well within the upper one-third of the states in apple, peach, and tobacco production.

Industry. Since the late 1970s the nonmetallic mining industry has been in decline, and only large operators have prospered. Severe cutbacks in steel and other heavy industries have been moderated by state aid and concessions from workers' unions. Thermal electricity plants have expanded but at a slower rate because of the drift from energy-intensive heavy manufacturing to light manufacturing and the service sector. With a renewed emphasis on tourism, related services have prospered.

Transportation. The larger cities and the state's perimeters are well served by transportation facilities. The rugged terrain of West Virginia limited early transportation and contributed to isolation and slow economic growth. It is still a formidable obstacle, but remarkable progress has been made. Interstate highways that cross the state have improved internal travel and economic development. The Appalachian Highway Corridor roads have been instrumental in completing the network of other federal and state roads. The major river systems of the western plateau provide 450 miles (725 kilometres) of navigable waterway. Huntington, located at the confluence of the Ohio and Guyandotte rivers, is the state's largest port. Airline deregulation has led to an overall reduction in scheduled air carrier service. Major airlines serve Charleston and Huntington, and commuter lines have helped fill the void. Three rail companies with a strong east–west orientation provide haulage over the 3,500 miles of line.

Administration and social conditions. *Government.* West Virginia's present constitution dates from 1872, but it has been amended many times. The governor is assisted by an elected secretary of state, auditor, treasurer, attorney general, and agriculture commissioner, in addition to a superintendent of schools appointed by the state board of education. The legislature consists of the Senate and the House of Delegates. Two senators are elected from each of the 17 senatorial districts; their four-year terms are so arranged that one of them must stand for reelection every two years. The House of Delegates is composed of 100 members, each serving a two-year term. Their representation is apportioned over 40 voting districts on the basis of population. The legislature meets annually for 60 days, but the governor is empowered to convene special sessions. At the head of the judiciary division is the Supreme Court of Appeals, consisting of five judges elected to 12-year terms. The heavily burdened 31 circuit courts try the major cases. Magistrate courts serve at the county level, and municipal courts are provided for in incorporated areas.

The Department of Public Safety provides law enforcement at the state level and maintains bureaus of criminal identification, motor-vehicle inspection, and accident prevention.

County and municipal government

County law enforcement is handled by a sheriff, sheriff's deputies, and a prosecuting attorney working with county and municipal police. Under the state constitution, each county is governed by three commissioners elected for six years. Other elected county officials are sheriff, prosecuting attorney, assessor, members of the board of education, and surveyor. Some towns have a mayor–council form of government, others a council–manager form. All officials except the city manager are elected.

Education. The public school system, directed by the state board of education, is under a county-unit plan allowing local supervision under county boards. The state-controlled system of higher education includes West Virginia University at Morgantown and Marshall University at Huntington. There are, in addition, several state colleges, community colleges, specialized technical institutions, and a number of private and sectarian colleges. The Vocational, Technical and Adult Education Division provides an important educational supplement with courses in agriculture, home economics, practical nursing, and technical, industrial, and distributive skills. The National Radio Astronomy Observatory at Green Bank is a unique educational and research facility. A national centre for the reception and study of radio waves that come from space, it is open to all users and has undergone expansion.

Health and welfare. The State Department of Health has responsibility ranging from collecting statistics on various diseases and health-control matters to overseeing all matters relating to public health and medical facilities. County and municipal boards of health assist at the local level. The state also oversees public assistance programs in such areas as child welfare, distribution of commodities and food stamps, and matters regarding old age, blindness, dependent children, disability, and general relief. There are hospitals for the mentally ill and numerous programs for mentally handicapped children and adults.

Cultural life. The early isolation of West Virginia resulted in the development and transmittal of strong, self-reliant local heritage relatively unaffected by circumstances from beyond the hills. Musical instruments, ballads, and handicrafts of earlier generations are still in evidence. Numerous fairs, craft centres, and collectors have assured the permanence of this cultural past. The making and playing of dulcimers, old-time fiddle contests, ballad singing, patchwork quilting, furniture caning, and primitive drawing, along with Elizabethan speech patterns and other remnants from the past, persist in the rural regions.

Traditions of folk arts and craft

In addition to the folk arts and crafts, the Huntington Museum of Art and the Sunrise Museums at Charleston have excellent cultural facilities. In local art exhibitions picturesque landscapes and townscapes of the state are shown in a majority of works, but often they indicate awareness of advanced artistic styles. The universities and colleges are cultural centres as well, fostering work in the visual arts, theatre, dance, and music.

A thriving tourist and recreation industry has developed around West Virginia's cultural heritage and its various historical and natural resources. Harpers Ferry, for decades almost a ghost town, is a national historical site. Ski slopes attract winter visitors. Plunging streams and river sites feature white-water activities and aquatic festivals. There are 33 state parks, including Pipestem State Park, in which an amphitheatre, a country store, and a craft centre are typical of installations to promote the state's heritage and to encourage the tourist industry. The resort at White Sulphur Springs is famous throughout the world.

HISTORY

Some 14,000 years ago Indian hunters entered the Ohio and Kanawha valleys in pursuit of mammoths. Around 9000 BC the Archaic people, with a small-game hunting, fishing, and gathering culture, occupied the area. Their successors, the Adena, or Mound Builders (*c.* 500 BC to *c.* AD 100), created numerous earthworks still visible in the Moundsville and Charleston areas. The Adena were absorbed by the Fort Ancient people, who dominated the territory until they were wiped out by the Iroquois League around 1650. Except for scattered villages the area that was to become West Virginia remained Indian hunting grounds and battlegrounds when Europeans arrived in the 1700s.

Colonial period and Virginia's dominion. The second charter of Virginia in 1609 provided for settlement of that colony's western frontiers. Exploration and trade were further encouraged by Governor William Berkeley after 1660. The Blue Ridge was reached in 1670, and in 1671 another expedition encountered the first westward-flowing stream, the New River, in southwestern Virginia. The expedition descended the river to Peter's Falls on the future Virginia–West Virginia border and claimed for England all the land drained by the New River and its tributaries. Subsequent

Trans-Allegheny frontier settlement was handicapped by such factors as mountain barriers, Indian resistance, conflicting English and French claims in the Ohio valley, disputed land titles, and a royal proclamation of 1763 prohibiting occupancy.

Separatist sentiment

Despite these obstacles, the population increased, and discontent with the government east of the mountains became endemic. The Vandalia colony, proposed in 1769, and the "14th State" movement for the establishment of a Westsylvania in 1776, indicate an early interest in a separate government for the Trans-Allegheny country. Dissatisfaction among the pioneers in that region mounted in the cultural, social, economic, and political realms. The frontier residents, who came from many areas, were distinctly different from the aristocratic eastern settlers. Furthermore, topographic differences rendered slavery economically unsound, and cultural heritage made it undesirable. Voting representation and taxation, however, decidedly favoured eastern Virginia.

Civil War and statehood. The advent of the American Civil War fueled new desires for a politically separate western area. At the Virginia secession convention of April 1861 a majority of the western delegates opposed secession. In subsequent meetings at Wheeling (May 1861), dominated by the western delegates, the ordinance of secession was declared an illegal attempt to overthrow the federal government. The second Wheeling convention (June) pronounced the Richmond government void, established a restored Government of Virginia, and provided for the election of new state officers. The restored governor, Francis H. Pierpont, secured federal recognition and maintained civil jurisdiction over the region until Congress consented to the admission of West Virginia to the Union on June 20, 1863. A condition of entry was the gradual emancipation of slaves in the region. The capital was permanently established at Charleston in 1885.

Civil War engagements were few in the state, although the war itself was in part precipitated by the seizure of the federal armory at Harpers Ferry in 1859 by a small band of men under the antislavery zeal of John Brown. Brown was captured by federal troops and subsequently was tried and hanged in Charles Town, but his exploits inflamed tensions between the nation's proslavery and antislavery factions. To the abolitionists of the North he became a martyr. During the Civil War nearly 32,000 soldiers enlisted in the Union army, and about 9,000 served the Confederacy, although some authorities maintain the latter figure to be low.

Industrial development

The modern period. West Virginia's industrial emergence, encouraged by railroad expansion, began in the 1870s. Its natural resources of timber, coal, salt, oil, and gas substantially contributed to the establishment of the modern industrial system of the United States. The labour troubles that flared in mining areas between 1912 and 1921 required the intervention of the National Guard (twice) and the U.S. Army (four times) to quell violence, but the right to organize labour unions, which was granted by national statutes in 1933 and 1935, brought a measure of peace to the state. West Virginia was one of the leading states in the percentage of its population serving in World War II, Korea, and Vietnam. The state received national political recognition in 1976 and 1980 by the election of a New York City native, John D. Rockefeller IV, to the governor's office. In 1984 he was elected to the U.S. Senate. Neither national political party has dominated the state for long periods, although the Democrats have tended to outnumber Republicans and usually had more than twice the number of registered voters. (S.E.C.)

THE MIDWEST

Illinois

Admitted as the 21st member of the Union on Dec. 3, 1818, Illinois has throughout the 20th century been profoundly divided. It lies within both the so-called old industrial belt and the fertile agricultural heart of the nation. The presence of Chicago, the nation's third largest city, creates sharp distinctions between the state's largely urban northeast and the more evenly balanced urban–rural population downstate. Because of its great length, Illinois exhibits both Northern and Southern regional characteristics. Still further contrasts derive from the racial and ethnic complexity of the population.

These internal divisions, while not unique to Illinois, perhaps became magnified through the state's critical role in the economic and political life of the nation. Rich in coal and oil reserves and ideally located for the acquisition of raw materials and distribution of finished goods, Illinois ranks among the top states in value of exports, agricultural income, and value added by manufacturing.

Transportation centre

Chicago is a railroad hub of the nation; its O'Hare International Airport is among the world's busiest, and Illinois highways and waterways are thick with commercial traffic. Politically, Illinois has continued to be a "swing state," its votes often mirroring fluctuating social tensions that underlie the growing, but unevenly distributed, economic prosperity.

Illinois encompasses 56,345 square miles (145,934 square kilometres) and stretches 385 miles (620 kilometres) from Wisconsin in the north to Cairo in the area that is known as "Little Egypt," which lies farther south than Richmond, Va. In addition to Wisconsin, the state borders Lake Michigan on the northeast, Indiana on the east, Kentucky on the southeast, and Missouri and Iowa on the west. Illinois was named for the Illinois Indians. The capital is Springfield.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief and drainage.* Illinois is drained by as many as 900 streams emptying mostly into the Mississippi River system. The Chicago and Calumet rivers, originally flowing into the St. Lawrence by way of Lake Michigan, have been altered through the construction of canals to drain into the Mississippi by way of the Illinois River. The Ohio River joins the Mississippi at the state's southern tip.

Flat prairies cover much of Illinois, with irregular plains in the western, northern, and southern sections. The unglaciated southernmost part of the state is in many ways out of character with the rest of Illinois. Shawnee National Forest, the only federal forest in Illinois, covers a great part of this region. Southern Illinois consists of gently sloping, open hills. Rolling hills in the northwest include the state's highest point, 1,235 feet (376 metres) above sea level. The statewide average elevation is about 600 feet. Water lies under all of Illinois in natural underground reservoirs. Chicago and many of its suburbs draw their water from Lake Michigan, but most of northern Illinois's water is pumped from underground wells. Some regions face a dwindling water supply; around Joliet the water table has been lowered more than 700 feet in the past 100 years. Most of the state's lakes are man-made.

The deep black soil of much of northern and central Illinois has unusual richness, and its quality for agriculture is among the finest in the world. The soils of the southern third of the state are far less suited for farming.

Climate. Because of its geographic position and its long north–south axis, Illinois experiences wide seasonal and regional variations in temperature, with typically cold, snowy winters and hot summers; extremes are somewhat ameliorated around Lake Michigan. Mean winter temperatures are about 22° F (−6° C) in the north and 37° F (3° C) in the south; summer equivalents are 74° F (23° C) and 80° F (27° C). Mean annual precipitation in the north is about 34 inches (864 millimetres) and in the south, 46 inches. The growing season varies from 205 days in the south to 155 days in the northernmost counties.

Plant and animal life. Illinois vegetation is separated into the Fayette prairie of northern and central Illinois and the oak–hickory forest of the western and southern regions. Before white settlers moved in, oak–hickory forests

prevailed also in the north. The settlers' needs for fuel and construction material and the lumbering industry stripped most of the trees, leaving only 10 percent of the forests in Illinois. Some 4,000,000 acres of forests remain, more than 700,000 of them in the Shawnee National Forest. The state's length gives it an unusual variety of Northern and Southern plant life. Both Northern and Southern wildflowers grow in Illinois, as do a variety of trees, such as white pine, tamarack, walnut, cypress, and tupelo gum.

Before 1800 wildlife in abundance roamed the prairies and forests, but the bison, bear, wolf, mountain lion, and elk have disappeared. Deer became extinct in 1910, but in 1933 the Department of Conservation placed small herds that resulted in a growing deer population. Game birds such as quail and pheasant are not as plentiful as in previous years, but waterfowl are abundant during spring and fall migration. Pollution has nearly wiped out many species of fish, but bullheads, carp, and catfish still abound.

Settlement patterns. Aside from the aforementioned distinction between the Chicago area and downstate derived from population patterns, Illinois can be separated into three broad regions that differ markedly in their economic and social characteristics. A highly urbanized band—with extensive farming areas in between—reaches across the state in the north from Chicago to the Rock Island–Moline complex on the Mississippi and includes Kankakee, Joliet, and Rockford. Most of the farmland, although decidedly rural, has easy access to a sizable urban centre. The region is characterized by heavy industry around Chicago and the other large centres with a large and rapidly expanding suburban complex of shopping centres, single-family dwellings, and apartment houses.

Economic and social regions

The central third of the state includes the cities of Springfield, Bloomington, Peoria, Champaign, Danville, Galesburg, Quincy, and Decatur. The economic base of the region is agriculture. Some cities (notably Peoria) support such industries as the manufacture of farm machinery and construction equipment; others are centred on institutions such as the state government complex in Springfield and the University of Illinois in Champaign–Urbana. The focus of communications and transportation is scattered among four or five metropolitan areas. The character of the people tends to remain rural or small-town, with a highly developed sense of tradition and history.

St. Louis, Mo., dominates southern Illinois. East St. Louis, Belleville, Alton, and Granite City are medium-sized cities, but they are located in only two counties, leaving the rest mostly rural. Because this region of Illinois was settled earliest, most of its communities have longer historical traditions than do their northern neighbours. Southern Illinois has coal mines, oil wells, and the Shawnee National Forest, which covers parts of 10 counties. The region is Southern mountain in character, and the pace of living is slower because of fewer cities and a depressed economy. Southern Illinois University in Carbondale has provided economic and cultural stimulation to the region.

The people. After most of the French had left Illinois following the French and Indian War, English settlers and colonists of Anglo-Saxon stock from Virginia, Tennessee, and Kentucky moved in. New Englanders and New Yorkers arrived by way of the Great Lakes or the National Road. When the United States experienced the great waves of European immigrants beginning in the 1840s, large numbers of Germans and Irish went to northern Illinois; and, from the 1880s until World War I, immigrants came from Poland, Hungary, Italy, Norway, Sweden, Austria, and Russia. In 1910 Germans ranked first in number among the foreign-born and Austrians and Hungarians second, followed by Russians, Scandinavians, Irish, and Italians. Although the Jewish population had become well established in the 19th century, its numbers swelled prior to World War II, particularly in the Chicago area. Following the war, whites from Appalachia and American Indians sought employment in the cities. More-recent immigrants have tended to come from diverse origins, with Asians and Latin Americans constituting increasingly influential sectors of the population.

The great ethnic mix

Blacks have lived in Illinois since the first slaves were imported in 1719, but their numbers remained low until the Civil War. In 1870, blacks numbered 29,000, and, by 1910, migrant blacks had settled in the southern counties and totaled 109,000. With World War I there began a steady flow of blacks to the major industrial centres, and by the late 20th century blacks in Illinois made up about 15 percent of the state's population, four-fifths of them in Chicago and Cook county.

The religious diversity in Illinois is reflected in the different origins of the people themselves. In the early 1800s Methodist circuit riders nourished tiny congregations throughout Illinois, and Methodists remain strong today. The Irish, some Germans, and later southern and eastern Europeans brought the Roman Catholic faith to the larger cities, and the Roman Catholic archdiocese of Chicago is the nation's second largest in membership. Also serving the city are scores of Eastern Orthodox and Protestant churches and Jewish synagogues.

Illinois has population characteristics similar to those of the nation as a whole: cities continue to lose whites to suburban areas, while both the number and percentage of blacks within the larger cities have increased.

The economy. The diversified nature of its economy—strength in manufacturing, agriculture, finance, mining, transportation, government, and services—makes Illinois a microcosm of the national economy. This diversity generally provides greater stability at times when other states with more narrowly based industries suffer.

The state and private business organizations give considerable attention to expanding Illinois's balanced economy. The Illinois Department of Commerce and Community Affairs has offices in foreign cities to stimulate the importation of Illinois products. The state offers services for the development of business enterprises by blacks and other ethnic minorities and disseminates information to private enterprises on new technological developments. Private organizations have played a significant part in attracting industry, in the development or rehabilitation of downtown areas, and in technological advancement.

Trade unions are strong in Illinois, both politically and economically, but neither they nor employer groups are strong enough to impose their will on the other. Mediation in labour–management disputes by politicians is a frequent occurrence, notably in Chicago, where union ties to political parties and leaders are accepted facts of life.

Resources. Illinois ranks second in the United States in reserves of coal overall and first in reserves of bituminous coal. Although Illinois is not the nation's leading coal producer, coal is the state's leading mineral commodity in value. Peak petroleum production was reached in 1940 and has declined since. Illinois leads in the mining of fluorspar, tripoli, and industrial sand.

Coal and oil dominate in the production of electrical power, but atomic energy plants have assumed increasing shares of the state's needs. Illinois has 12 such plants, more than any other state. The Argonne National Laboratory, near Lemont, and the Fermi National Accelerator Laboratory, in Batavia, are major research and development installations of the U.S. Department of Energy.

Developments in nuclear power

Industry. Illinois ranks high among the states in the manufacture of fabricated metals, food products, rubber products, and electrical machinery, as well as in its number of printing and publishing establishments. It is also among the top producers of nonelectrical machinery, which accounts for the largest share of its foreign exports. Industrial parks are scattered throughout the state, the greatest concentration being in the Chicago metropolitan area.

Agriculture. Illinois's greatest natural resource is its rich, black soil. Farms cover more than 80 percent of the state's area. Long the nation's major soybean producer, Illinois from year to year trades places with Iowa for first-place rank in corn. Illinois is also noted for pork and dairy products, grains, and meat animals. Family-owned farms account for the greatest percentage of farms in Illinois.

Finance. A state prohibition against branch banking has caused a proliferation of independent banks in Illinois. This issue long has produced complex political battles in Springfield, with frequent charges and occasional exposures of political graft connected with it. Observers see opposition to change coming primarily from small banks

Banks and other financial institutions

Cornfield on a farm near Macon in central Illinois.
Grant Heilman/Grant Heilman Photography

downstate that fear elimination by the huge Chicago banks and from the currency exchanges that operate with high service charges, especially in inner-city areas that cannot support an independent banking institution.

In addition to its banking strength, Illinois is a major insurance centre. Chicago is the seat of the seventh district of the Federal Reserve Bank as well as of the Midwest Stock Exchange and the Chicago Board of Trade. Although well below the New York and American exchanges in volume, the Midwest Stock Exchange experiences considerable trading. The Board of Trade is the nation's oldest and largest commodity market, dealing in contracts for grains, soybeans and their products, silver, plywood and lumber, livestock, and dairy products. Other exchanges include the Chicago Mercantile Exchange.

Transportation. Illinois is known as the transportation centre of the United States. Few comparable areas are served by so many means of transportation. The state's rail network, radiating from Chicago, is among the most extensive in the nation. Chicago is a major centre for Amtrak service and also maintains an extensive commuter rail service. Chicago is also one of the principal hubs of the national road system, and nearly all of the state's major urban areas are served by at least one interstate highway.

Water transportation became more efficient when Lake Michigan was connected to the Mississippi River in 1848 by means of the Illinois and Michigan Canal, linking the Chicago and Illinois rivers. The St. Lawrence Seaway stimulated the expansion of the Port of Chicago. Oceangoing freighters dock at Calumet Harbor in South Chicago. Chicago has two major airports, O'Hare International and Midway, and a third on the lakefront, Meigs Field, serving small planes. More than 700 airports are located throughout the state.

Administration and social conditions. *Government.* The first state constitution was adopted in 1818, a second in 1848, and a third in 1870, which was to remain in effect for 100 years. The constitution adopted in December 1970 added new concepts to the Illinois Bill of Rights, forbidding discrimination on the basis of race, creed, colour, national ancestry, or sex in employment or in the sale or rental of property and discrimination against the physically or mentally handicapped.

Major provisions of the 1970 constitution

The executive branch consists of the governor, lieutenant governor, secretary of state, attorney general, treasurer, and comptroller, all elected to unlimited four-year terms. Under the new constitution the governor was granted powers to reorganize state government, and the governor's veto power was augmented by the authority to reduce appropriations and to object to certain portions of legislation without having to veto an entire act.

Illinois's General Assembly comprises a 59-member Senate and a 118-member House of Representatives. Senators serve four-year terms and representatives two-year terms. In 1980 a referendum abolished the unique electoral framework of the House of Representatives, under which each district elected three representatives, with the district's minority party virtually assured of one seat; representatives are now elected from single-member districts.

The judiciary is headed by the seven-member Supreme Court, whose justices serve 10-year terms. There are also Appellate Courts, Circuit Courts, and the Court of Claims, which hears claims involving the state. Although widespread support has been voiced for adoption of the Missouri Plan, under which judicial candidates and incumbent judges are reviewed by nonpartisan boards and then appointed by the governor, the partisan election of judges continues. Judges run for retention on their record, the voter designating "yes" or "no."

Illinois has three levels of local government—county, township, and municipality—plus many special districts. Counties are classed as township and nontownship, with Cook county, containing Chicago and most of its major suburbs, in a class by itself. All counties elect a number of administrative officials. The 85 township counties are governed by boards of commissioners elected from districts or at large, the 17 nontownship counties by three-member boards elected at large, and Cook county by a 17-member board, 11 from Chicago and six from the suburbs.

Townships act primarily as road-maintenance and general-assistance units. The annual town meeting, a gathering of all qualified voters, is still a feature of local government remaining from earlier centuries. Municipal government usually is of the mayor–council type, though other forms are permitted; villages utilize a president–trustee system.

Overall, Illinois has more than 6,400 units of local government, resulting in overlapping administrative, educational, park, fire, sanitary, sewage, drainage, and other special districts. Most were formed to circumvent restrictions in the old constitution but have become self-perpetuating. Patterns of taxation are similarly mixed. State taxes include those on personal and business income, cigarettes, liquor, and retail sales. Real-estate taxes contribute the major local support for schools and other services, though the state supplies "no-strings-attached" grants to municipalities and counties from income-tax revenues.

Education. The Illinois State Board of Education was created in 1975 after the 1970 constitution directed that responsibilities for public elementary and secondary education be transferred from an elected superintendent to an appointed board of citizens. The board's 17 members are appointed by the governor with the consent of the Senate. The issue of funding Chicago schools long has been a bitter point of city–state contention.

In the field of higher education, Illinois offers a wide array of opportunity and is often cited for academic excellence. The University of Chicago (chartered 1890) is respected as among the finest institutions of higher learning in the nation. Northwestern University (1851), in Evanston, has a distinguished faculty in several areas, as do the Illinois Institute of Technology, in Chicago, and Southern Illinois University, in Carbondale and Edwardsville. In addition to the last-named institution, the state system includes eight state universities as well as the University of Illinois, with campuses at Champaign–Urbana and Chicago, and more than 50 community colleges. Sectarian and nonaffiliated private colleges are scattered throughout the state.

Institutions of higher learning

Health and welfare. Health and welfare assistance and services are provided by both state and local government, most of which receive funds from diverse federal agencies, although representatives of the low-income groups most requiring the services have been in frequent conflict with local officials over the question of who is to administer the funds and the programs.

Chicago's public housing is primarily a municipal re-

sponsibility, with federal aid. Both the former Model Cities Program, funded by the federal government, and the Chicago Housing Authority have met with opposition to proposals to locate public housing in predominantly white neighbourhoods. Welfare assistance to poor families, dependent children, and other groups has met resistance in the legislature in Springfield, whereas independent attempts to establish free neighbourhood medical clinics in low-income black and Spanish-speaking areas of cities have encountered local hostility and lack of adequate funding.

Thus, although medical facilities throughout most of Illinois are among the finest in the nation, and Chicago is a centre for medical and psychiatric services and training, many Illinoisans are served inadequately. And, despite the fact that income per capita in Illinois is among the highest in the nation, the burdens of public assistance seem far from solution or even agreement on goals.

Effects of Chicago

Cultural life. Inevitably the influence of a major artistic centre tends to dominate the cultural reputation of Illinois. By 1900, Chicago architects were designing commercial and private buildings that became models for modern architecture throughout the world. In the 1910s and '20s the city was a hub of literary activity, and today the holdings of its public and private institutional libraries are enormous. Its Art Institute, Museum of Science and Industry, Field Museum of Natural History, and other civic landmarks have collections and research facilities among the most complete in the world. Before the development of Hollywood around 1910, Chicago was the centre of the American film industry, and today its theatrical community offers a broad spectrum of standard and avant-garde works. Dance is widely available, and its symphony orchestra and opera company are among the premier American musical organizations.

Communities outside the Chicago area have thriving cultural lives as well, often revolving around the theatre, music, art, or various science departments of the many colleges and universities or around community theatre or musical organizations. Belleville boasts the second oldest symphony orchestra in the nation, founded in 1867. The Eagle's Nest Art Colony, founded in Oregon in 1898 by the sculptor Lorado Taft, included many well-known Illinois artists; it was acquired by Northern Illinois University in 1950. The Illinois Arts Council was created by the state in 1965 as the primary agency to fund statewide or local programs in the several arts. It is supported by the state and the National Endowment for the Arts, a federal agency.

Points of historical importance dot the state. Among old cities on the Mississippi are Galena, which preserves the home of President Ulysses S. Grant, and Nauvoo, founded in 1839 by the Mormons and their point of departure in 1846 on the trek that took them to Utah. New Salem, near Springfield, is a preservation of the community of log cabins in which Abraham Lincoln spent much of his young manhood. Throughout central Illinois the Lincoln Trail joins places associated with the president, including his home in Springfield and the sites of his debates with Senator Stephen A. Douglas. The Spoon River Trail in north central Illinois leads through the country made famous by the poet Edgar Lee Masters. Scenic areas include the Mississippi Palisades State Park and Apple River Canyon State Park in the northwest, Starved Rock State Park in north central Illinois, and the forests of the south. Oak Park, home of the pioneer modern architect Frank Lloyd Wright, contains much of his early work.

Among Illinois's finest recreational offerings are the sandy beaches of Lake Michigan, from Chicago to the Wisconsin border, and the forest preserves. Although Illinois has virtually no wilderness areas, many camping sites are located throughout the state, and boating and fishing are avidly pursued on the state's many lakes and streams.

Newspapers and the broadcast media

About 70 daily and 700 weekly newspapers are published throughout Illinois. The largest of these papers, the *Chicago Tribune,* became a nationally recognized symbol of the political and social conservatism of the Midwest under the long reign of publisher Robert R. McCormick, and it continues to have wide distribution throughout the Midwest. Its point of view is offset somewhat by the *Chicago Sun-Times,* the second largest daily in Chicago. The *Chicago Daily Defender* is published primarily for the black community. Southern Illinois is influenced also by newspapers and broadcasts from St. Louis.

Chicago is the third largest publishing centre in the nation, exceeded only by New York City and San Francisco. Much of its publishing is specialized in the areas of education, encyclopaedias, medicine, and business.

HISTORY

Archaeologists have found evidence dating from around 8000 BC of a Paleo-Indian culture in southern Illinois. The Mississippian people, whose religious centre was Cahokia in southwestern Illinois, constituted probably the largest pre-Columbian (around AD 1300) community north of Mexico in the Mississippi floodplain. Indian tribes in Illinois were all of the Algonquian stock. The Kickapoo, Sac, and Fox roamed in the north; the Potawatomi, Ottawa, and Ojibwa (Chippewa) dominated the Lake Michigan area; the Kaskaska, Illinois, and Peoria stalked the central prairies; and the Cahokia and Tamaroa roamed the south.

Settlement. The first Europeans to visit Illinois were the French explorers Louis Jolliet and Jacques Marquette in 1673, when they explored the Mississippi and Illinois rivers. Near present-day Peoria, Lasalle established the first French foothold, Fort-Crèvecoeur, and built Fort-Saint-Louis near Ottawa. After the French and Indian War in the 1760s, France ceded to Britain its claim to lands east of the Mississippi. The following years were uneasy—British policy was unfavourable to the area's economic development, Indians resented the British, and settlements were without civil government. By 1773 the number of settlers had declined to about 1,000 plus a few hundred slaves.

In 1778 the American capture of Kaskaskia, the British seat of government, made Illinois a county of Virginia. The first settlement on the site of Chicago was made in 1779 by the black pioneer Jean-Baptist-Point Du Sable. On July 4, 1800, the Northwest Territory was divided, and the Illinois country was made a part of Indiana Territory; Illinois Territory was formed in 1809 by dividing Indiana Territory, and Illinois attained statehood nine years later.

Early years of statehood. In 1818, two-thirds of the population lived along the eastern and western edges of southern Illinois and primarily engaged in the fur trade. The final conflict with the Indians was the Black Hawk War in 1832.

Southern and central Illinois remained the more heavily settled areas of the state during the early 19th century. In 1848 the Illinois and Michigan Canal was completed, linking points on the Illinois River on opposite sides of the Mississippi–Lake Michigan watersheds. With rail expansion many towns became prosperous. The National Road, leading westward from Maryland and terminating at Vandalia, brought many settlers to Illinois.

Slavery as a constitutional issue

The Illinois Constitution of 1818 gave blacks the status of indentured servants, and slavery would have been legalized except for fear that such a move would prevent admission. In 1824 Illinois voters rejected a proposal for a constitutional convention whose implicit purpose was to legalize slavery. Following the heavy influx of Yankees into northern Illinois during the 1830s and '40s, which offset the Southern attitudes, abolitionist sentiment translated itself into the Constitution of 1848, which abolished slavery and forbade the importing of slaves.

When the Civil War broke out, northern Illinois remained loyal to the Union and to the Illinoisan in the White House, Abraham Lincoln. A movement to ally southern Illinois with the Confederacy failed. Some 250,000 Illinoisans fought for the Union; among them was its most able general and a future president, Ulysses S. Grant.

Economic and social maturation. Chicago's great fire of 1871 was only a temporary deterrent in the city's progress toward becoming an industrial colossus. Its mills, rail yards, and slaughterhouses were filled with both European immigrants and freed blacks who had come to Illinois beginning in the 1860s. Until well into the 20th century, Illinois was a main focus of the American labour movement. Two events in Chicago, the Haymarket Riot of 1886 and the Pullman strike of 1894, became landmarks in the militant rise of the unions.

At the same time, Illinois was becoming a pioneer in social legislation, with a state board of health created in 1877; a compulsory school-attendance law in 1883; a "sweatshop act" providing for factory inspections and restrictions on child labour; and an eight-hour-day, 48-hour-week work limit for children, both enacted in 1893. The World's Columbian Exposition of 1893 was America's first international exhibition of the nation's vast technological and scientific strides during the 19th century.

The 20th century. During the decades up to and including the 1920s and '30s, the name Chicago became an international byword for bootleg liquor, gangsterism, and syndicate crime. Downstate Illinois was also notorious as a region of violence. "Bloody Williamson" county was the site of a feud, beginning in 1868, among five families of Tennessee and Kentucky origin. A dispute over a card game in a tavern near Carbondale grew into an eight-year vendetta fought by ambush or nighttime murder in barnyards, bars, and country stores. This violent tradition continued into the 1920s with the antiblack crusades of the Ku Klux Klan, the coal strikes, and the wars among the Shelton, Birger, and other bootleg gangs.

"Bloody Williamson" county

Amid the violence and scandals that rocked state and municipal governments in Illinois, there was tremendous economic and cultural growth. A reorganization of state government in 1917 brought more than 100 independent agencies and commissions under the governor and became a model for many other states. Chicago became America's second largest city in the 1880s, and in 1933–34 its Century of Progress Exposition drew attention again to further industrial achievement. In 1942 the world's first controlled atomic chain reaction was set off at the University of Chicago, ushering in the atomic age.

Since the Civil War, intense competition between the Republican and Democratic parties has characterized the political life of Illinois. This factor and the state's large electoral vote make it a major battleground in presidential elections. The three distinguishable political regions are Chicago, which is heavily Democratic; Chicago's suburban metropolitan area and the rich farmlands of north and central Illinois, which are strongly Republican; and southern Illinois, which may swing one way or the other. Today, the two parties have almost equal strength statewide.

Both the Democrats and the Republicans in Illinois are highly organized, from precinct to state levels. During the state's history both parties have been so frequently the target of corruption and fraud that Illinois politics has gained a checkered national reputation. From 1955 until his death in 1976, Mayor Richard J. Daley of Chicago built up enormous statewide—and nationwide—power in the Democratic Party, largely through his administrative control of all city and, effectively, Cook county departments and their patronage. If the governor is Republican, he almost always leads the party but is ordinarily unable to command the huge bloc vote possible in Chicago.

James R. Thompson, a Republican from Chicago, was first elected governor in 1976 and was reelected for four consecutive terms, a record in the history of the state. During most of that period he was faced with a Democratic-controlled House and Senate. As a result, Thompson used his extensive veto powers—including total veto, line item veto, appropriation-reduction veto, and the amendatory veto—to influence legislation. (R.T.L./J.M.Ca./V.D.S.)

Indiana

Indiana, though officially an eastern north central state of the United States, is perhaps the most Southern in character of all Northern states. This is largely a reflection of the early settlement of the region by immigrants from the Southern hills, who brought slavery and a hearty distrust of the federal government. Indiana's 36,185 square miles (93,720 square kilometres) make it, except for Hawaii, the smallest state west of the Appalachian Mountains. The capital has been at Indianapolis since 1825, nine years after Indiana, the name generally thought to mean "land of the Indians," was admitted on Dec. 11, 1816, as the 19th state of the Union.

Indiana is a manufacturing state; its northern half lies in the mainstream of the industrial belt stretching from Pennsylvania and New York to Illinois. Many of its people, nevertheless, continue to cherish an image derived from 19th-century America: largely white, dedicated to the Protestant ethic of sobriety and hard work, oriented to the small town and medium-sized city, and interested in maintaining the prerogatives of local self-determination. It is not by coincidence that Indiana's federal aid is one of the lowest per capita of any American state or that the Indianan's nickname, the Hoosier, remains a symbol in the nation's lore for a kind of homespun wisdom, wit, and folksiness that harkens back to what is popularly regarded as a less-hurried and less-sophisticated period of history.

An overview of the state

Indiana sits, as its motto states, at "the crossroads of America." It borders Lake Michigan and the state of Michigan on the north, Ohio on the east, Kentucky on the south, and Illinois on the west, making it an integral part of the Midwest. Its northwestern cities form an industrial, economic, and social continuum with neighbouring Chicago. Their heavy black populations and black political aspirations contrast strikingly with life in the cities and towns on the Ohio River. The state is at once Northern and black, Southern and white-dominated, with all the problems attendant on both circumstances. Though generally considered a conservative and Republican stronghold, Indiana has voted into both state and national office an almost equal number of liberals and Democrats.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief and drainage.* Indiana forms part of the east central lowlands that slope downward from the Appalachian Mountains to the Mississippi River. Approximately five-sixths of its surface was modified by glacial action, leaving a vast quantity of excellent soil material and extensive deposits of sand and gravel. The more eroded southern part of the state gives way to the central plain, an extremely fertile agricultural belt with large farms, and then to the mostly flat glacial lake basin and moraine region of northern Indiana. The highest elevation is along the Ohio border, at 1,257 feet (383 metres) above sea level, while the low point, 320 feet (98 metres), is in the southwest where the Wabash River enters the Ohio. About 90 percent of the land lies between 500 and 1,000 feet.

Natural environment

The general slope and drainage pattern is toward the southwest, though an almost imperceptible groundswell in the northeast forms a St. Lawrence–Mississippi water divide. The Wabash, the Ohio, and the east and west forks of the White River follow this slope, forming part of the Mississippi basin. In the north the St. Joseph River meanders into Lake Michigan, while in the east the Maumee flows northeastward into Lake Erie. The northern half of the state is dotted with many small lakes, including several of the state's largest. Nearly all of the forested land is commercially owned. Among the dramatic features of the landscape are the sand dunes along Lake Michigan, most of which have been removed from the public domain by industry and private homes. This situation was remedied somewhat with the dedication in 1972 of Indiana Dunes National Lakeshore. The most scenic part of the state is the south central region around Brown county.

Climate. Indiana has four distinct seasons and a temperate climate, usually escaping extremes of cold and heat. The mean temperature in January ranges from about 37° F (3° C) in Jeffersonville on the Ohio River to 25° F (−4° C) in South Bend in the north. In July the range is from about 78° F (26° C) to 73° F (23° C) in corresponding regions, and precipitation varies from 44 inches (1,120 millimetres) in the south central region to 35 inches in the north. Snow may fall over a six-month period and averages more than 20 inches annually, with the cities along the northern border often reporting more than 100 inches. The climate of northwestern Indiana is modified greatly by its presence in the lee of Lake Michigan. Cold air passing over the warmer lake water in fall and winter induces heavy precipitation, and winter snowfall, especially, is several times greater than in other parts of the state. In addition, average daily temperatures are warmer in the fall and cooler in the spring as a result of this "lake effect." Fall is perhaps the most delightful season in Indiana,

with colours provided by maples, oaks, tulip trees (yellow poplars), and a wide assortment of other trees, whereas the spring is generally erratic and unstable. Spring is also the season of the greatest number of tornadoes, Indiana being part of a belt of Midwestern states in which the frequency of these severe storms is unusually high.

Plant and animal life. Indiana is typical of eastern north central America in its variety of trees, birds, and small game animals, and, like its neighbours, it has through the years thinned out the numbers of many of these, including the great variety of life once common to the sand dunes. Steady growth of agriculture, urban areas, and industry and the consequent pollution have taken a steady toll of natural life. Pollution of both air and water is particularly severe along the southern tip of Lake Michigan.

Settlement patterns. The three major regions of Indiana, which generally follow physiographic divisions, are: the northern region of industry and truck gardening; the fertile central plains; and the large southern region, forested and the site of caves and limestone quarries.

Industrial northwest

The nighttime skies of northwestern Indiana are alive with the belching volcanoes of its steel furnaces, and during the daytime such cities as Gary and Hammond are often darkened by clouds of smoke and airborne industrial wastes. Because of this dramatically apparent air pollution, the contamination of inland and lake waters, and the steady encroachment of industry upon the dunes and other lakefront areas since the early years of the 20th century, the area has been a national target of conservationists. Southward to the Wabash valley are rich farmlands, obtained largely by draining and deforesting marshes. To the east, South Bend is an important manufacturing city and a noted educational centre. The northeastern part of the state is more forested and pastoral, although Elkhart and Fort Wayne are major industrial cities.

Indianapolis, a city designed after Versailles, Fr., and Washington, D.C., dominates the central plains. It is a strongly conservative city whose growth has occurred largely through immigration from rural areas and annexation. A railway and highway hub, Indianapolis also serves the surrounding farming belt as packer and distributor, and it is a major industrial city as well.

The advent of industry and railroads ended southern Indiana's early dominance, which was based on the river traffic of the Ohio and Wabash. The region's major city, Evansville, also serves adjacent areas of Kentucky and Illinois, and between it and Terre Haute to the north lie most of the state's oil and coal deposits. Many handsome examples of pre-Civil War architecture are found in the river towns. An old buffalo trace used by pioneers moving from Kentucky to the western prairies leads from New Albany, across the Ohio from Louisville, Ky., to the Wabash at Vincennes. Southward from Bloomington is a vast limestone belt underlain by numerous caves, which makes the state a major limestone producer. Brown county remains largely a backwoods area where log cabins abound on hillside farms.

©Cathy Melloan/EB Inc.

Beach along the south shore of Lake Michigan in Indiana Dunes State Park, northern Indiana, with (right) the steel mills of Gary in the background.

The people. Hoosiers are predominantly white, native-born Americans of native-born parents, most of whom trace their ancestry ultimately to England, Scotland, Ireland, and Germany. Significant concentrations of ethnic minorities, however, are found in the larger metropolitan areas. Blacks constitute about 70 percent of Gary's population, 30 percent of East Chicago's, and more than 20 percent of Indianapolis'. Citizens of Polish descent form the largest white ethnic group in South Bend; and, along with Hungarian, Belgian, Italian, and Mexican groups, they are numerous throughout the north. The state's overall lack of ethnic and linguistic mix helps to account for the continuing strength of Hoosier localism.

White Protestant domination

More than two-thirds of the people of Indiana who are part of an organized religion are Protestant, a figure considerably above the national average. Roman Catholics, who make up a sizable percentage of the population in northern Indiana, are concentrated largely in the urban areas with large continental European and Irish ethnic groups, particularly South Bend. Jews constitute a small percentage of the state's population and live almost exclusively in urban centres. The Amish people constitute a small group located in the northeast, in and around Middlebury, Nappanee, and Goshen. They conduct a model farm at Amish Acres in Nappanee. Mennonites, who also live in this area, have established a college in Goshen.

About two-thirds of the state's residents live in urban areas. More than one-fifth of the people are concentrated in the Indianapolis metropolitan area, and another large percentage are in the Gary–Hammond–East Chicago complex. The national pattern of deserting the central city for the suburbs has generally occurred throughout the state, with South Bend, for example, losing population while its surrounding county has grown. Of the counties losing population, most are located in the south.

The economy. Indiana's economy is concentrated in three sectors: manufacturing, services, and retail trade. Manufacturing accounts for about one-third of all jobs in the state and generates more than 40 percent of all personal income. The availability of labour and essential materials, as well as the state's location within 800 miles (1,280 kilometres) of 40 of the nation's 50 largest consumer and industrial markets and its high rank in interstate highway mileage, have all contributed to the growth of manufacturing in Indiana. Heavy industrialization, however, has made the state's economy vulnerable to recession.

Sources of economic strength

The American Railway Union, America's first industrial (as distinct from craft) union, was founded in Terre Haute in 1893 by Eugene Debs, five-time Socialist candidate for president. The following year it was involved in the Pullman strike that brought the intervention of federal troops and Debs's imprisonment. Since then Indiana has had its share of labour strife, especially in the steel industry, but, in general, business, labour, and government work to maintain an atmosphere attractive to industry.

Resources. Indiana is a major producer of U.S. building stone, quarried in the Bedford–Bloomington region. Bituminous coal from the southwest is the state's leading source of energy for heating and generating electrical power. Natural gas must be brought in by pipelines, though during the 1880s Indiana's "Gas Belt," stretching from central Indiana south to the Ohio River, was the world's largest producing field. Few attempts were made to conserve the gas, however, and by 1898 the supply was virtually exhausted. The Indiana and Michigan Electric Company's nuclear generating station in southwestern Michigan began operation in 1975.

Agriculture. Along with forestry and fisheries, agriculture employs about 3 percent of the labour force. Technological advances have meant that, despite drops in total farm acreage and number of farms, production has increased. Indiana ranks in the top quarter nationally in cash receipts from farm marketing. Major crops include corn (maize), soybeans, and wheat, with tomatoes the principal vegetable crop. Hogs are the most numerous livestock, with cattle, sheep, and poultry also important.

Industry. Since the 1850s manufacturing has grown to be the dominant source of income for the state, with the steel industry a major component. Elkhart is well known for its production of musical instruments, and Fort Wayne produces a large proportion of the world's supply of diamond tools. Overall, Indiana ranks among the nation's top 10 states in manufacturing.

The highway system

Transportation. Signs on the Indiana Toll Road proclaim the state to be the "Main Street of the Midwest," perhaps a fair estimate of its position in interstate transportation, whether by highway, waterway, air, or rail. Indianapolis is served by more major highways than any other American city, and some of the nation's largest moving companies have their headquarters there.

Responsibility for road construction and maintenance rests with city, county, state, and federal governments. County and state highway departments are subject to a patronage system whereby changes in administration bring on political dismissal and employment that result in instability in management. Nonetheless, Indiana ranks high nationally in road mileage per square mile of area, and almost all of its rural roads are paved. Though quantity may sometimes surpass quality of highway mileage, virtually all intrastate passengers and much commercial produce travel by road. Indiana has a comparatively dense network of railroad trackage. All lines running east from Chicago and St. Louis pass through the state. As in other states, however, the Amtrak system that went into operation in 1971 has sharply reduced passenger service. Commercial air service is available in major Indiana cities, and there are more than 300 public and private airports in the state.

The Ohio River, linking Indiana with the Mississippi River system, carries more low-cost freight than does the Panama Canal. The Port of Indiana harbour, on Lake Michigan about 10 miles (16 kilometres) east of Gary, was created artificially; opened in 1970, it connects Indiana with world commerce by way of the St. Lawrence Seaway.

Administration and social conditions. *Government.* Indiana's executive, legislative, and judicial structures are similar to those of other states, with some marked differences. The governor is elected for a four-year period and can serve no more than two terms in a 12-year period. Thus, gubernatorial influence on the General Assembly is often weak during the second half of an administration. The governor has veto power over legislation, but the veto can be overridden by a simple majority of the two houses. The governor's authority is wielded largely through his power to appoint and remove heads of nearly all departments, commissions, and governing boards of institutions. Several thousand jobs are subject to patronage.

The General Assembly

The bicameral General Assembly includes 50 senators, serving four-year terms, and 100 representatives, serving for two years. They may be reelected, but there is often a high turnover. In 1970 the voters approved annual sessions for the two bodies. The state constitution requires that the legislature reapportion itself according to population every six years. This law was ignored, however, from 1923 to 1963, during which time the rural areas exerted an influence far out of proportion to their declining population. Under pressure from the U.S. Supreme Court, the state eventually achieved a reapportionment based on the "one man, one vote" principle in 1965.

The judiciary is headed by the Supreme Court, which is composed of five judges appointed by the governor and a judicial nominating commission after a screening procedure. A new judge serves for two years and then, if retained, for a term of 10 years. The Court of Appeals consists of three regional divisions with three appellate judges each and a fourth judge elected at large. There are also circuit, superior, municipal, and county courts.

The four principal types of local government are the county, township, municipality, and school district. Townships can serve a dual capacity as school and civil townships, but overall their importance has been greatly reduced; welfare is now the primary function of townships. Boards of county commissioners have executive and legislative powers, while county councils are concerned almost exclusively with fiscal affairs. City voters elect a mayor and common council. In 1969 Indianapolis merged with Marion county.

Indiana is a two-party state, the Republicans having a slight advantage since the last quarter of the 19th century, especially in control of the governor's office and the General Assembly and in presidential voting. Voting trends show about one-third of the counties to be Republican, one-third Democratic, and one-third doubtful. In 1988 Indiana elected its first Democratic governor in 20 years.

At the national level the state can claim one president, Benjamin Harrison, the grandson of William Henry Harrison, and five vice presidents. In 1940 the Indiana native son Wendell L. Willkie was the Republican candidate for president.

Education. The state's educational system is headed by a board of education and a superintendent of public instruction. Indiana ranks in the lower half of the states in outlay per pupil, although it spends about one-third of its general revenue on education. The state's elementary and secondary educational systems are good, but they are not considered to be innovative or outstanding.

Leading universities

In the realm of higher education, on the other hand, Indiana has made notable achievements. The three leading universities of the state are Indiana, in Bloomington; Purdue, in West Lafayette; and Notre Dame, near South Bend. Indiana University, founded in 1820, has become noted for its work in several fields, including English, foreign languages, biology, medicine, and law, and its university press rates among the nation's finest, especially in the arts. The university's School of Music has become internationally known; among its performance series are works staged annually by the Opera Theater. Purdue University, organized in 1869 as a land-grant college, is one of the nation's leading engineering and agricultural schools.

The University of Notre Dame, dating from 1842, is widely regarded as the leading Roman Catholic university of the United States. Its faculty in a dozen graduate programs has become recognized at the national level, helping to overcome its image as an institution devoted solely to athletics. Notre Dame, originally a men's school, in 1972 began enrolling women from St. Mary's College, also located near South Bend.

Health and welfare. Hoosiers as a group do not take well to the idea of supporting public welfare programs. Indiana is among the more prosperous states in the nation, but it fails to apply its wealth to public programs. For example, not until the 1970s did the average weekly unemployment benefit in Indiana exceed the national average.

In comparison with the rest of the United States, Indiana ranks near the bottom in allotment per capita for public welfare. It spends less than 10 percent of its general revenue on welfare and only slightly more than 10 percent on health care and hospitals. Indiana is well below the national average in the number of recipients of general assistance per 1,000 population. Distrust of federal programs has tended to militate against making maximum use of aid that could be available, thus necessitating continued taxes at the state and local level that are a greater burden on the aged, on those with fixed incomes, and on the working and lower-middle classes. The problem that the state confronts in financing public welfare programs is as much ideological as it is financial.

Cultural life. The fine arts flourish in most of Indiana's major cities and even in some of the smaller towns. The Indianapolis Symphony Orchestra has a respected place among the nation's orchestras, and the city also boasts the Indianapolis Museum of Art and the Civic Theater, the nation's oldest continuously operating theatre. South Bend and Fort Wayne also have symphony orchestras, and one of the best known art colonies in the United States is located in Nashville, in the heart of Brown county.

The popular arts

Indiana's contributions to the popular arts in the United States include the Hoosier poet James Whitcomb Riley, novelists Booth Tarkington, Lew Wallace, and Kurt Vonnegut, satirist George Ade, and the World War II chronicler of the foot soldier, Ernie Pyle. Some of the country's most popular songs have been written by such Hoosiers as Hoagy Carmichael ("Star Dust"), Cole Porter ("Begin the Beguine"), J. Russel Robinson ("Margie"), Albert von Tilzer ("Take Me Out to the Ball Game"), and Paul Dresser ("On the Banks of the Wabash, Far Away"), brother of the novelist Theodore Dreiser. Among the most notable of Indiana comedians have been Herb Shriner and Red Skelton.

Almost every citizen seems to participate in Hoosier Hysteria, the state's annual high-school basketball tournament. Notre Dame, well known for its football talent, vies annually with Purdue and Indiana to provide Hoosiers with exceptional intercollegiate athletics. Indiana University has also become a mecca for basketball and for some of the world's greatest swimmers. Indianapolis is internationally known for the Indy 500, an auto race held annually on Memorial Day. The first race was held in 1911, while the city was still an automobile-manufacturing centre. The entire month of May has become devoted to the race, with such attendant events as a major professional golf meet. Indianapolis is also the home of the professional football Colts, and is the site of the annual U.S. Clay Court Championships, which attract top international tennis players. In 1987 it became the second U.S. city to host the Pan-American Games.

Indy 500

Hoosiers fond of the outdoors enjoy the state parks and forests and the many reservoirs, nature preserves, and wildlife areas. Indiana has many museums and historic sites, including the Levi Coffin House, a stop on the Underground Railroad, in Fountain City, and the Whitewater canal, with a covered aqueduct, in Metamora.

HISTORY

Prehistory and exploration. Archaeologists have discovered remains of the earliest known inhabitants at Angel Mounds, an archaeological site on the Ohio River near Evansville. Early historical records show that Algonquian Indians organized tribes of the area into the Miami Confederation, which fought to protect the lands from the unfriendly Iroquois. Other important Indian tribes were the Potawatomi and the Delaware. In the 17th century the French made treaties with the Iroquois allowing them to trade with the Miami Confederation.

In 1679 Robert Cavelier, Lord de La Salle, traveled by boat from Michigan down the St. Joseph River. To the south, traders from the Carolinas and from Pennsylvania settled on the Ohio and the Wabash river shores, threatening the French traders, to whom the region was a means of connecting Canada and Louisiana. To protect the route to the Mississippi, the French built Fort-Miami (1704); Fort-Ouiatanon (1719), near present-day Lafayette; and Fort-Vincennes (1732), one of the first permanent white settlements west of the Appalachians.

In 1763 the area, part of what came to be known as the Northwest Territory, was ceded to England, which forbade further white settlement. The prohibition was largely ignored, and in 1774 Parliament annexed the lands to Quebec. During the American Revolution Virginia, Connecticut, and Massachusetts made claims on the land, and in 1779 George Rogers Clark secured the area for the rebelling colonies by leading his troops on a surprise march from Kaskaskia to Vincennes.

Territorial period. The Northwest Territory was ceded to the United States by the Treaty of Paris, ending the Revolution in 1783, and in 1784 the first U.S. settlement was established at Clarkville, in the southern part of the state. Warfare between the Indians and the whites continued until 1794, when General Anthony Wayne defeated the Indians in a battle near Fallen Timbers, near the present-day Ohio–Indiana line, and the Indians were forced to make land concessions. Increasing numbers of white immigrants from Southern states entered the area after 1800, leading to renewed Indian resistance. In 1811 the last major encounter, the Battle of Tippecanoe, was fought near Lafayette, with General William Henry Harrison the victor. Between 1820 and 1840 the major Indian tribes abandoned the area. The Ordinance of 1787 creating the Northwest Territory prohibited slavery, but it did not abolish slavery already in existence, and in 1800 the territory had at least 175 slaves. With the end of Indian resistance came rapid settlement and in 1816 statehood. The territorial capital, Corydon, became the first state capital.

Indian warfare

Statehood. The patterns of rural life and local autonomy were established in the first half of the 19th century as settlement progressed from south to north. The Utopian community of New Harmony, on the Wabash River in the southwest, was settled by George Rapp in 1815 and taken over by Robert Owen in 1825. In 1801 the first college was founded in Vincennes, and in 1820 Indiana University was chartered. A single-car, horse-drawn railroad arrived in Shelbyville in 1834. The constitution of 1851, which remains the framework of state government, made it nearly impossible for the state to go into debt, reinforced the powers of local government, and created a tax-supported public school system. Article XIII prohibited the entrance of blacks into the state, but this was struck down by the U.S. Supreme Court in 1866 as being in conflict with the federal Civil Rights Act of that year.

Development of rural-industrial dichotomy

The period from 1850 to 1900 was one of agricultural and then industrial growth. The Civil War gave impetus to industrialization, and the northern part of the state emerged as a major sector. With the founding of the steelmaking city of Gary in 1906, midway between the iron ore of Minnesota's Mesabi Range and the coal of southern Indiana and Illinois, and with the subsequent development of automobile manufacturing in South Bend, Indiana moved from an agricultural to an industrial base. The isolation, independence, and spirit of Jeffersonian and Jacksonian democracy that underlay the constitution of 1851, however, continued to leave their mark upon the state. The document was written when towns and villages were days rather than minutes and hours apart. It was not until 1970 that annual rather than biennial meetings of the legislature were approved. Other features of the constitution remain impediments to effective management of 20th-century social and political problems, and the ideology of localism is still deeply ingrained. (W.V.D'A./Ed.)

Iowa

Iowa is a north central state of the United States, forming a bridge between the forests of the east and the grasslands of the high Prairie Plains to the west. Its gently rolling landscape rises slowly as it extends westward from the Mississippi River, which forms its entire eastern border. The state is bounded on the north by Minnesota, on the east by Wisconsin and Illinois, on the south by Missouri, and on the west by Nebraska and South Dakota. Its area is 56,275 square miles (145,753 square kilometres). Iowa, named for the Iowa (or Ioway) Indians who once inhabited the state, was admitted as the 29th state of the Union on Dec. 28, 1846. Des Moines has been the capital since 1857.

The popular image of Iowa—one of corn (maize) and hogs, flat prairies, and conservative people—is not altogether incorrect, but it masks both a subtle variety and the fact that Iowa and its people are very much in a middle position economically, politically, and geographically. With 90 percent of its total land area devoted to farming, Iowa is a major breadbasket of the United States and of the world. In addition, a large part of its industry is directly related to agriculture, and the rural population is still considerable. Iowans are strongly Republican in most years, but they exhibit a lively independence when they feel that the times dictate a different tack. Iowa has not shared the full benefits that have accrued from economic and demographic expansion elsewhere in the nation. Economic downswings that have afflicted other regions affect Iowa to the extent that they involve agriculture.

PHYSICAL AND HUMAN GEOGRAPHY

The land. Iowa's terrain and rich soils are the products of the continental ice sheets that periodically covered the

Glaciation

state during the Pleistocene epoch, between about 1,600,000 and 10,000 years ago. Glacial drift deposited by the earliest ice sheets filled the preglacial stream valleys, and little evidence of them remains.

The Illinoian ice sheet covered a small area of southeastern and extreme eastern Iowa, and in so doing it diverted the Mississippi and created a valley along its western front that can still be seen. Some 20,000 to 25,000 years ago the Wisconsin ice sheet moved southward in a lobe that ended at about the site of the present city of Des Moines. The Des Moines lobe began its final retreat about 13,000 to 14,000 years ago. Accompanying the last two stages of glaciation were extensive deposits of windblown silt, or loess, which in the western portion of Iowa were derived from the glaciation of the Great Plains to the west. As the ice sheets retreated, tremendous quantities of drift carried by the melting waters were deposited in the valleys. These various deposits form the basis of the Iowa landscape and make up the parent materials of the present soils.

Relief, drainage, and soils. The most varied relief anywhere in Iowa is in the northeastern part of the state, which was covered by the earliest ice sheets. There tributaries of the Mississippi cut deeply into the underlying bedrock. The Mississippi bluffs stand 300 to 400 feet (90 to 120 metres) above the valley, and the network of tributaries creates a scenic and hilly landscape.

Most of the state is underlaid by pre-Illinoian drift, which has been eroded for at least a few hundred thousand years by a relatively dense network of streams. Lakes or swamps that were left by the ice have long since been drained by natural erosion, and the result is a rolling landscape of great uniformity throughout most of the state. Near the Missouri River valley on the western border, the loess was piled 80 to 100 feet over the underlying drift surface, producing a line of bluffs 100 to 200 feet high. The highest point in the state, 1,670 feet (509 metres) above sea level, is in the northwest. The broad, flat uplands—which form the popular image of Iowa—are found mainly in the Des Moines lobe, a gently sloping, poorly drained drift plain that covers 12,300 square miles (32,000 square kilometres) in the central and north central portions of the state. Most of Iowa's lakes are in the northwestern part of this lobe.

Fertile soils

Most of the soils of Iowa, formed under prairie vegetation, are thick, dark in colour, and rich in organic matter and minerals. Only in the rough northeast and along the dissected river valleys of the south and southeast are there lighter-coloured and less-fertile forest soils.

Climate. Iowa's climate reflects the state's position deep in the interior of the continent. Winters are cold, with January temperatures averaging about 14° F (−10° C) in the northwestern section and 22° F (−6° C) in the southeast. Snowfall is light compared with the amount received in states to the east and north. Summers are warm and more humid, with daytime temperatures averaging 74° F (23° C) in July but varying from region to region. Maritime tropical air masses from the Gulf of Mexico bring frequent thunderstorms, with precipitation in June four times that of the winter months. Precipitation ranges from less than 28 inches (710 millimetres) in the northwest to more than 35 inches in the southeast.

Plant and animal life. Countless species of wildflowers once covered the prairies; and, though most of Iowa's virgin timber was cut long ago, almost 1,500,000 acres (600,000 hectares) are still forested. The only native evergreen is the red cedar, once found in profusion along the Cedar River. The state's streams are well stocked with dozens of species of fish, and trapping of muskrat and raccoon for furs is still widespread. The ring-necked pheasant—imported early in the 1900s—and quail are the major game birds, replacing the nearly extinct wild turkey. Small animals and a variety of other birds are also found.

Settlement patterns. Although Iowa is not a featureless plain, the relative homogeneity of physical characteristics has led geographers and other social scientists to use the state as an example of large-scale uniformity. Quarter sections of 160 acres formed the basis of much of the original settlement pattern. Consequently, farmsteads and the smaller towns generally are evenly spaced in the form of a grid, and most of the roads in the state follow a north–south or east–west line. Farmhouses amid the square or rectangular patterns usually have a row of trees serving as a windbreak and providing shade from the midcontinental sun. The largest city, Des Moines, is sited approximately in the middle of the state, above the Red Rock Reservoir of the Des Moines River. The other large cities are on the Missouri and Mississippi rivers at the western and eastern boundaries or on the Cedar River in the east.

The people. Iowa was settled largely by immigration from states lying directly to the east of it and from northwestern Europe. Until 1850 the southern third of Iowa received many settlers from the border states of the South, particularly Kentucky; but the influx from Ohio, Indiana, and Illinois and the New England and Middle Atlantic states was more important in the northern area. Settlers from Europe took on greater significance after 1850. The single most numerous group came from Germany, but Britain and Ireland were well represented. In the later years of the century, many Scandinavians settled throughout the western and central parts of the state. By 1915 there were few foreign-born people in southern Iowa, except Croatians and Italians in the coalfields and Dutch near Pella. The larger of Iowa's cities, particularly those on the Mississippi, attracted a variety of groups.

Sources of immigrants

Several ethnic and religious groups—an example is the Czech population of Cedar Rapids—are still present in Iowa. Among several experiments in communal living, the only survivors of the first years of pioneer hardship are the Amana colonies, a religious group originally from Germany that migrated to Iowa from Buffalo, N.Y., in 1855. This group changed to a corporation in 1932 and has been successful in maintaining its integrity while modifying its economy to fit the times. The strong religious and social traditions of the Amish groups living south of Iowa City and near Independence have come into conflict with modern society over state education laws. Mormons fled through Iowa on their way to Utah to escape persecution in New York and Illinois, and one large group remained behind at Lamoni, in southern Iowa. Quakers were important in the Springdale–West Branch area east of Iowa City in the mid-19th century. This area was an important link in the Underground Railroad, which helped slaves escape from the South before the Civil War. John Brown often visited Springdale, and President Herbert Hoover was born in West Branch of Quaker parents. West Branch is the site of the Hoover Presidential Library.

The only notable immigration into Iowa during the 20th century has been that of blacks to the larger urban centres, particularly Des Moines and Waterloo. Despite state and national civil rights laws, most blacks are concentrated in the decaying urban areas. Most Indians moved westward after federal land purchases in the 19th century, but some returned to purchase a small reservation—the Mesquakie Settlement—near Tama.

Major religious groups

Most of Iowa's population is Protestant (primarily Lutheran and Methodist), because major immigration was from northwestern Europe. Roman Catholics are strong in the northeast, in the Dubuque area, and in the larger cities. In its outlook southern Iowa is more fundamentalist; this had such social ramifications as the prohibition of liquor by the drink until 1963.

Iowa's population is evenly distributed, a factor that, together with the relative uniformity of the state's physical conditions, has made Iowa an excellent location for the testing of geographic and economic theories. Among the largest cities are Des Moines, Cedar Rapids, Davenport, Sioux City, Waterloo, and Dubuque. Most of the remainder of the population lives in scattered, evenly spaced small towns or in dispersed farmsteads. The rural population has declined since the mid-20th century, and Iowa overall experienced population loss in the 1980s.

The economy. The state attempts to aid industrial development and improve the economic situation in Iowa in a number of ways, particularly through trade missions and the activities of the Department of Economic Development. Corporate income taxes contribute a small part to revenues. The government's debt is low, and the overall labour picture is relatively bright. Unemployment rates and work stoppages tend to lag behind national trends.

Harvesting corn on a farm near Alden, north central Iowa.
Thomas Hovland from Grant Heilman

Iowa ranks at about the median for the United States in family income, but this is largely due to the fraction that is derived from agriculture. The cost of living is generally less than that in metropolitan centres of the East and Far West but above that of the South and Southwest.

Agriculturally based economy

Agriculture. The popular image of Iowa as an agricultural state is entirely correct. Iowa's agriculture is based on corn and soybean production and on the feeding and selling of animals. Iowa ranks among the top states in the nation in total value of all livestock, in hogs, and in cattle and calves. It is also one of the top three states in value of agricultural exports. The fertile land often produces crop surpluses that contribute to depressed farm prices, a persistent problem in the 20th century.

Industry. Iowa is located on the western fringe of the American manufacturing belt, and, although its manufacturing, trade, and service sectors exceed farming in income, much of the former is devoted to food processing or to the manufacture of agricultural machinery. In only a few instances does Iowa contribute significantly to the national economy in areas not related to agriculture. The production of electronic materials in Cedar Rapids, household appliances in Newton, refrigeration equipment in Amana, tires in Des Moines, writing instruments in Fort Madison, and rolled aluminum in Bettendorf are a few exceptions. Exploitation of mineral resources, except for portland cement and gypsum, plays a relatively minor role in the state's economy.

Transportation. In the 1920s Iowa developed an extensive rural road system designed for the low population density. It now has an excellent system of surfaced roads. The amount of the state's network of railroad track in active use has decreased over the years, and many Iowa towns have lost all railroad service. The state also has more than 240 public airports. Inland waterway traffic is important along the Mississippi River, and a channel nine feet deep runs up the Missouri to Sioux City.

Administration and social conditions. *Government.* Iowa's constitution at the time of admission in 1846 proved to be unsuitable, and a second version was drafted and ratified in 1857. This remains the fundamental law of Iowa, though it has been amended numerous times. The constitution provides for a separation of governmental powers into executive, legislative, and judicial components.

Executive branch

In the executive branch the governor, lieutenant governor, secretary of state, auditor, treasurer, secretary of agriculture, and attorney general are elected for four-year terms; all are eligible for an unlimited number of terms. A number of commissions, boards, and departmental executives are appointed by the governor, though most employees of the state departments are under a civil service system. The Iowa Civil Rights Commission investigates charges, holds hearings, and gives decisions on complaints of discriminatory practices in public accommodations, housing, employment, and education.

The bicameral General Assembly meets every year; longer sessions are held in odd-numbered years, when major budget items are decided. The House of Representatives has 100 members elected for two-year terms, while the Senate has 50 members elected for four-year terms. Both bodies are reapportioned every 10 years to ensure compliance with the "one man, one vote" decision of the U.S. Supreme Court.

The state judiciary is headed by the Supreme Court, which has considerable jurisdiction over the lower courts. The nine members of this body elect their own chief justice. Justices are appointed by the governor, are subject to a confirming popular vote one year later, and after an eight-year term may declare their candidacy for another term. There are 14 judicial districts in the state, with the number of judges varying according to population and caseload. Most larger cities have municipal courts; the others have police and mayor's courts. Justices of the peace are elected in those townships that lack municipal courts.

Local authority is vested in each county's board of supervisors, under which serve the elected auditor, sheriff, recorder, treasurer, and county attorney. The county government collects municipal, school, county, and state taxes; manages welfare; and operates the road system in cooperation with the State Highway Commission. Municipalities have only those powers that have been specifically granted to them by the General Assembly. This was an issue of considerable contention when the legislature was controlled by rural forces, but after reapportionment in the 1960s the urban-rural discord was reduced. Most of the smaller incorporated towns have a mayor–council form of government, whereas most of the larger cities have a council–manager or commission administration.

Republican voting record

Iowa's political tradition has been Republican. Between 1848 and 1968 only seven Democrats represented Iowa in the U.S. Senate. The Democrats failed to elect a candidate to the governorship until the Great Depression of the 1930s. Disquiet over farm prices, however, has elicited a substantial Democratic vote on several occasions.

Education. Iowa's first school was opened in 1830. Long known for excellence in education, Iowa ranks above more than half of the other states in expenditure per pupil. The University of Iowa (founded 1847), in Iowa City, is especially noted for the programs in fine arts, and Iowa State University of Science and Technology (1858), in Ames, has shown national and international leadership in the basic sciences, agriculture, veterinary medicine, and related fields. There are also a great number of other public and private institutions of higher learning; nearly all of the latter have religious affiliations.

Health and welfare. Welfare is managed on the county level, as are many health services, though federal and state funds support these activities. Health facilities are adequate in the larger cities and especially in the University of Iowa medical centre in Iowa City, but rural areas suffer from a lack of hospitals and doctors. Hospitals are being upgraded, often with federal support. Compared with other states, Iowa ranks low in the number of doctors per person, and efforts to lure doctors into rural practice have proved unsuccessful: about two-thirds of the doctors trained in Iowa's medical schools establish their practices in other states.

Cultural life. A widely dispersed population with small urban centres makes it difficult for Iowans to support many of the cultural amenities that exist in large urban settings. Traveling shows, including theatre and dance, symphonies, and guest artists visit many places in the state each year. The major cultural centres are the universities and colleges. The fine arts are notably supported at the University of Iowa, where the regional painter Grant Wood

did much of his work and where the Writer's Workshop enjoys national esteem. Art museums of significance are found in Iowa City and Des Moines. Such towns as Cherokee and Decorah have museums emphasizing the area's presettlement character or the early European settlers, while Davenport, Des Moines, Cedar Rapids, Sioux City, Dubuque, and Fort Dodge have museums or art galleries.

School sporting events

In a region generally lacking large urban centres, sporting events furnish much of the cultural life. The University of Iowa has long been one of the national leaders in basketball and football attendance. In every college town in the state, football weekends form the centre of the autumn social season. High school basketball and wrestling tournaments evoke great community enthusiasm near the end of the long, cold winters. Outdoor sports of all types are extremely popular, with hunting, fishing, boating, and camping especially prevalent.

Folk traditions are maintained in the Amana colonies, with their Oktoberfest; in the Dutch community of Pella, with its annual tulip festival; among the Czechs of Cedar Rapids; and in other localities.

HISTORY

Prehistory. The earliest inhabitants of what is now Iowa (Paleo-Indians) probably occupied ice-free land during the time when the Des Moines lobe was covered by glaciers. The earliest archaeological evidence of settlement, however, dates from around 8,500 years ago. The hunters and food gatherers of this period existed at the subsistence level, enduring the periodic droughts that continue to plague the region today. Even after the advent of sedentary agriculture in western Iowa around AD 800, entire villages occasionally disappeared. In eastern Iowa, effigy mound builders occupied settlements from about 300 to the 17th century. Most of the early Indians were of the Siouan language family, although Algonquian-speaking tribes were important in eastern Iowa after the 17th century, often displacing the western tribes in bloody conflicts. The Iowa (Ioway) tribe was virtually annihilated shortly before the advent of dense white settlement. All the Indian tribes ceded their lands through treaty and purchase in the 1830s and '40s. The last purchase was of Dakota (Sioux) lands in northern Iowa in 1851.

From territory to statehood. The first Europeans to reach Iowa were probably the French explorers Louis Jolliet and Jacques Marquette in 1673. Permanent white settlement, however, did not take place until the early 1830s, though Spanish land grants were occupied in the late 1700s, principally to exploit the lead-mining potential around the site of Dubuque. In the interim, both pioneers and Indians moved through the area exploring or hunting. The combined French and Indian history can be seen in geographic names throughout the state: for example, Des Moines, Dubuque, and Le Mars; Ottumwa, Keokuk, and Onawa.

The area that includes the modern state of Iowa was included in the Louisiana Purchase from France in 1803, and during the War of 1812 a U.S. garrison was driven from Fort Madison on the Mississippi River. Following

Rapid settlement

the purchase of eastern Iowa from the Sauk and Fox Indians in the 1830s, U.S. settlers rapidly moved in to till the land. The Territory of Iowa was established in 1838, with a population of 23,242. In 1846 Iowa was admitted to the Union as part of a compromise between the slaveholding South and the free North. By 1860 there were nearly 675,000 people in the state, and with the construction of railroads the frontier was pushed farther westward. The population of Iowa more than tripled during the 1850s, and the Spirit Lake Massacre in 1857 marked the final instance of Indian hostility in the state. The years immediately prior to the Civil War were Iowa's frontier days, however, with lawlessness, vigilantes, and lynchings accompanying the unsteady beginnings of a settled society.

Iowa was deeply involved on both sides of the issues that led to the Civil War, to which the state contributed more troops in proportion to its population than any other state. No battles were actually fought in Iowa, though a Confederate guerrilla raid from Missouri occurred in 1864.

Economic stabilization. The end of the Civil War, railroad expansion, and the removal of the Indian threat opened the prairie to settlement by massive waves of immigrants from states to the east and also from Europe. By 1900 claims to the land had filled the state, and the population showed a slight decline in the decade that followed. The wide use of barbed wire permitted diversified agriculture, and the draining of wetlands began the development of an efficient agricultural production that often threatened the financial stability of the state with too plentiful a harvest. Corn was the basis of Iowa's agriculture from the beginning, nearly all of the crop being fed to livestock.

World War I created short-term demands for maximum production and high prices, and since then the state has had recurring agricultural surpluses, low prices, and high land values. The various economic panics and depressions of the 19th and 20th centuries were only temporary impediments to this pattern of growth. In the past century Iowa politicians have appeared most prominently on the national scene when farm crises have been major issues.

The last significant case of exploitation of natural resources occurred in the coalfields of southern Iowa, beginning in the mid-19th century and reaching its peak in the first two decades of the 20th century. Most of the coal was quickly exhausted, however, and the miners moved on, leaving behind decaying towns and a deteriorating landscape.

After World War I the population growth of Iowa slowed. Attempts were made to entice industry into the state to diversify the economy, as animal feeding had diversified agriculture half a century before. Such attempts were not entirely successful, and Iowa's economy has continued to be governed by fluctuations in agricultural markets.

(N.E.S.)

Kansas

Lying amid the westward-rising landscapes of the Great Plains of the North American continent, Kansas became the 34th state of the United States on Jan. 29, 1861. In that year the capital was located in Topeka by popular vote, outpolling nearby Lawrence by some 2,700 ballots. The state's 82,277 square miles (213,098 square kilometres) are bounded by Nebraska on the north, Missouri on the east, Oklahoma on the south, and Colorado on the west. The state's name was derived from that of the Kansa, or Kaw, Indians.

The geographic centre of the 48 coterminous states of the nation is marked by a limestone shaft and a flag located in a pasture near Lebanon, Kan., close to the Nebraska border. Some 40 miles (65 kilometres) to the south is the magnetic, or geodetic, centre of the terrestrial mass of North America; this is the reference point for all land survey in the United States, Canada, and Mexico.

Central location

Kansas was once seen as the agricultural heartland of the nation. After 1952, however, industry began to contribute more to the economy than did the wheat fields and cattle ranches. Wichita, the state's largest city, is known locally as the Air Capital of the World because it produces more general aviation aircraft than any other city.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief, drainage, and soils.* Kansas has been characterized as a featureless plain, but its topography, while rarely spectacular, is varied. The land rises slowly but steadily from 700 feet (213 metres) above sea level in the southeast to 4,039 feet (1,231 metres) near the Colorado border. The far western section consists of high plains with few natural trees and appears flat and endless. Actually these plains are creased with shallow gullies, called draws, the product of millennia of erosion. Here are some of the state's most striking geologic formations. Castle Rock, south of Quinter, consists of chalk spires rising high above the level plains. Monument Rocks, a few miles to the west, resemble sphinxes. Near Jetmore is Horse Thief Canyon, a miniature of the Grand Canyon.

Under irrigation, southwestern Kansas has come to produce truck crops and sugar beets. Northeastern Kansas, once covered by the glacier that crept over most of the northern United States, is hilly and timbered, with many

creeks and springs. The southeast, lying near the foothills of the Ozark Mountains, is rough and covered in parts with scrub oak. In south central Kansas near Medicine Lodge are the Gypsum Hills, which resemble the mesas of the Southwest and are named for the gypsum found in them. In east central Kansas the Flint Hills stretch from north to south; gentle, rolling, largely treeless, and covered with bluestem, they provide the only extensive unplowed tract of true prairie remaining in the United States.

The principal rivers are the Kansas and the Arkansas. Tributaries of the Kansas are the Big Blue, the Republican, the Solomon, the Saline, and the Smoky Hill, all in northern Kansas. The Arkansas flows into the state from Colorado and winds through southwestern and south central Kansas, continuing through Oklahoma and Arkansas to the Mississippi. Tributaries of the Arkansas are the Cimarron, the Verdigris, and the Neosho (Grand).

Millions of years ago much of Kansas was the floor of an inland sea. The land was built up by the deposit of soil and vegetable matter from streams feeding the sea. This residual soil is among the most fertile in the world, and in it prehistoric fossils of great importance have been found.

Fertile soil

Climate. The climate of Kansas is temperate but continental, with great extremes between summer and winter temperatures but few long periods of extreme hot or cold. The annual average temperature is 55° F (13° C). The growing season ranges from mid-April to mid-September. Normal annual rainfall ranges from less than 20 inches (500 millimetres) in the west to more than 40 inches in the southeast.

Plant and animal life. Buffalo grass is native in the west and central areas of the state, bluestem around the Flint Hills, and bluegrass in the east. Wildflowers of many kinds are found in all parts of the state, and sunflowers grow in profusion (Kansas is popularly known as the Sunflower State). The cottonwood grows throughout Kansas, while in the northeast there are many oak, walnut, and maple trees, as well as cedar and elm. Western Kansas abounds in quail and pheasant and has the largest population of prairie chickens (grouse) in North America. Deer, once almost extinct, were protected by law for many years and have multiplied to the degree that hunting is again allowed. The once proliferating buffalo that roamed the plains are now found only in parks and zoos.

Settlement patterns. Most western Kansas farms or ranches are large, covering not less than one section (a square mile, or 640 acres [259 hectares]) of land, though a farmer's holdings may not always be contiguous. Eastern Kansas began with small farms, some of no more than 40 acres; but these have grown. A Kansas law forbids other than family corporations for farming purposes. Most of the small towns are modern and well kept, with paved streets and full utilities. Many of the small cities, especially in the west, present unexpected cultural and commercial resources, perhaps because they often lie far apart and draw from large trade territories. In the east the cities are older, closer together, and generally less progressive, though most of them are attractive, with broad, well-shaded residential streets and downtown shopping facilities.

Wichita, the largest city, has the state's largest buildings, biggest industries, and most venturesome businesses. In Topeka, where state government once was the largest industry, more people now are employed in services. Kansas City, Kan., merges with its larger neighbour, Kansas City, Mo., and contains a significant part of the industrial complex of that region, as does neighbouring Johnson county. Leavenworth, the state's oldest city, is built around institutions, including an army post at Fort Leavenworth, a federal prison, a state penitentiary (in the bordering city of Lansing), and a veterans' hospital. Lawrence, home of the state's largest university, depends on the school for its economy, though the city has worked successfully for industry. Most of the other cities depend on farm trade and agriculturally related business.

Kansas suffered during most of its history from two kinds of regionalism: one that pits rural against city dwellers; the other, the east against the west. The two are related in that none of the state's principal cities is in the west. More thinly populated than the east, western Kansas has always feared and fought eastern domination, while the east often has ignored the west. Wichita, one of Kansas' four metropolitan areas, contains nearly one-fifth of the state's population. The Kansas City–Lawrence–Topeka area of northeastern Kansas, containing three metropolitan areas, is more populous and is the centre of much industry. Rivalry between these two urban areas is obvious in the state legislature. People from the rural areas, mostly farmers, ranchers, and owners of small businesses, as well as residents from the smaller towns, have tended to distrust the cities, often bringing about an impasse in the legislature.

Regionalism

The people. Kansas' early settlers were principally antislavery New Englanders of Anglo-Saxon stock. After the Civil War and with the building of the railroads, many central Europeans were attracted by the promise of jobs laying track and of free land when the jobs were finished. Small communities populated by citizens of predominantly Russian, Bohemian, German, or Scandinavian ancestry still dot the state. The original languages have largely disappeared, though here and there church services are still conducted in German or Swedish, and a few communities hold festivals each year at which the old folkways, foods, and languages are featured. During World War II there was an influx of servicemen and aircraft workers, many of whom remained. The state is largely Protestant, with large communities of Methodists, Baptists, and Lutherans. Virtually every sect is represented in the state, including such rare groups as the Amish and the Dunkard Brethren. Roman Catholics are slightly more than one-fourth of all religious adherents.

Because of insufficient employment opportunities, Kansas loses a considerable number of its young people to other states. The birth rate, however, produces a slight natural increase in population in most years. The most conspicuous demographic trend is from the farms to the cities. As further technological advances in farming are made and individual landholdings increase in size, this trend undoubtedly will continue.

The economy. The national trend away from manufacturing and toward service industries has been experienced to a lesser degree in Kansas, which remains above the national average in percentage of employees in manufacturing. Small and medium industries account for increasing percentages of the overall numbers of employees. The availability of a reliable work force is one of the advantages the state has to offer to prospective employers.

Resources. Kansas has abundant farmland, large mineral resources, a good labour force, a healthy retail trade, ample electrical power, plenty of water, and a central location. It is among the 15 top mineral-producing states. Oil and gas production is declining, but Kansas leads the nation in production of helium and is a major producer of portland cement, stone, clay and clay products, sand, salt, gravel, zinc, bituminous coal, and lead, and its chalk supply is virtually limitless. In petroleum refining capacity Kansas remains among the top states. Kansas has imposed only a small severance tax (laid at the time of severance, or extraction from the ground) on oil or gas for the purposes of conservation and pollution prevention.

Agriculture and industry. Both agriculture and manufacturing contribute significantly to Kansas' economy—the former contributing many raw materials for the latter. The production of its farms and ranches places Kansas first among the U.S. states in wheat and first in sorghum grains. It also ranks high in wild hay, beef, and hogs.

Manufacturing and processing plants produce everything from airplanes to zinc castings. Wichita is a major producer of camping gear; it also manufactures heating and air-conditioning equipment, snowmobiles, and a variety of other products. In addition to ranking first in the world in production of general aviation aircraft, Wichita also is an important manufacturing centre for military aircraft. Other plants in the state turn out baby foods, pet foods, prefabricated houses, mobile homes, greeting cards, tires, paint, and dishwashers. The state has a right-to-work law that forbids compulsory unionism. Most cities issue revenue bonds to encourage new industry. Kansas still leads the nation in wheat milling, despite some decline, and is among the top producers of processed beef.

Varied manufacturers

Grain elevators at Mingo in northwestern Kansas.
Grant Heilman/Grant Heilman Photography

Military installations. Two long-established army posts have contributed significantly to the state's economy. Fort Leavenworth, with its renowned Command and General Staff College, dates from 1827. A major outpost in the early Indian wars and during the Civil War, it has offered sophisticated training to international military officers for many years. Fort Riley, near Junction City, was established in 1853 and was also an Indian outpost. In the 20th century it has been an important infantry-training centre and is the home of the famous 1st Infantry Division (The Big Red One). McConnell Air Force Base at Wichita is a part of the Strategic Air Command and one of four bases in the nation to house the B-1B bomber.

Transportation. Kansas has an excellent system of railroads for east–west transport but, except in the east, has less adequate north–south lines. The same may be said of its highways. Exceptions are the state's single toll road, the Kansas Turnpike, which runs southwesterly from Kansas City to the Oklahoma line south of Wichita, and Interstate 135. Although Kansas has more than 350 airports and is served by several airlines, the only major airport with transcontinental service is in Wichita.

Administration and social conditions. *Government.* Under the constitution adopted in 1859, Kansans elect a governor, lieutenant governor, attorney general, and secretary of state; most other state officers are appointed. The legislature comprises 125 representatives and 40 senators, elected for two-year and four-year terms, respectively. The legislature holds an unlimited session in odd-numbered years and meets for no more than 90 days in even-numbered years. Each of the 105 counties elects commissioners, a county attorney, a treasurer, and other officers. Judges of the 31 judicial districts are elected, but the seven justices of the Supreme Court and seven judges of the Court of Appeals are appointed by the governor from a panel presented by a Supreme Court nominating commission. The justices are subject to the approval of the voters.

Innovations in government and law

The first legislative council in the United States was inaugurated in Kansas in the 1930s. It was an interim body designed to work between legislative sessions at analyzing and drafting laws. Several other states later adopted legislative councils. The 1969 legislature provided for prefiling of bills between sessions, a change that persuaded the legislature that the council was no longer necessary. It was replaced in the 1971 session by the Legislative Coordinating Council, made up of the leadership of both houses.

In 1933 Kansas enacted a "cash basis law," which requires that no state money be expended until it has been raised and appropriated by the legislature. Bonds have been issued only for capital improvements, such as state buildings and highways, in which case they are retired by user fees.

Kansas once was known as the most Republican state in the nation, but it now has a sizable Democratic minority, a growing independent vote, and a small Libertarian contingent. The first legislature, in 1861, gave women the right to vote in school elections. In 1887, women's suffrage was extended to city and bond elections, and in that year the country's first woman mayor was elected in Argonia. The state constitution of 1861 granted women equal rights to own property and to have control of children. Universal suffrage was granted in Kansas in 1912. Kansas ranks high among the states in the proportion of women holding public office.

The Farmers' Alliance and the People's (Populist) Party both had their origin in Kansas, and in the 1890s they played an important part in the politics of the Midwest. Kansas pioneered the direct primary election, and a Kansas senator introduced the resolution in the U.S. Congress that put direct election of U.S. senators into the federal Constitution.

Kansas was the first state to adopt the constitutional prohibition of alcoholic beverages. The prohibitory amendment was added to the state constitution in 1880 and was not repealed until 1948. In 1986 voters approved a constitutional amendment permitting the sale of liquor by the drink in establishments that do at least 30 percent of their business in food sales. In the same election parimutuel wagering and a state lottery were approved.

Education. A landmark civil rights case of the 20th century, *Brown* v. *Board of Education,* originated in Topeka in 1951, when the clergyman father of a nine-year-old black girl led her to the door of an all-white school. She was denied enrollment, and the decision that was handed down by the U.S. Supreme Court in 1954—basically stating that segregated, even "separate but equal," education is inherently unequal and must be eliminated with all due speed—became the basis for most of the civil rights decisions that have been applied to schools since that time.

In the mid-1960s Kansas abolished its office of state superintendent of public instruction and substituted the Department of Education headed by a commissioner and an elected state board of education. There are some 300 public school districts throughout the state. A number of two-year colleges and vocational-technical schools are operated by the communities in which they are located.

Universities and colleges

Kansas has six state universities. Fort Hays State University, Pittsburg State University, and Emporia State University offer liberal arts degrees. The University of Kansas is located in Lawrence, Kansas State University in Manhattan, and Wichita State University in Wichita. Kansas State, recognized as having one of the country's leading agricultural colleges, was the first land-grant university in the United States. The state's medical school is part of the University of Kansas Medical Center, with its campus at Kansas City. In 1971 the School of Medicine established a second campus at Wichita to expand its clinical teaching facilities. There are two law schools, one at the University of Kansas, the other at Washburn University of Topeka, a municipal school. In addition, there are several church-affiliated, private four-year colleges in Kansas, all offering liberal arts degrees.

Health and welfare. A Department of Health and Environment is responsible for health information and education and has supervisory authority over environmental problems, water and waste management, air quality, and radiation, as well as food, drugs, lodgings, vital statistics, and general health concerns. A Department of Social and Rehabilitative Services operates mental hospitals at Topeka, Osawatomie, Larned, and Norton and offers services in geriatrics, public health nursing, nutrition, maternal and child health, and mental retardation. The Menninger Foundation and Clinic are at Topeka.

Since 1862 Kansas has had some form of public assistance for the needy. The Department of Social and Re-

habilitation Services offers financial assistance and special education. Vocational and rehabilitational services also are provided for the handicapped. There are schools for the mentally retarded at Parsons, Winfield, and Norton. A children's home has operated at Atchison since 1855; there is a state treatment centre for mentally disturbed children at Topeka and a girl's reformatory at Beloit.

The state has a fair housing law and a civil rights commission that hears grievances and attempts to mediate them.

Cultural life. The citizens of Kansas resent the suggestion that they live in a cultural desert, but the assertion is at least partially true. Most of the larger cities have amateur theatre groups, while Topeka and Wichita support symphony orchestras. The numerous colleges and universities in the state provide a concentration of art and music in many small communities that otherwise would have no comparable activities. In the sparsely populated areas of western Kansas, however, a large number of the small communities have few cultural institutions except a public library. Wichita, however, has several art museums and a cultural and civic centre with two theatres, an exhibition hall, and a convention hall. The extreme eastern areas of Kansas look to Kansas City, Mo., for cultural attractions. In the mid-1960s the Kansas Arts Commission was formed; funded by the state, it seeks to encourage the development of the arts, often providing money for communities or organizations that want to develop cultural events. The University of Kansas has an outstanding museum of natural history and an art museum. The Eisenhower Center at Abilene, boyhood home of the 34th U.S. president, has a museum and a library containing the papers and memorabilia of his presidency and military career.

Traditional commemorations

In addition to an art museum, the small community of Lindsborg has a biennial folk festival, the Svensk Hyllningsfest, which honours the Swedish pioneers who settled the town. It features Swedish costumes, traditional food, folk dances, and displays of the arts and crafts of local artisans. Wilson has a Czech festival each year. Examples of eccentric folk sculpture are found in Lucas, where a self-taught artist, working in wet concrete, sculpted his own idea of the Garden of Eden and other biblical stories.

HISTORY

Indians, explorers, and settlers. Archaeological exploration has uncovered evidence of Indian cultures that existed in Kansas for many centuries before the Europeans settled on the land. From about 1200 to 1500 there had been a thriving agricultural society in the area of the Republican and Big Blue rivers.

The first known European explorers were Spaniards under Francisco Vázquez de Coronado, who in 1541 rode northward from Mexico seeking the gold of the legendary Seven Cities of Cíbola. Juan de Padilla, a priest with the expedition, founded the first mission in the territory, possibly north of present-day Wichita. The territory was claimed for France in 1682 by Robert Cavelier, Lord de La Salle. During the 18th century French fur traders had a flourishing exchange with the Indians in what is now the northeastern part of the state.

The region passed to the United States as a part of the Louisiana Purchase in 1803. The explorer Zebulon Montgomery Pike passed through Kansas in 1806 and described it as the "Great American Desert"—a false image that still persists. Kansas was thoroughly explored during the following decades, but westward-bound settlers and miners passed through it without staying.

Creation of Kansas Territory

From 1830 to 1854 Kansas was in an area designated as Indian Territory, where tribes who occupied lands wanted by whites were relocated. The Kansas–Nebraska Act of 1854, however, created two territories and opened both to settlement, allowing residents to determine whether their future states would be free or slave. The rush began, and Kansas became a major breeding ground for the American Civil War as North and South each attempted to send the most settlers into the new territory. Most early settlement was near the eastern border, and free staters were harassed constantly by Border Ruffians from Missouri. One incident was the sacking of Lawrence by Southern guerrillas in 1856. The abolitionist John Brown, with his sons and a few other men, retaliated by dragging five of their proslavery neighbours from their homes and killing them. Proslavery forces attempting to avenge this massacre were captured by Brown, who became a hero to the Northern sympathizers. Hundreds of such incidents won the territory the name Bleeding Kansas.

Statehood. Kansas entered the Union as a free state in 1861. Before and after the Civil War sporadic fighting occurred between the settlers and the Indians. In 1867 a peace treaty was signed in which the Indians agreed to sell their land; in return, the United States agreed to build homes for them in what is now Oklahoma and to provide money, food, and clothing. The U.S. Congress did not honour the treaty, and when the Indians returned they found their land occupied by white settlers. Further sporadic battles continued until the last Indian raid, in 1878.

Early settlers in wooded eastern Kansas lived in log cabins, but in the west they had only dugouts or sod houses. The unpredictable weather, the recurring Indian raids, the droughts and dust storms, and the grasshopper invasions discouraged many early settlers. One of the heroes of this era was William Mathewson, known as the original Buffalo Bill, who hunted buffalo for the settlers all of one winter without pay, providing meat by the wagonloads. The coming of the railroads in the late 1860s and the '70s made first one village and then another into boisterous cow towns. Texas cattlemen drove herds northward to Caldwell, Wichita, Dodge City, Ellsworth, Newton, and Abilene to reach the railhead. Although this development brought prosperity to Kansas and created a persistent image, the cow-town era lasted less than a decade.

The Mennonites arrived in 1874, bringing trunks full of hand-selected grains of Turkey Red wheat. This excellent strain was the basis of the abundant crops that became an important part of the Kansas economy. Many of the Mennonites' descendants remain as prosperous farmers.

Expansion of agriculture

By about 1890 most of the land was occupied, and Kansans settled into a life dominated by agriculture. World War I produced a great demand for food, and more and more prairie was plowed and put into production, which led to temporary prosperity but contributed directly to the dust storms that devastated the state in the 1930s. World War II contributed to Kansas' growing eminence in aircraft and brought many people from Oklahoma and Arkansas to work in Wichita's aircraft plants.

The decades of the 1970s and '80s were characterized by a slow but steady growth in population, one of the lowest unemployment rates in the nation, and a steady increase in the number of Kansans employed in mining and in health care and other service industries. Kansas remains a Republican stronghold, but, for 20 of the 30 years between 1957 and 1987, Democrats occupied the governor's office, suggesting that some political as well as economic change has been taking place in Kansas. (C.G.P.)

Michigan

Since its admission on Jan. 26, 1837, as the 26th state of the Union and the fourth to be carved from the Northwest Territory, Michigan has become a mainspring in the economic life of the United States; the name of its largest city, Detroit, has become a byword throughout the world for the American automotive industry. The state also has retained its prominence in agriculture, and, because of its many inland lakes, its borders on four of the five Great Lakes, and its many wilderness tracts, Michigan has evolved into one of the nation's leading tourist regions.

The Upper and Lower peninsulas

Michigan is the only one of the 49 continental states to be split into two large land segments: the sparsely populated but mineral-rich Upper Peninsula slices eastward from northern Wisconsin between Lakes Superior and Michigan, and the mitten-shaped Lower Peninsula reaches northward from Indiana and Ohio. The two landmasses have been connected since 1957 by "Big Mac," the five-mile (eight-kilometre) Mackinac Bridge across the Straits of Mackinac, which separate Lake Michigan on the west from Lake Huron on the east. Between Lake Huron and Lake Erie, in the southeast, the Lower Peninsula is sep-

arated from the Canadian province of Ontario by Lake St. Clair and the St. Clair and Detroit rivers. The state's name is derived from an Ojibwa (Chippewa) Indian word meaning "large lake." Although its 58,527 square miles (151,586 square kilometres) rank the state only 23rd nationally in size, the inclusion of Great Lakes waters over which it has jurisdiction raises the figure to 97,102 square miles (251,495 square kilometres), placing it 10th.

More than 70 percent of the state's residents live in urban areas, with a heavy concentration in the industrialized centres of the Lower Peninsula. This factor, together with a broad array of ethnic and national stocks among the people and a high number of lesser-skilled workers attracted to Michigan by the union-dominated labour scene, has created in many cities the typical marks of economic progress and poverty existing side by side, with a sometimes tenuous social stability. The state government, in the capital of Lansing, coordinates a vast network of programs attempting to reduce these contrasts, and it has provided a system of public higher education that is among the most diversified and renowned in the nation.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The mildly rolling countryside of Michigan appealed to the early agricultural settlers. Elevations are not high. The highest point in the Lower Peninsula, near Cadillac, rises only about 1,600 feet (488 metres). Elevations rise to over 1,900 feet only in the western Upper Peninsula.

Several physical features of Michigan are appealing. The sand dunes on the Lake Michigan shore are used annually by thousands of vacationers, and state forests, parks, and wildlife areas, containing more than 4,000,000 acres (1,619,000 hectares) of wooded land, include varied landscapes that have helped Michigan to become a major tourist attraction of the Midwest.

Drainage. Michigan's rivers, the majority of them in the southern part of the Lower Peninsula, drain vast interior areas. Most of them are shallow and narrow. Their navigability and the ease of bridging them encouraged settlement. Several of the rivers, especially in the Upper Peninsula, have falls to serve as a basis for waterpower.

Inland lakes

Michigan's 11,000 inland lakes range in size from a few acres to the 20,000 acres (8,000 hectares) of Houghton Lake in the north central Lower Peninsula. The shores of many lakes are ringed with summer cottages, as are the shores of the Great Lakes in many places. Two recreational areas in the state include the Pictured Rocks National Lakeshore on Lake Superior and the Sleeping Bear Dunes National Lakeshore on Lake Michigan.

About 500 islands dot the lakes and rivers. Belle Isle, a public recreation centre, and Grosse Ile, largely residential, are well-known features of the Detroit River. Mackinac Island is a resort on which motor vehicles are prohibited. Isle Royale, a virgin wilderness of nearly 900 square miles, is a national park in western Lake Superior.

Soils. Michigan soils vary with the diversity of climate, landforms, and water conditions. Clays and loams in the southern Lower Peninsula permit extensive agriculture, while unproductive sandy soils dominate in the northern Lower Peninsula. The Upper Peninsula has a few acres of fertile loam, but most of the soil is sandy and similar to the northern Lower Peninsula. Bog or muck soils, which were created as inland lakes became filled with organic matter, are found throughout the state, especially in the southeastern Lower Peninsula. These soils are particularly important for vegetable production.

Climate. The Great Lakes cool the hot winds of summer and warm the cold winds of winter, giving Michigan a milder climate than some other north central states. Although the Upper Peninsula is cooler, the temperature ranges in the far northern and far southern cities do not differ excessively. In the far north, Sault Sainte Marie has an average January temperature of about 14° to 16° F (−10° C to −9° C) and an average July temperature of about 64° F (18° C). In Detroit, in the southeast, the respective averages are 26° F (−3° C) and 73° F (22.8° C). The greatest amount of moisture is received in southern Michigan, and the state's average is about 31 inches (787 millimetres). The central portion of the state has less precipitation than does the Upper Peninsula, and the coastal strip along Lake Michigan receives an unusually large snowfall from westerly storm fronts moving across the lake. The growing season ranges from approximately three months in the Upper Peninsula to as long as six months in the more southerly portions of the Lower Peninsula.

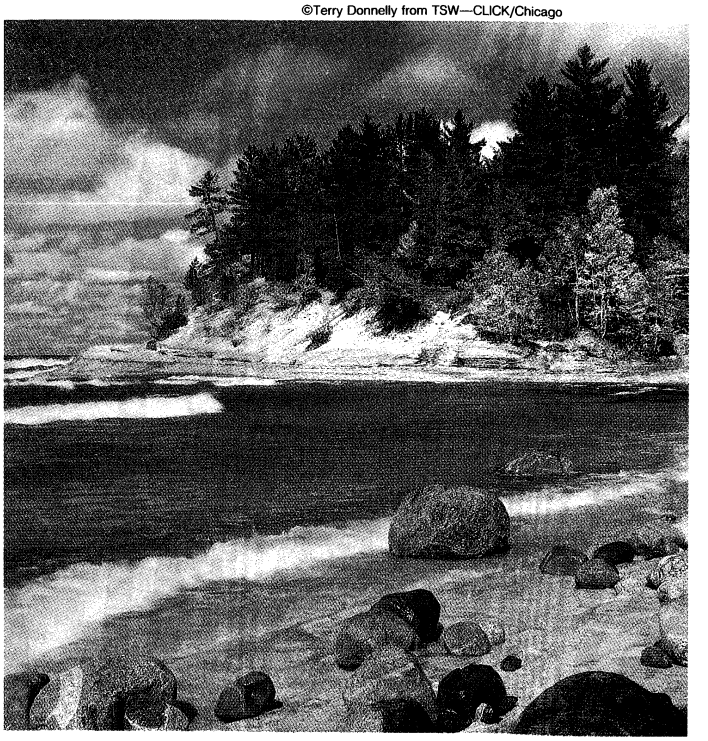
Plant and animal life. Almost all of Michigan was once heavily wooded, with genuine prairies or clearings found only in the southwest. Hardwood timbers included the hickory, ash, oak, and hemlock, though the white and Norway pine were the most common timber in the north. Animals native to the area are numerous. Whitefish, lake trout, and salmon in abundance swim in the Great Lakes, and many of Michigan's streams contain various edible trout. The Department of Natural Resources operates hatcheries and encourages tourism around the inland lakes, where perch, pike, and bass abound. The beaver was sought eagerly by early traders, and other fur-bearing mammals were also found in large numbers. Deer and bears, as well as quail and ducks, remain numerous in many counties.

Geographic divisions

Settlement patterns. Many of Michigan's people think of their state as divided into the Upper and Lower Peninsula, but the two physiographic divisions, the Superior Upland and the Central Lowland, follow this plan only in part. The Upland comprises the rugged western region of the Upper Peninsula, where the abundance of copper and iron ores has made the area economically dependent on the mining industries. Ironwood, Iron Mountain, Hancock, Houghton, and Marquette are among the larger cities. Agriculture is insignificant, but tourism and recreation offer possibilities for diversifying the economy.

Geographic and historical forces unite the eastern counties of the Upper Peninsula with the Lower Peninsula north of the Bay City–Muskegon corridor to create the area known as northern Michigan. There the white pine forests were exploited. The sandy soil, with adequate moisture, produces an abundant yield of potatoes and grain. Fruit growing along the Lake Michigan shore north to Traverse City has succeeded because the lake's influence prevents early killing frosts. The towns in the north serve as regional centres for tourists and farmers and often as headquarters for governmental services. Many of the larger communities have attracted small-scale manufacturing. Cities in-

©Terry Donnelly from TSW—CLICK/Chicago



Shore of Lake Superior near the mouth of the Mosquito River in Pictured Rocks National Lakeshore, Upper Peninsula, Michigan.

clude Sault Sainte Marie, Petoskey, Ludington, Manistee, Cadillac, and Traverse City.

South of the Bay City–Muskegon corridor a fertile clay soil and a longer growing season permit a wider variety of crops, especially grains. The southern agricultural counties produce much of the state's agricultural wealth. The region also contains industrial concentrations and numerous large cities, including Detroit, the sixth largest U.S. city, Flint, Lansing, Grand Rapids, and Battle Creek.

The people. The diverse backgrounds of Michigan's people have been important in shaping the state's character. The percentage of foreign-born and first-generation citizens is small, however, even in Detroit.

Germans were the most numerous of the early non-English-speaking immigrants. Detroit had a large German community by the mid-1830s, as did several rural counties by 1850. The large Irish population was basically urban, although Irish farmers were found in southern Michigan and by 1860 in the Upper Peninsula. Dutch influences are still observable in western counties around Holland, where Dutch settlers pioneered successfully in 1847. Finns and the Cornish have been important in the economic and cultural life of the Upper Peninsula. Early Polish immigrants settled in rural areas until the 1890s, when a large number of Poles became concentrated in Detroit. The city's present-day population includes many people of Polish ancestry. More recently Hispanics, Asians, and immigrants from the Middle East have contributed to Michigan's ethnic mix.

The growing black population

The most significant population phenomenon of the 20th century has been the growth of the black population, from fewer than 16,000 in 1900 to well over 1,000,000 by the late 20th century. More than 60 percent live in Detroit, which has become nearly two-thirds black. Many blacks have moved to suburbs with long-established black neighbourhoods. They also have moved into older neighbourhoods and central city areas where the prices of homes are not exceptionally high. The majority of the newer Detroit suburbs, however, have very small black populations.

Michigan's religious history differs somewhat from that of many of the other Midwestern states. Because the first European settlers in Detroit were Roman Catholic, many immigrants of that faith were attracted to the city even before the large Irish, Italian, and Polish immigrations of the 19th century. Detroit was made a diocese in 1833 and an archdiocese in 1937. Other dioceses were established at Marquette, Grand Rapids, Lansing, Saginaw, Gaylord, and Kalamazoo. Of the Protestant denominations, Lutheranism has had many German and Scandinavian adherents, while Methodism has been important in both rural and urban Michigan. The first Dutch settlers were members of the Christian Reformed church, which was opposed to the state church of the Dutch. The Michigan group was highly independent and conservative in its doctrine and social mores. There also are large numbers of Presbyterians, Baptists, and Episcopalians. Altogether, Michigan has almost 200 Protestant denominations, some with only small memberships. The first Jewish immigrants to come to Michigan were of German background. In 1851 Detroit Jews founded a synagogue. Synagogues throughout the state reflect all forms of Judaism.

The economy. Michigan's economy, following a period of dependence in the 19th century on mining and lumbering, began in the second decade of the 20th century to be dominated by the automobile industry. Despite contributions from agriculture, tourism, forestry, and industries, producing goods such as office furniture, cereal, oil, chemicals, and pharmaceuticals, Michigan remained tied to the fortunes of the auto companies. The oil embargo of the late 1970s, combined with a dramatic increase in imports of foreign cars and a national economic recession, caused an economic crisis in Michigan. Between 1979 and 1982 the state's unemployment level climbed above 15 percent, the highest in the nation. Since then the auto industry has made a modest recovery, and Michigan's government and business leaders have initiated programs to expand the state's manufacturing base, to attract new high-technology firms, and to promote the service sector of the economy. Nevertheless, the automotive industry continues to dominate the economy, accounting for about one-third of all manufacturing employment and one-third of the total value added by manufacture.

Dominance of the automobile industry

The state government maintains many research divisions that offer assistance to various segments of the economy. The Department of Agriculture administers divisions of animal health, marketing, and soil conservation. The state tax structure attempts to avoid discrimination against the private sector. Michigan has had a workers' compensation law since 1912. Extensive programs of social legislation, including unemployment compensation, were undertaken in Michigan before World War II and have since been broadened.

Resources. Michigan is one of the nation's leading producers of iron ore, accounting for nearly one-third of the total national output, but changes in the U.S. steel industry have severely affected iron ore mining in the state. In 1984 all mining activities in the Menominee Range ceased, although production continued in the Marquette Range in Marquette county. Some deposits of gypsum for cement and of salt have not been touched, and the search for petroleum and natural gas continues.

Agriculture. Despite a continuing decline in the number of farms, Michigan's fertile soils make it a major agricultural state. The future of Michigan's agriculture is, however, dependent in part on prices. Many farmers work part-time in industry or for the government, an important factor in the economy of many rural communities. Corn is Michigan's major field crop, but the state is best known for its fruit production. Michigan leads the nation in the production of cherries and is a major producer of apples. The state usually is also the leading producer of dry beans in the United States. Christmas trees are another important agricultural product. Michigan forests also produce large quantities of pulpwood products, and the timber industry is undergoing a revival.

Fruit growing

Industry. Many reasons have been advanced for the rise of the automotive industry in Detroit. The city long had been noted for the manufacture of carriages, wagons, bicycles, and marine engines, and it had a large number of skilled and semiskilled labourers and an available supply of investment capital. Other cities, however, offered inducements equal to those of Detroit, and during the pioneering phase of the industry Detroit had a number of rivals.

Personalities played a major role in making Detroit the world's automotive centre. The industry began with Ransom E. Olds of Lansing, whose father manufactured gasoline engines. Olds's success by 1901 focused the attention of automotive people on Detroit. Henry Ford brought even greater fame. Organizing the Ford Motor Company in 1903, he was by 1908 confining production to the standard, low-priced Model T. He emphasized ease of repair, garage service, and the utility of his product. W.C. Durant of Flint recognized that the automobile would be purchased by persons who desired transportation rather than by faddist motorists. He hoped through the formation of a company with large-scale capital to speed up technological advances and thus capture a large portion of the untapped middle-class market. The General Motors Corporation stands as a testimony to his thinking.

The industry paid high wages, and many automotive workers became members of the middle class. Because of the high wages there was little interest in unionism until the Great Depression of the 1930s.

The automotive industry also played an important role in the growth of Flint, Pontiac, and Lansing. Grand Rapids attained prominence through its furniture industry, and Battle Creek is known nationally for its cereals.

Tourism

Tourism brings Michigan millions of dollars annually, depending on the protection of forests, rivers, lakes, and shorelines from careless development and pollution.

Transportation. The first railroad in Michigan, the Erie and Kalamazoo, was completed between Toledo and Adrian in 1836, and by 1870 the state had more than 1,600 miles of rail. The peak of 9,000 miles in 1910 has since been reduced by well over half. In contrast, the gains in motor vehicle ownership have been staggering. State and local governments have combined to give Michigan a

modern system of state highways, county roads, and city streets. The interstate express system has been built largely with federal assistance.

Air passenger service in Michigan began in 1926. Detroit Metropolitan Airport services millions of airline passengers annually, and there are some 20 other major airports throughout the state and several small commuter airlines. Much of the airfreight of metropolitan Detroit is handled at the Willow Run Airport, a facility constructed as a bomber plant during World War II.

The waterways carry tremendous tonnage, with many of the state's exports shipped from Detroit's harbour. Ores and other bulk materials destined for the interstate trade are generally sent by water.

Administration and social conditions. *Government.* The constitution of 1963 included a number of provisions to streamline government and make it more responsive to the problems of a modern industrial and urban society. Amendments may be submitted to the electorate by the legislature or by initiative petitions, but all amendments must be approved by a referendum of the voters.

Executive power is vested in the governor, who serves for four years. The governor is nominated by a primary election, but the lieutenant governor is chosen by party convention. Administrative commissions appointed by the governor are responsible to the executive and to several advisory commissions. The majority of the important governmental services are combined under departments responsible to the governor. A few of these bodies, such as the State Highway Commission, must be bipartisan.

The legislative branch comprises the Senate of 38 members elected for four-year terms and the House of Representatives of 110 members elected for two-year terms. The legislative districts are redefined by a special bipartisan commission after each federal census. The highest court is a seven-member state supreme court. The state also has a court of appeals, circuit courts, probate courts, and courts of limited jurisdiction that are specified by the legislature.

Local government

Michigan has more than 2,600 local governmental units, including counties, municipalities, townships, school districts, and such special districts as park authorities. Although the majority of counties are governed by a board of supervisors, the home-rule privilege allows larger counties to entrust management to county commissioners. Extensive privileges of home rule are authorized for the cities as well. School districts are classified by population and enjoy differing privileges of government.

The precinct is the primary unit of political party organization, and the precinct delegates carry considerable importance in the annual party conventions, where candidates are nominated for lieutenant governor, attorney general, secretary of state, and the members of the boards that govern the state system of higher education and the State Board of Education. Justices of the Supreme Court are nominated on a nonpartisan ballot. The state party conventions also select delegates to the national presidential conventions.

Black political interest grew out of the Civil Rights movement, and civil rights groups have encouraged the political awareness of black voters. Blacks have been nominated for major state and local offices by both parties; Coleman Young was elected the first black mayor of Detroit in 1973. William Lucas, nominated in 1986 by the Republican Party, was the first black to run for governor of Michigan; he was defeated by incumbent James Blanchard. Unions have been very active in Michigan politics, and the United Automobile Workers has endorsed candidates at the municipal, state, and national levels.

The personal and corporate income tax is the major source of general revenue, though federal aid and receipts from the sales tax and property tax play a major role in assuring the financial structure of the state. Gasoline and vehicle tax revenues are reserved exclusively for highways.

Education. Many of the legislature's major appropriation bills are for support of Michigan's numerous public institutions of higher learning, including many community colleges. Large sums, however, are returned to the local districts for public elementary and secondary schools. Many of these schools, both in the inner-city areas and in the suburbs, are frequently studied to identify problems, and experiments in improved methods of instruction are often conducted by educators and psychologists from the state's universities. Adequate and equitable funding for school districts is a continuing concern. Beginning in 1988, Michigan offered state residents an opportunity to invest in the Michigan Educational Trust Fund. The program guarantees tuition for an investor's child who later enrolls in a Michigan college or university.

Higher education

In 1817 Judge Augustus Woodward, one of the major figures in the state's early history, conceived the idea of a "Catholepistemiad," an academy of universal knowledge. This was achieved in some measure in 1837, when the University of Michigan opened in Ann Arbor. It has since come to be widely regarded as one of the nation's finest universities, with programs on both the undergraduate and graduate levels. In 1849 a teacher-training institution, which later became Eastern Michigan University, began instruction at Ypsilanti. In 1855 the Agricultural College of the State of Michigan, now Michigan State University, was established in East Lansing. The latter shares much of the wide regard of its rival in Ann Arbor, having moved far beyond its early identification with agriculture in many areas of research and scholarship. The Michigan Technological University at Houghton, a state institution, was established in 1885 as the Michigan Mining School. In 1956 the state acquired Wayne University, a Detroit municipal university. Wayne State University, as it was renamed, has fostered much educational experimentation, and in the 1960s its campus and physical plant became landmarks in U.S. educational architecture through the designs of the American architect Minoru Yamasaki.

Michigan is also the home of widely recognized specialized schools. In 1927 the School of Music was founded in Interlochen, the forerunner of the National Music Camp that now offers instruction in music, dramatics, and related arts. The Cranbrook and Kingswood schools in Bloomfield Hills, designed by Eliel Saarinen, pioneered advanced art courses for students of high-school age.

Health and welfare. The Department of Public Health regulates the operation, construction, and licensing of health care facilities, including hospitals, nursing homes, homes for the aged, and long-term care units. Michigan devotes a major portion of its annual budget to social programs. A psychiatric hospital in Kalamazoo that received patients as early as 1859 was a forerunner of a number of other institutions caring for the needy. Since 1936 many of the state welfare services have followed a federal formula in transmitting aid to units of local government. Payments to recipients of old-age assistance, for example, and aid to dependent children, the blind, and the disabled have required special personnel at the state level, though disbursal of most funds is done locally.

Cultural life. Detroit dominated the cultural life of early Michigan, and the multifaceted backgrounds of its population gave it a cosmopolitan atmosphere. The state's first traveling theatrical companies performed in Detroit, and an opera house was erected there before the Civil War. In 1819 the Young Men's Society was organized by Lewis Cass, who led many early civic endeavours, to promote debates, lectures, and general intellectual life.

The cultural milieu

The pioneer farmers, however, had little time for cultural pursuits. The community dance that eased the strains of plowing or harvesting and the county fairs that provided a ritual summation of the year's achievements were important. Michigan's culture, then, tended to become broadly based on popular life, having many of the aspects of "mass culture." In the major parks the city of Detroit promoted band concerts and, later, symphony concerts to bring other types of music to thousands. The Detroit Institute of Arts, maintained municipally, has always emphasized exhibits that appeal to a broad public.

Michigan celebrated its sesquicentennial in 1987 with historical events as well as programs that explored contemporary issues. The Michigan Historical Commission has designated sites of historic importance, such as the location of early settlements and the first home of the Ford Motor Company. Many cities emphasize Homecoming Day, usually the anniversary date of the incorporation

Folk and regional festivals

of the community. Local folk festivals have been given a greater emphasis. Holland's Tulip Time Festival, held each May, has become an event of more than local importance. The annual Bavarian Festival, in Frankenmuth, appeals to others besides those of German background. Traverse City sponsors a popular National Cherry Festival. Several cities in the north commemorate the lumbering era with Paul Bunyan Days. Many ethnic groups in Detroit and other cities sponsor folk festivals that recall their cultural ties to Europe. The Museum of African-American History in Detroit is a continuing reminder of the black population's distinguished heritage.

The Detroit Institute of Arts, founded in 1885, holds one of the nation's major collections. The Muskegon Museum of Art, founded in 1911, the Kalamazoo Institute of Arts, and the Grand Rapids Art Museum have won recognition. Lansing is the home of the Michigan Historical Museum, famous for its military and Indian collections, while many county museums commemorate local history. In 1988 the state museum, state library, and state archives moved into a new building in Lansing that showcases the state's cultural resources.

The state library was founded in 1828. A state library board was created in 1936, and major public libraries are found in the larger cities. Among the outstanding special libraries are the William L. Clements Library at the University of Michigan, specializing in American history; the Burton Historical Collection of the Detroit Public Library, specializing in local history and genealogy; the Clarke Historical Library at Central Michigan University, with collections on state and regional history; and the Gerald R. Ford Museum in Grand Rapids and the Gerald R. Ford Library in Ann Arbor.

Outdoor recreation in Michigan is dominated by woods and water. As early as the 1830s, the Great Lakes were a favourite vacation lure for Eastern travelers. Today residents of southern Michigan and tourists from other areas flock to the state's shores and forests to swim, fish, hike, and hunt. Beginning in 1919, the state developed a state park system, and there are now some 90 state-operated parks and recreation areas. In addition, Michigan's 3,900,000 acres of state forest and 25,000,000 acres of national forest constitute the largest expanse of public forestland in any state east of the Rocky Mountains.

Organized team sports began with the establishment of baseball teams in several Michigan cities during the late 1850s and early 1860s. In 1881, when the Detroit team began to compete nationally, the state's continuing love affair with the Tigers began. Football is the predominant college sport, with the heated traditional rivalry between the University of Michigan and Michigan State University being the highlight of each season. Professional football, hockey, and basketball teams from the Detroit area claim loyal fans.

HISTORY

In the 17th century the Indian population of what is present-day Michigan was estimated at between 12,000 and 15,000. The majority of these Indians, including the Ottawa, Ojibwa, Miami, and Potawatomi, belonged to the Algonquian linguistic group. A lesser number, located primarily in southeastern Michigan, were Huron and Wendat (Wyandot). The Ottawa and Ojibwa aided the French in the development and expansion of the fur trade. The Ottawa, with their commercial interests, had developed a type of canoe that was highly serviceable in the Great Lakes area. The Potawatomi Indians were identified more with the culture of the woods. The Huron were the most advanced in their agricultural practices. All the Indians of the Michigan area lived in small communities and were unfamiliar with the concept of private property.

French heritage

Settlement. Étienne Brulé, the first European to visit the area (1622), was the forerunner of numerous explorers, missionaries, and fur traders who paved the way for French control over Michigan. The oldest community in Michigan is Sault Sainte Marie, founded in 1668 at a site where in 1641 missionaries held services for 2,000 Ojibwa. In 1701 Antoine de la Mothe Cadillac established Detroit as a fur-trading centre and administrative post; it soon became the leading French community in the entire Great Lakes area. The French, and later the English and Americans, also maintained a fort at the Straits of Mackinac.

In 1760 the French garrisons were surrendered to an English force, and in 1763, by the Treaty of Paris, England acquired jurisdiction over Canada and the French empire east of the Mississippi River except for New Orleans. Under English rule Michigan remained a part of Canada. During the American Revolution Detroit was a major supply centre for British troops, who raided the Kentucky country until 1779, when the British general Henry Hamilton was captured.

U.S. territory. Although Michigan had been awarded to the United States in 1783, the British refused to leave Detroit and other major military posts until 1796. In 1787 it was made a part of the newly created Northwest Territory. Indian opposition to U.S. rule in the area was ended by the victory of Anthony Wayne at the Battle of Fallen Timbers, near present-day Toledo, Ohio. After 1796 the Americanizing of the regions was accomplished within a few years. Detroit became the capital of the Michigan Territory, which was separated in 1805 from Indiana. Although the first governor, William Hull, surrendered Detroit to the British early in the War of 1812, American rule was restored late in 1813 by the victory of Commodore Oliver Hazard Perry at the Battle of Lake Erie.

Process of Americanization

The real growth of Michigan Territory began soon after the war, with the new governor, Lewis Cass, encouraging settlement and promoting growth. New modes of transportation were even more significant. In 1818 steamship navigation linked Detroit and Buffalo, N.Y., inaugurating a new era in lake transportation. Cass's crude highway chain from Detroit to Chicago, Saginaw, and Port Huron helped to establish the patterns of settlement in the interior. Completion of the Erie Canal in 1825 made Detroit a leading distribution point for settlers seeking new homes in the Great Lakes area.

Statehood and growth. Michigan was anxious for statehood so that it might undertake a more ambitious program of internal improvements. The first constitution was enacted in 1835, but statehood was delayed until 1837 by the Toledo War, a boundary dispute with Ohio. In return for relinquishing its claims to the mouth of the Maumee River, at Toledo, Michigan was awarded the western half of the Upper Peninsula as well as the eastern portion, which was historically part of the territory.

The state grew rapidly through the 1840s and '50s. Thousands of prospective agricultural settlers, including many foreign-born, established new homes in the state. Detroit and other leading cities profited, and in the 1840s the rich iron and copper resources of the Upper Peninsula became known. The state capital was moved from Detroit to the more central location of Lansing in 1847. Tension over the slavery issue resulted in the formation of the present-day Republican Party at Jackson in July 1854.

Throughout the American Civil War, Michigan made major contributions to the Union cause, losing some 14,000 of its 90,000 men who served. A black regiment from Michigan included enlistees from many states and from Ontario. The Republican Party became dominant after the war. In the 1890s Hazen Pingree, mayor of Detroit and subsequently governor, as well as other leaders, implemented progressive legislation.

Increasing economic tempo

The 20th century. Before 1900 all of Michigan's 83 counties had been settled, and agriculture, lumbering, mining, and manufacturing created a stirring economic tempo. Throughout the 20th century, however, the economy has been dominated by the automotive industry. During World War I, industrial production at all levels was intensified. The emergence of new problems connected with urban and industrial growth was recognized by features of the state's third constitution, approved in 1908.

The Great Depression was unusually severe in Michigan, the industrial products of which were not among the necessities of life; unemployment and deflation were far above the national averages. In 1932 Michigan departed from the Republican fold, thereafter becoming one of the doubtful, or swing, states, and organized labour became a powerful political and economic factor. In 1937

the United Automobile Workers became the bargaining agent for production workers at General Motors Corporation, and by the outbreak of World War II it was the dominant union in all automotive plants. During the war Detroit became known as the Arsenal of Democracy, and industrial production continued at a peak afterward to restock the nation with new cars and other war-depleted consumer goods.

Since World War II Michigan has experienced both racial polarization, as attested by the Detroit riots of 1943 and 1967, and strong efforts to equalize opportunity for minorities, the handicapped, and women. The 1963 Michigan constitution was the first in the nation to provide for a Department of Civil Rights.

The postwar years were also a period of explosive development in the suburbs and rapid expansion of the state's highway system. Inner cities, however, beginning in the late 1950s, declined in population, industries, and services. Detroit's Renaissance Center, a high-rise riverfront hotel, retail, and business development, stands as a symbol of the state's dedication to making its cities attractive and livable.

The severe recession of the late 1970s and early 1980s caused widespread unemployment, business failures, and cuts in state government services. Since then, government, business, and education have cooperated in their efforts to attract new industry, broaden opportunities for young people, strengthen the work force, and promote the expanding tourist industry. (S.Gl./Ri.J.H.)

Minnesota

Minnesota lies near the heart of the North American continent. Its waters flow southward through the Mississippi River to the Gulf of Mexico, eastward through the Great Lakes to the Atlantic Ocean, and northward via the Red and Rainy rivers to Hudson Bay. Minnesota, which became the 32nd state on May 11, 1858, received its name from the Dakota (Sioux) word for the Mississippi's major tributary in the state, which means "Sky-Tinted Waters." St. Paul is the state capital.

The state's 84,402 square miles (218,601 square kilometres) are bounded on the north by the Canadian provinces of Manitoba and Ontario, on the east by Lake Superior and Wisconsin, on the south by Iowa, and on the west by South and North Dakota. A small extension of the northern boundary makes Minnesota the most northerly of the 48 coterminous U.S. states. This irregularity is the result of a general boundary agreement with Great Britain before the area had been carefully surveyed.

Minnesota is a land of extensive woodlands, fertile prairies, and innumerable lakes, more than 12,000 of which are larger than 10 acres (four hectares) in area. The nearly 5,000 square miles of inland fresh water are a dominant feature of the Minnesota life-style. Its climate is continental, with cold winters and warm summers. About one in four Minnesotans is at least part Scandinavian, but Germans constitute the single largest ethnic group in the state. In the past the Minnesota economy has been dominated by the production and processing of its timber, iron ore, and agricultural resources. While agriculture remains important, the state's economy has become much more diversified since World War II, with the rapid growth of specialized manufacturing and services.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief and drainage.* Minnesota's elevations range from 602 feet (184 metres) above sea level at Lake Superior to 2,301 feet (701 metres) at Eagle Mountain, about 12 miles from Lake Superior's north shore. Most of Minnesota has been covered by glaciers several times, and the land's surface has been shaped by the alternate freezing, thawing, and movement of those glaciers. Prominent reminders are the rolling farmlands, thousands of lakes, steep hillsides, and flat glacial lake and outwash plains. Minnesota's rich prairie soils developed on the finely ground mineral materials left by the retreating glaciers. The majority of Minnesota's lakes are located in the areas of glacial moraine, where glaciers deposited hills of sand and gravel. Lakes of more than 100 square miles in area are Red Lake, Mille Lacs Lake, Leech Lake, and Lake Winnibigoshish; Lake of the Woods and Rainy Lake are shared with Canada. With some 160 miles (260 kilometres) of coastline, Minnesota shares Lake Superior, the largest freshwater lake in the world, with Ontario, Michigan, and Wisconsin.

Effects of the glaciers

The largest glacial lake plain (more than 100,000 square miles) was formed by Lake Agassiz, which held the meltwaters as the latest glaciers retreated northward some 8,000 years ago. The southern part of the former lake bed lies along the Minnesota–North Dakota border and is known as the Red River valley. Red Lake and Lake of the Woods, as well as Lake Winnipeg in Canada, are all remnants of this huge body of glacial meltwater. Its southward drainage created the wide valley of the Minnesota River, the flow of which eventually reversed as ice blockage to the north melted. In northeastern Minnesota there are stream valleys and deep, clear lakes that were scoured by glaciers from the granite bedrock. Extreme southeastern Minnesota was the only part of the state to escape glaciation. There, streams have cut their way through layers of limestone, leaving extensive caverns beneath the surface and steep, rocky bluffs rising high above the valleys.

Soils. The most fertile soils in Minnesota are in the grasslands of the south and west, rich in organic matter and high in soluble minerals. Soils that have formed beneath the coniferous forest in northeastern Minnesota are light coloured, acidic, and low in organic matter, while those formed in the hardwood areas are intermediate in colour and natural fertility.

Climate. Great variations of temperature and growing season occur not only seasonally but also from one part of the state to another. This reflects the fact that Minnesota stretches from the edge of the subarctic forest to the heart of the Corn Belt. During June and July the sun shines more directly over southern Minnesota than over the equator, while in the north frost is possible in any month.

In July average daily maximum temperatures range from 85° F (29° C) in southern Minnesota to 70° F (21° C) along the shore of Lake Superior. Average daily January highs range from 25° F (−4° C) in the south to 15° F (−9° C) in the north; minimums are from 5° F (−15° C) to −5° F (−21° C). The average frost-free periods vary from less than 90 days in parts of the north to more than 160 days in parts of the south. The average annual precipitation ranges from more than 30 inches (750 millimetres) in the southeast to less than 20 inches in the northwest. Average seasonal snowfall varies from less than 40 inches in the western part of the state to more than 70 inches in the extreme northeastern tip. Many parts of Minnesota have continuous snow cover for at least 90 days, from about mid-December to mid-March.

Plant and animal life. Minnesota stands astride one of the major physical geographic boundaries in the world, the sharp transition from forest to prairie in the heart of North America. The natural vegetation of Minnesota may be divided into three general categories: needleleaf forests, hardwood forests, and tallgrass prairie. The needleleaf forests originally occupied the northeastern third of the state and included pine, spruce, and fir, with tamarack in the bog areas. A belt of hardwoods extends from southeastern Minnesota northeasterly to the Canadian border, passing through the Twin Cities and lying immediately to the south and west of the coniferous forest. The hardwood forest is known as the big woods and averages some 40 to 80 miles in width. It consists primarily of oak, maple, and basswood, with ash, elm, cottonwood, and box elder along the stream valleys. South and west of the hardwood forests lies the tallgrass prairie.

Belts of forest and prairie

Mammals commonly found throughout the state include deer, foxes, raccoons, porcupines, minks, weasels, skunks, muskrats, woodchucks, and squirrels. Black bears, moose, elk, wolves, coyotes, lynx, bobcats, otters, and beavers are found almost entirely in the north. Common year-round birds include chickadees, woodpeckers, grosbeaks, nuthatches, cardinals, sparrows, and jays. Favourite migratory songbirds include robins, orioles, thrushes, meadowlarks, and red-winged blackbirds, the state's most

common bird. Migratory waterfowl include ducks, geese, gulls, coots, herons, and egrets. The common loon is the official state bird. In addition to ducks and geese, other game birds include grouse, quail, partridge, wild turkeys, and imported ring-necked pheasants. Important raptors include hawks, eagles, owls, and ospreys.

The walleye is designated as the state fish and is the most popular object of anglers. Other important game fish include the northern pike, muskellunge, bass, lake trout, crappie, sunfish, and eelpout. Brown and rainbow trout thrive in many streams. The deep, cold waters of Lake Superior contain lake trout, whitefish, coho and chinook salmon, steelhead, smelt, herring, and ciscoes. The timber rattlesnake is found in several southeastern counties.

Settlement patterns. To serve Minnesota's growing agricultural, forestry, and mining activities in the 19th century, a network of towns emerged across the landscape. In the latter half of the 20th century many of the smaller communities atrophied, while the larger communities extended their influence over wider areas. In general, population densities are greatest in the east and south, declining toward the north and west.

During a century of white occupancy, virtually all of the prairies were cultivated. The coniferous forestlands, mostly cut by 1920, have become covered again by aspen, birch, and jack pine, while much of the big woods has been cleared for crops and pasture. Minnesota reached its peak in cultivated farmland in 1945. Since then the agricultural frontier has retreated, and farms have been abandoned in the less fertile areas in north central and northeastern Minnesota. The big woods area has become primarily a dairying area. Within 100 miles (160 kilometres) or so of the Twin Cities, dairying has intensified, but, beyond that, dairying has declined in importance. The prairie areas of southern and southwestern Minnesota have a characteristically Corn Belt crop and livestock agriculture.

The people. The New England Yankees of English, Scottish, and Irish descent who first settled Minnesota were entrepreneurs who helped establish the institutions and many of the traditions that remain important in Minnesota. The first major immigrant groups in the latter half of the 19th century were Germans, Swedes, and Norwegians who cut the trees, built the railroads, and became the farmers, tradesmen, and professionals.

German settlers dominated the push up the Mississippi, continuing into the central and south central parts of the state. Norwegian settlers moved westward across the southern tier of counties, forming the major ethnic group in west central Minnesota and the Red River valley. The major areas of Swedish settlement are in several counties immediately north of the Twin Cities and scattered locations in west central and northwestern Minnesota. Substantial numbers of Finns live in northeastern Minnesota, Poles in southeastern and central Minnesota, Bohemians south of the Twin Cities area, Irish across the south, French and French Canadians just north of the Twin Cities and in northwestern Minnesota, Dutch and Flemish in parts of southwestern Minnesota, Icelanders in northwestern Minnesota, and Danes, Welsh, and Swiss in scattered pockets.

The ethnic blend

Each ethnic group brought with it its religious traditions. Central and south central Minnesota are heavily Roman Catholic, reflecting the German, Polish, and Bohemian populations. Other Germans, as well as most Scandinavians, are Lutheran. Many ethnic clusters have retained a degree of homogeneity in the rural areas. They have been sources of population movement to urban areas, but they have attracted few in-migrants to alter the original stock.

By 1890 most of the good agricultural land had been claimed. Thus, most immigrants who arrived during the next few decades sought a livelihood in the Twin Cities area or on the iron ranges, where employment opportunities were expanding. The Twin Cities area, in particular, grew rapidly as the state's major centre for cultural contact and variety. These later immigrant groups included Finns, Italians, Slovakians, Croatians, Serbs, Greeks, Jews, Ukrainians, Russians, American blacks, and Hispanics, as well as a continued flow of northern Europeans.

The American Indian population is primarily Ojibwa, about one-half of whom live in the Twin Cities; most of the remainder live on reservations in rural Minnesota. Slightly more than 1 percent of the state population is black, about 90 percent of whom live in the Twin Cities. Hispanics constitute a sizable community. Since the mid-1970s, refugees from Vietnam, Laos, and Kampuchea have been added to the Twin Cities area population.

The economy. The economic growth of early Minnesota was related closely to the exploitation of its primary natural resources: soils, iron ore, and timber. These activities, in turn, stimulated the growth of such ancillary activities as railroads, processing of natural resources and agricultural products, and services. During the late 1960s and early '70s, activities based on natural resources, as well as railroads and associated manufacturing, began to decline. Agriculture, however, is still Minnesota's largest industry.

Agriculture. Agriculture in Minnesota falls into four categories that are closely related to climate and soil type. Minnesota's most valuable and productive farmland lies across the southern quarter of the state, mostly an area of dark, fertile prairie soils and hot, humid summers. Corn (maize), which occupies the greatest acreage, is fed to cattle and hogs; soybeans are the major cash crop. In the Red River valley the growing season is shorter and the humidity is lower than in southern Minnesota, making small grains and specialty crops more profitable. Major crops are wheat, barley, sugar beets, sunflowers, potatoes, and flax.

Dairying dominates the hilly big woods region from southeastern to west central Minnesota. Milk and milk products are the major sources of farm income in this region, with feed crops being important. Soybeans and potatoes are grown as supplemental cash crops. Large-scale turkey production is important in several localities. In northeastern Minnesota, where soils are thin and acidic, agriculture is much less important than in other parts of the state.

Mining. Iron ore accounts for more than 90 percent of the value of all minerals produced in Minnesota. The Mesabi Range, the largest of three iron ranges in the state, began production in 1892 and at its peak produced one-fourth of the world's iron ore. By the late 1950s, however, most of the high-grade natural ores of the Mesabi had been depleted. A process was developed at the University of Minnesota for extracting iron from the abundant but low-grade taconite rock. To encourage the heavy capital investment required, the voters of Minnesota approved a constitutional amendment in 1964 that guaranteed the taconite industry a tax-free period of 25 years. This resulted in a brief revitalization of the iron ore industry in Minnesota. In the 1980s, however, low-cost foreign steel reduced American steel production, and lower-cost foreign ores replaced Minnesota taconite in American steel mills.

Iron extraction

Other mining activities in Minnesota include granite and limestone quarrying and sand and gravel extraction. There are no coal, oil, or natural gas resources in Minnesota, and geologic formations are such that the discovery of these minerals is highly unlikely.

Industry. Minnesota's earliest industries facilitated exploitation of the state's natural resources with the manufacture of implements, machinery, tools, and hardware. Ancillary activities included processing, packaging, transporting, financing, insuring, and providing the necessary infrastructure for these industries, all of which developed into a complex interdependent economic network. The agricultural servicing industries have diversified into a wide range of consumer products and have grown into major corporations with worldwide markets. The major financial institutions, insurance companies, and merchandisers likewise have diversified and expanded their market areas.

The major changes in Minnesota's economic structure have come from high-technology and non-resource-based industries. Other growing sectors of the economy are printing and publishing, health care, scientific instruments, industrial chemicals, and recreational equipment.

The lumbering industry declined rapidly after the turn of the century as the pine forests were depleted and much of the natural regrowth of aspen and birch had limited commerical value. In the latter half of the 20th century, however, Minnesota's forest industry has been revitalized with the growth of the wood pulp and waferboard indus-

tries. Pine, balsam, and spruce are harvested for pulpwood, while aspen, once considered a "weed" tree, has become the preferred species for waferboard manufacturing and accounts for about 70 percent of the commercially harvested wood in Minnesota. From the 1880s to about 1920, Minneapolis was known as "the mill city," producing more flour than any other city in the world. After 1920 Buffalo, N.Y., surpassed Minneapolis because of its proximity to eastern markets. While flour is no longer produced in Minneapolis, the major milling companies retain their headquarters there.

The milling industry

Transportation. The movement of people and goods in Minnesota and the upper Midwest is centred on the Twin Cities area. Regional and transcontinental rail and highway systems radiate outward from the Twin Cities, tying the towns and hamlets into one interdependent network. The rail system of northeastern Minnesota brings iron ore and taconite products for transshipment by boat at the Lake Superior ports of Duluth and Superior. Wheat from the Dakotas and Montana also has been an important product transshipped from rail to boat at Duluth. Since the opening of the Great Lakes waterway to ocean vessels in 1959, products of the upper Midwest are carried directly to locations throughout the world.

River transportation was the first important mode for the movement of both people and goods in many parts of the state. Barges on the Mississippi carry bulk products to and from the major inland ports at St. Paul and Minneapolis. Carried upstream are such bulk products as coal, oil, and salt; grain, sand, and gravel are transported downstream.

The Twin Cities area, also the air hub of the upper Midwest, is served by several commercial airlines. The Minneapolis–St. Paul International Airport is supplemented by a satellite network of additional airports around the state.

Administration and social conditions. *Government.* Minnesota's first state constitution was ratified by Congress at the time of statehood in 1858. In 1974 this document was restructured and reworded to improve its utility without altering its meaning. In cases of constitutional law the original document is the final authority.

Minnesota's constitution provides for an executive branch comprising a governor, a lieutenant governor, a secretary of state, an auditor, a treasurer, and an attorney general. These six state officials are nominated with political party designation and elected by statewide ballot. There are more than 100 administrative departments and independent agencies, boards, commissions, and other bodies.

The state's bicameral legislature consists of a 67-member Senate and a 134-member House of Representatives that meet in regular session in odd-numbered years. Senators are elected to four-year terms, representatives to two-year terms.

The judicial branch comprises the Supreme Court, district court, probate court, county courts, municipal courts, and justices of the peace. Nine justices constitute the Supreme Court; each is elected for a six-year term.

The state's 87 counties range in size from 155 square miles of land area to 6,092 square miles and in population from less than 5,000 to about 1,000,000 people. Counties and municipalities provide most of the local governmental services, but townships assume some authority for planning and zoning and for maintenance of public works, parks, and hospitals. A number of school districts have been consolidated, and special districts have been established to provide for waste management, water supply, fire protection, parks, airports, soil and water conservation, and other interjurisdictional needs.

The Twin Cities Metropolitan Council, the members of which are appointed by the governor, is responsible for the development of certain areawide services that local government is unable to provide, including sewage and water systems, transportation, and major land uses. It plays a coordinating and regulatory role among the more than 350 local governmental jurisdictions within the Twin Cities area.

Minnesota politics

Politics and social issues. From the beginning, Minnesota politics has been characterized by recurring waves of protest and reform that have spawned such national groups as the Grange, the Greenbackers, the Antimonopolists, the Farmers Alliance, the Populists, the Prohibitionists, and the Nonpartisan League. Each of these movements brought about social reforms and influenced the major political parties. Both major Minnesota parties in the late 20th century are amalgams from this tradition. The Democratic-Farmer-Labor Party was formed in 1944 between the more traditional Democrats and the reformist Farmer-Labor Party. The Independent-Republican Party was created in an effort to attract more of the substantial but diverse independent vote in Minnesota.

The political environment emerged from the traditions of the original New Englanders who brought their town-meeting form of government to this new frontier. That foundation was reinforced by the Scandinavian and German immigrants, with their ambition and high regard for education. Government has always been accepted as the legitimate means for public decision making in Minnesota, and business has played an important role as a strong participant in public decisions.

The traditions of citizen involvement can be seen in the many neighbourhood and community organizations and ad hoc issue-related groups in the state and in the relatively large number of Minnesotans of national political prominence. Minnesota has been a leader in such national movements as those to guarantee the rights of women, homosexuals, and American Indians.

Education. Approximately 90 percent of Minnesota's elementary and secondary students attend public schools. School districts vary widely in size and resources, with the larger and wealthier generally located in the major urban centres. Small rural school districts have often consolidated—or collaborated without consolidating—in order to provide a full range of curriculum opportunities and essential services. On average, more than half of local school district revenues come from the state, with support ranging widely based on local needs. Minnesota ranks among the top states in the proportion of its students graduating from high school, in standard test scores for high school graduates, and in the proportion attending higher educational institutions.

The University of Minnesota was established in 1851, with its main campus in Minneapolis and a smaller campus in St. Paul. Smaller branches of the university are located in Duluth and Morris, with four-year programs, and in Crookston and Waseca, with two-year programs.

The state university system operates campuses in several locations. One of these, Metropolitan State University, in the Twin Cities, is a "college without a campus" that utilizes diverse physical facilities throughout the area to bring higher education to the neighbourhoods. In addition, the state operates large systems of community colleges and of post-secondary vocational-technical schools. Several private four-year colleges supplement the public system.

Health and welfare. Minnesota is high among the states in the quality of health and welfare services. The state's high standard of general medical services, its extensive children's health and welfare programs, and its innovative approaches to health maintenance, drug- and alcohol-abuse treatment, and care for the elderly have all been praised.

The Twin Cities and Rochester serve as national health care centres. The Mayo Clinic in Rochester has served patients from around the world since the late 19th century. The University of Minnesota Hospitals in the Twin Cities area has been a pioneer in medical research, while numerous hospitals across the state provide an effective network of medical care.

Cultural life. Cultural life in Minnesota is highly diversified and seasonal. Many activities are oriented toward the outdoors; they include swimming, boating, canoeing, camping, hunting, and fishing. Popular winter sports include downhill and cross-country skiing, snowmobiling, and ice fishing; ice hockey is most commonly played indoors. St. Paul celebrates winter with its annual Winter Carnival, while Minneapolis celebrates summer with its Aquatennial. The State Fair is a major summer attraction of the Twin Cities. Community festivals are abundant throughout the state year-round.

Prevalence of outdoor life

Whereas "outstate" Minnesota is the outdoor playground

Ice Palace at Winter Carnival, St. Paul, Minn.
©Phillip M. James from TSW—CLICK/Chicago

for the state, the Twin Cities area serves as the centre of cultural institutions. The best-known musical organizations are the Minnesota Orchestra, which was formed in 1903 as the Minneapolis Symphony Orchestra, the St. Paul Chamber Orchestra, and the Minnesota Opera Company. Civic orchestras and colleges and universities throughout the state make substantial contributions to the arts within their communities and regions.

The Twin Cities area has several resident professional theatres. The best known is the Tyrone Guthrie Theatre Company, formed in 1963. The Children's Theatre Company is nationally recognized as one of the finest of its kind. The Minnesota Dance Theatre is the most prominent resident dance company in the Twin Cities.

The Minneapolis Institute of Arts, the Walker Art Center, and the Minnesota Museum of Art are among the most important art museums in the state. Other major museums are the Minnesota Science Museum, the Bell Museum of Natural History, the Minnesota Historical Society, the American Swedish Institute, and the Planetarium of the Minneapolis Public Library.

Sinclair Lewis and F. Scott Fitzgerald are celebrated Minnesota-born authors. Minnesota's pioneer days are remembered in the books of Laura Ingalls Wilder and Ole Rölvaag.

HISTORY

Until the middle of the 19th century, two major Indian tribes occupied what is now Minnesota: the Ojibwa (Chippewa) in the north and east and the Sioux (the popular name for the Dakota) in the south and west. Between the time of European exploration and statehood, the Ojibwa occupied the forested areas of the state and pushed the Sioux southward and southwestward onto the prairie. Indians of tribes from as far away as the Appalachians and the Rocky Mountains met in a sacred place of peace in southwestern Minnesota to quarry a hard red rock that was used for making peace pipes; today this area is preserved as the Pipestone National Monument.

Diversity of Indian tribes

European settlement. Investigation of the Kensington Stone, found in west central Minnesota in 1898 and bearing inscriptions allegedly made by Norsemen who penetrated the continent in the 14th century, has proved it to be a forgery. The earliest verifiable Europeans in the area were 17th-century French explorers who were searching for a Northwest Passage. The first white settlement was made where the French fur traders known as voyageurs had to leave Lake Superior to make a nine-mile portage around the falls and rapids of the Pigeon River. Before the American Revolution this outpost, known as Grand Portage, was the hub of an enormous commercial empire stretching 3,000 miles from Montreal to Canada's northwestern wilderness. It was the inland headquarters of the North West Company, which trapped beaver and marketed their pelts, and the meeting place each July and August for fur buyers and sellers. Grand Portage became U.S. territory after the Revolution but did not pass into American hands until 1803, when the North West Company moved 30 miles up the Lake Superior shore to Fort William (now Thunder Bay), Can. Today Grand Portage is a national monument, and part of the fur traders' route east of International Falls has been preserved as Voyageurs National Park.

The first permanent U.S. settlement was at Fort Snelling, a military outpost established in 1819 overlooking the junction of the Mississippi and Minnesota rivers; the site has been restored as a state park. Immigration into the region was slow during the first half of the 19th century, but, once the value of the vast forestlands of northern and central Minnesota was realized, lumbermen from New England led a large wave of permanent settlers.

Territory and state. That part of Minnesota east of the Mississippi River was part of the original Northwest Territory, which came under the jurisdiction of the Ordinance of 1787; the part of the state that lies west of the Mississippi was part of the Louisiana Purchase in 1803. Minnesota became a territory in 1849, its boundaries at that time reaching as far west as the Upper Missouri River, but most of its approximately 4,000 white settlers were located in the Fort Snelling–St. Paul area in the eastern part of the territory. The lumber industry developed rapidly, and major sawmills were soon built at Stillwater, on the St. Croix River, and at the Falls of St. Anthony, on the Mississippi River. The two villages at the falls were merged in 1872, with the village of St. Anthony on the east bank being absorbed into the larger and more aggressive city of Minneapolis on the west bank.

Ties with Canada were important during the early settlement period. In 1811 a colony had been established in the lower Red River valley, near modern Winnipeg. As there was little effort to mark and enforce the international boundary, goods and people flowed unhindered between the two countries. Immigrant groups that came into Minnesota via this route were Canadians of English, Scottish, Scotch-Irish, and French extraction. Because it was much easier to supply this area from Minnesota than from eastern Canada, supplies were shipped from St. Paul via St. Anthony to Fort Garry and other Red River valley settlements. As a result of this lucrative trade, people from both sides of the border sought U.S. annexation of western Canada, known as Rupert's Land. This received little support in Congress from Southern states concerned with maintaining the sectional balance. Great Britain effectively undercut any Canadian desire to defect to the United States with the British North America Act of 1867, which brought about the formation of the Dominion of Canada, giving Canada self-governing authority. The efforts of Minnesota expansionists ended in 1870 when Canada established the province of Manitoba and sent troops to Winnipeg.

When Minnesota became a state in 1858, its boundaries were cut back from the Missouri River to the Red River. In 1861, Minnesota was the first state to send volunteers for the Civil War. Meanwhile, attention at home concerned the Sioux Uprising, one of the bloodiest Indian wars in the country's history. The Sioux Indians who had not been driven from the state were confined to small reservations. The federal government had forced the sale of some of these lands, reversing earlier treaty agreements. Driven further by crop failures and starvation, the Sioux attacked isolated farmsteads. In only a few weeks more than 500 civilians, soldiers, and Indians were killed.

Sioux Uprising

The most rapid period of settlement in Minnesota was during the 1880s, when homesteaders rushed into western and southwestern Minnesota. In the same period, lumber-

ing was at its peak and flour milling, using power provided by the St. Anthony falls, was becoming important. Both Minneapolis, as the lumber, milling, and retail centre, and its neighbouring city of St. Paul, as the transportation, wholesaling, finance, and government centre, tripled in population during the 1880s. The rivalry between the two cities became particularly intense after the census of 1880, when Minneapolis surpassed St. Paul in population.

Commercial iron ore production began in Minnesota in 1884 at Soudan, on the Vermillion Range. After the huge iron reserves of the Mesabi Range were discovered at Mountain Iron in 1890, large-scale production began, and the population along the Mesabi Range and in the Lake Superior port cities of Duluth and Superior grew rapidly during the next two decades. Most of the valuable pine, balsam, and spruce in central and northeastern Minnesota had been cut before 1900, after which time the lumbering industry declined rapidly. Wood products remained important in northern and northeastern Minnesota.

Tapping the ore beds

Adaptation and change. Since permanent settlement took hold in central North America, Minnesota has evolved from a frontier outpost to an integral part of the national and global economy. With its traditions of political activism, it has sought to influence those conditions it could and to adapt creatively to those it could not.

Mechanization of the resource-based economy has meant that fewer people could produce more. As a result, rural populations have declined since about 1920, and people increasingly have sought employment opportunities in the urban centres, particularly the Twin Cities area. The automobile, as successor to the railroad, has strongly influenced this pattern of development because people can now travel great distances with ease.

The attitude toward the environment has shifted from one of exploitation to more skillful management of the forests, water, soil, and wildlife. Remaining pristine wilderness areas, such as the Boundary Waters Canoe Area in northeastern Minnesota, are now guarded by many with a loving passion. The large areas of the state given over to parks, forests, and wildlife refuges attest to the high priority given to environmental management in Minnesota.

During the 20th century the development of Minnesota's economic focus from the regional and national level to that of the world has made its citizenry more conscious of the global community. Increasing exposure to global markets has made the Minnesota economy more vulnerable to fluctuations in international economic conditions and has required new forms of adaptation. Even the sometimes bitter parochial rivalry in the past between St. Paul and Minneapolis has mellowed with the acknowledgment of their common interests and competitors, and, as the core cities have become vastly outpopulated by their suburbs, a growing sense of metropolitan and state identity has developed. (N.C.G.)

Missouri

Near the centre of the coterminous United States, Missouri is the meeting place of the timberlands of the East and the prairies of the West, of the cotton fields of the South and the cornfields of the North. It has represented the political and social sentiments of a border state since its admission as the 24th member of the Union on Aug. 10, 1821. The question of its admission as a slave or free state produced in Congress the Missouri Compromise (1820), which regulated the spread of slavery in the western territories.

Role in U.S. history

Missouri was the westernmost state of the nation until the admission of Texas in 1845, and for decades it served as the eastern terminus of the Santa Fe and Oregon trails. For the West, St. Louis, Missouri's largest city, long was the closest contact with the more settled society and the culture of the East, and for the East the state had a reputation as the chief gateway to the West.

The Missouri River cuts across the state from Kansas City in the west, through Jefferson City (the state capital) in the centre, to just above St. Louis in the east, where the river joins the Mississippi. (Missouri was the name of a group of Indians who lived in the area; the French gave the name to the river, and it was later transferred to the state.) With the exception of Tennessee, Missouri has more neighbouring states than any other U.S. state. To the north lies Iowa; across the Mississippi River to the east, Illinois, Kentucky, and Tennessee; to the south, Arkansas; and to the west, Oklahoma, Kansas, and Nebraska. The area of Missouri is 69,697 square miles (180,516 square kilometres). Slightly more than half of Missouri's population lives in the two major cities, St. Louis and Kansas City, and their surrounding counties.

Missouri ranks high in the degree of urbanization and the amount of industrial and economic activity, though it maintains a vigorous and diversified agriculture. The rugged Ozark Plateau is a scenic beauty, and many lively folk traditions persist among its communities. Missouri retains numerous conservative characteristics of the rural life that predominated prior to the 1930s. Its nickname, the Show-Me State, suggests a tradition of skepticism regarding change. In Missouri the Democratic Party does not necessarily represent more liberal political philosophies than does the Republican, and the latter has made inroads into the traditional Democratic orientation. Continuing low tax bases prevent the elaboration of social services, a problem felt most acutely in the two major cities, which have had an increasing loss of wealth to the suburbs, coincident with greatly expanding needs of the cities.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* North of the Missouri River, in the glaciated section, Missouri's landscape is characterized by gently rolling hills; open, fertile plains; and well-watered prairie country. South of the Missouri, except in the extreme southeastern corner of the state and along the western boundary, the land is rough and hilly, with some deep, narrow valleys and clear, swift streams. It is a region abounding with caves and extraordinarily large natural springs. Much of the land is 1,000 to 1,400 feet (300 to 425 metres) above sea level. Near the western border, however, the elevation rarely exceeds 700 to 800 feet, and in southeastern Missouri, a part of the alluvial plain of the Mississippi, it is less than 500 feet. The St. Francois Mountains in the eastern Ozarks exhibit igneous granite and rhyolite outcroppings, while the rest of the state is underlain by sedimentary rocks, mainly limestones, dolomites, sandstone, and shale. Missouri is tectonically stable except for the southeastern portion, where small earth tremors occur. The possibility of another devastating earthquake of a magnitude comparable to those centred at New Madrid in 1811–12 cannot be discounted.

Drainage and soils. Drainage and soil conditions permit farming in all of Missouri's counties, although the Ozark Plateau only supports livestock farming because of the region's thin soil. Northern Missouri is generally well drained, much of it covered by rich glacial and loessial soils. The alluvial soils in the bottom lands along the many rivers and streams, which some soil experts believe to be the most extensive in the nation, also add to the farming potentiality. Except for the rivers that flow generally southeasterly into the Mississippi, many through Arkansas, the Missouri drains most of the state.

Ozark Plateau

Missouri's many lakes have been created by damming rivers and streams. The Lake of the Ozarks is a large artificial lake, with an area of approximately 60,000 acres (24,300 hectares) and a shoreline of 1,375 miles (2,213 kilometres). Most have been built primarily to furnish hydroelectric power and to prevent flooding, but they provide Missouri with excellent recreational resources. There also are many natural springs, some of them among the largest in the world.

Climate. Missouri is susceptible to the influences of cold Canadian air; of warm, moist air from the Gulf of Mexico; and of drier air from the southwest. Although winds are variable throughout the year, summer winds generally blow from the south and southwest and winter winds from the north and northwest. Rainfall, usually sufficient for crops, varies from around 34 inches (864 millimetres) in the north and northwest to 50 or more inches in the extreme southeast. About one-third of it falls from April to June. Heavy snows are unusual, most snow occurring between December and February. Missouri lies

in "Tornado Alley," the zone of maximum occurrence, and has an average of 27 tornadoes annually. Maximum January temperatures range from about 36° F (2° C) in the north and northwest to 48° F (9° C) in the southeast. The extreme northwest usually has cooler summers than the southeast, but summer temperatures well above 100° F (38° C) may occur in any part of the state.

Plant and animal life. Originally, about two-thirds of Missouri was forested, and the remainder was covered with prairie grasses. The river bluffs and valleys of the Ozark Plateau have a wide variety of unusual plants, including fameflower, royal catchfly, Trelease's larkspur, coneflower (*Rudbeckia maxima*), gayfeather, and fringed poppy mallow. Trelease's larkspur and coneflower are found in the wild in Missouri. Elk, deer, bison, and bears once were plentiful, as were such smaller animals as beavers, otters, and mink. After settlement and the development of agriculture, most of the larger animals, with the exception of deer, disappeared, and animals with valuable fur were trapped until near extinction.

Settlement patterns. Missouri's regions reflect the ethnic, religious, and political persuasions of the residents. The "Bootheel" in the extreme southeast was settled by planters from the South and was appended to Missouri at the time of statehood through the great influence of one planter; it is the centre of Missouri's cotton culture, which has declined due to cotton disease and is being replaced by the cultivation of soybeans, wheat, grain sorghum, and rice. The Ozark Mountains area, whose rugged terrain is unsuited to extensive agriculture, has been among the poorer regions of Missouri, but it constitutes one of the great tourist attractions of the state. "Little Dixie" is a block of counties that lies generally north of the Missouri River and extends westward along its banks to the middle of the state. It was initially settled by persons sympathetic to the South. Some of the finest examples of antebellum residences are found there. South of Little Dixie, on the bluffs and uplands south of the Missouri River and west of St. Louis, is a concentration of German settlements, known traditionally as the "Missouri Rhineland." In the western part of the state, north and east of the Missouri River, is historic "Mormon Country." There, followers of Joseph Smith settled around 1831, first at Independence and subsequently in other counties, until they were driven out by hostile neighbours. In the centre of the state, around Boonville, Franklin, and Columbia, is the "Boone's Lick Country," where the frontiersman Daniel Boone and his sons moved from Kentucky to make salt.

The "Bootheel," the Ozarks, and other regions

Human settlement has altered Missouri's landscape significantly. Only one-third of the state remains forested, mostly on the hills and slopes of the Ozarks. Nearly all of the prairie land has been brought under cultivation. The damming of streams has produced numerous lakes and reservoirs, and in the southeast, drainage systems have converted former swamps into one of the state's richest agricultural regions. Agriculture in the state traditionally has been characterized by family-owned farms, but, as elsewhere in the nation, the number of farms has been decreasing, while acreage and productivity per farm have increased. Urbanization also has reduced the amount of agricultural land. Although urban settlements are scattered throughout the state, Kansas City and St. Louis are Missouri's important centres of commerce and manufacturing and the nuclei of large metropolitan areas that extend into Kansas on the west and Illinois on the east.

The people. After the early French settlers, immigration came largely from states to the east and northeast, as well as from the South, with the implantation of its type of economy and society in the Bootheel and in Little Dixie. The first immigrants from abroad—particularly Germans, Irish, and English—came in great numbers after 1820. By 1860 large groups of Germans had settled in Missouri, mainly in St. Louis and just to the west, while many Irish had settled in the city. Between 1860 and 1890 the immigration from Ohio, Illinois, and Indiana exceeded that from the South, while an increasing number of immigrants from Germany arrived, settling mostly in urban centres. Subsequently, St. Louis and Kansas City attracted sizable communities of Italians and Greeks as well as Poles and Jews. By World War II more than 20 different European ethnic groups had settled in rural Missouri.

Patterns of immigration

While Missouri's population has grown modestly, people continue to leave the state, a pattern more characteristic of heavily rural or economically less developed states. The northward migration of many blacks from the rural South has altered the racial composition of the population. From 1940 to 1960 the white population increased by 11 percent, whereas the nonwhite population increased 62 percent. By the late 20th century blacks made up more than 10 percent of the state's total population, largely concentrated in St. Louis and Kansas City.

It is not unusual that a crossroad state should exhibit great religious diversity. The Roman Catholic church, which was dominant until the Louisiana Purchase, remains powerful, particularly in the St. Louis and Kansas City areas. After 1803 the chief Protestant denominations as well as many smaller sects were established. Baptists and Methodists predominate, and various Pentecostal groups are well represented throughout the state. Jewish communities have flourishing congregations in the larger cities.

The economy. Although agriculture has remained important as an income-producing activity, services, manufacturing, and wholesale and retail trade have forged ahead since World War II. Missouri has become the commercial and industrial leader among all its adjacent states except Illinois and Tennessee. In some types of manufacture, particularly in the production of aerospace and transportation equipment—including automobile assembly—Missouri ranks among the leading states in the nation. Recreation and tourism have surpassed agriculture in economic importance. Much of this growth has been in the larger towns and cities and near the large Ozark lakes.

The state Department of Economic Development includes several divisions and commissions that have had a significant influence upon the state's economic development. Local chambers of commerce and private financial groups also have stimulated economic growth.

Workers in Missouri generally have enjoyed the benefits of an expanding economy, but income per capita is below the national average. Unions have had great influence in increasing the salary levels of teachers, clerical workers, and those in trades and crafts. In spite of favourable comparisons with other states in income and revenue, Missouri ranks near the bottom in terms of state tax revenue per capita. As a result, it has suffered from a lack of sufficient revenue to meet the needs and services that modern governmental agencies are expected to provide.

Resources. The state's variety of resources includes lead and iron ore, barite, limestone, timber, and hydroelectric power. Mineral-rich Missouri leads the nation in lead production, and deposits of lead and zinc, as well as iron ores, continue to be discovered. Iron production, however, has ceased, and lead production has decreased because of declining demand. About two-thirds of the forest stand lies in the Ozarks, and since the 1950s industry has made increasing use of these resources. The acreage of many marginal farms in the Ozarks is being given over to recreation, one benefit of which is their return to forest.

Minerals

Agriculture and industry. A wide range of crops are grown, including soybeans, the state's most valuable crop, wheat, corn, cotton, rice, and tobacco. More than half of the state's total farm income derives from the sale of animals and animal products, mainly cattle, hogs, and dairy products. In parts of northern Missouri the combined number of cattle and hogs amounts to more than nine times the number of human inhabitants.

Manufacturing is led by the production of aerospace and transportation equipment, followed by the processing of food and the production of chemicals.

Finance and trade. Missouri ranks high among the states in commercial and savings and loan bank assets. Federal reserve banks are located in St. Louis and in Kansas City, and the regional offices of the Internal Revenue Service in Kansas City serve much of the Midwest. Kansas City and St. Louis have always been important trading and commercial centres for large regions reaching into neighbouring states, and they rank among the foremost grain and cattle markets of the nation. Farm-related

products and automotive sales are the leading sources of revenue for wholesale and retail trade, while business demands predominate the services sector.

Transportation. The major flows of traffic within the state are from the east to west along the Missouri valley and southward along the Mississippi. Missouri is served by several interstate highways. Its railroads are linked with most of the nation's major trunk lines, and St. Louis, Kansas City, and Jefferson City are served by Amtrak passenger service. Since 1910 the gradual abandonment of competing parallel lines and short lines built by mining and lumbering companies has led to a considerable reduction in Missouri's railroad mileage. The Mississippi and Missouri rivers, providing more than 1,000 miles of navigable waterways within the state, connect waterborne traffic with New Orleans. St. Louis and Kansas City are regional air hubs.

River transportation

Administration and social conditions. *Government.* Missouri is governed under its fourth constitution, ratified in 1945, but the basic structure of government has remained constant since the first constitution of 1820. Governors are elected for four-year terms and may succeed themselves. They have the power of "item veto," by which they may strike individual provisions from any appropriation bill, except those for public school support or payments on the public debt. Legislative power is vested in the General Assembly, composed of the Senate and the House of Representatives. The House has 163 members, elected for two-year terms, and the Senate has 34 members elected for four-year terms. Each senator represents equivalent population units, whereas each county has at least one representative, regardless of its population. During the early 1960s voters in 42 rural counties had nearly 10 times as many representatives as citizens living in St. Louis, St. Louis county, and Kansas City and surrounding Jackson county. Redistricting was completed in 1971 to establish equality of representation.

The judicial system is similar to that of most states, with a Supreme Court as the highest tribunal. Below it are the Court of Appeals and 44 circuit courts. An unusual feature of Missouri's judicial system is a method of selecting judges by merit, known as the Missouri Plan and adopted by several other states. Under the plan the governor fills a vacancy in the court by appointing one of a three-member panel selected by a nonpartisan judicial commission. The appointment must be confirmed in a separate nonpartisan ballot in the first general election after the judge has been in office 12 months. The plan applies only for the Supreme Court, the Court of Appeals, and circuit courts in metropolitan St. Louis and Kansas City. In counties outside the two large metropolitan areas, circuit judges and associate circuit judges are elected by voters in partisan elections. Political partisanship is still a consideration in the governor's selection of appointees and in the selection of commission members.

Missouri Plan

The city and the county are the most important units of local government. The state has 114 counties plus the city of St. Louis, which is independent of surrounding St. Louis county. Counties are administered by a county commission consisting of three elected commissioners. Counties with a population of more than 85,000 are permitted to adopt their own charters. Missouri was the first state in the nation to permit cities to adopt their own governing charters, and there are now more than 20 cities across the state with home-rule charters. Most Missouri cities have the mayor–council form of government.

Missouri voters tend to favour Democratic candidates, but they have elected Republican governors and returned Republican majorities to the General Assembly on a number of occasions. Both parties contain liberal and conservative factions. The Democratic Party is somewhat stronger in the two large metropolitan centres, while the Republican Party is strongest in southwestern Missouri and in the rural northern counties.

Education. Since 1945 many small school districts have consolidated into fewer large ones, and school enrollments and revenues have declined significantly since the early 1970s. Missouri has lagged behind other states in support of public education, while at the same time efforts to desegregate schools in St. Louis and Kansas City have been an additional financial burden.

Higher education has expanded both in the public and private institutions. The University of Missouri has campuses in Kansas City, St. Louis, and Rolla, in addition to the main and oldest campus in Columbia. There are also several state colleges plus Lincoln University, founded originally for blacks but now enrolling whites. Among the private institutions are Washington University and Jesuit-run St. Louis University, both in St. Louis. Financial support of higher education consistently has been inadequate.

Colleges and universities

Health and welfare. The state Department of Social Services and the Department of Mental Health, both established in 1974, provide services for the ill and the indigent. In cooperation with federal agencies, St. Louis and Kansas City have undertaken massive urban-renewal programs to help relieve inner-city problems.

Though costs of living have risen, especially in the larger population centres, income per capita also has risen. Pockets of poverty exist in depressed rural areas and in city slums, but Missouri generally has not had the severe poverty of states with more heavy industrialization or a greater amount of subsistence farming. The disparities between rich and poor are greatest in and around St. Louis and Kansas City. Because the metropolitan areas of St. Louis and Kansas City cut across state lines, their problems of metropolitan government and management are compounded.

Cultural life. Diversity characterizes Missouri's cultural milieu, from the centres of fine art, music, and theatre along the St. Louis–Kansas City axis to the folk culture and native crafts of the Ozarks.

The state has furthered cultural opportunities through the Missouri State Library, established in 1946, and the Missouri State Council on the Arts, created in 1965. The state library has been responsible for the rapid growth of county and regional libraries. The larger cities have their own library systems. The council on the arts has stimulated communities to expand their cultural resources.

From the state's beginnings, the arts have flourished in Missouri. In painting, George Caleb Bingham and Thomas Hart Benton have been preeminent. If expatriate poet T.S. Eliot, a St. Louis native, is disqualified, Mark Twain remains Missouri's most distinguished literary figure, world-renowned for his immortalization of mid-19th-century life in Hannibal and along the Mississippi. There are schools of art and design in St. Louis and Kansas City, and music flourishes in both cities. St. Louis' Gateway Arch, designed by Eero Saarinen, is a spectacular example of the diverse architectural styles in evidence throughout the state.

The Ozarks abound in folk traditions, tales, and ballads. This region was settled primarily by pioneers from the southern Appalachians, who brought with them traditional songs and ballads, some of which were brought over from England and Scotland in the 17th century. Native crafts once practiced out of necessity by the pioneers have begun to flourish again in response to the interest of tourists and also because of recognition of the intrinsic merit of the objects made. Quilting, woodworking, basketmaking, and pottery are among some of the most important crafts, and their development is encouraged through the Missouri Federation of Arts and Crafts.

Folk traditions and crafts

Besides its universities and colleges, Missouri has outstanding cultural institutions, among them the St. Louis Symphony Orchestra, the second oldest U.S. civic orchestra and one of the major musical ensembles in the nation. In Kansas City the Nelson-Atkins Museum of Art owns one of the finest collections of Asian art in the Western Hemisphere, and the Linda Hall Library has an outstanding scientific collection. Independence is the home of the Harry S. Truman Library and Museum. Fulton, where the British leader Winston Churchill made his famous "Iron Curtain" speech, has a collection of Churchilliana in the Church of St. Mary the Virgin, Aldermanbury, whose stones were reassembled there after its destruction in the World War II bombings of London.

High school, college, and professional sports are popular. Kansas City is the home of the professional baseball Royals and the football Chiefs. St. Louis is the home of

the baseball Cardinals. Increased leisure time and mobility have stimulated an enormous interest in recreation, and Missouri has developed a superb system of state parks and historic shrines and memorials that are attractive to residents and visitors alike. Numerous man-made lakes afford fishing and waterskiing, while the clear, cool rivers of the Ozark Plateau offer pleasure to canoeists, fishers, and campers. The Mark Twain National Forest provides ideal habitat for game animals and songbirds. The Current and Jacks Fork rivers are protected and managed as wild and scenic waterways by the Ozark National Scenic Riverways.

The Missouri Press Association, established in 1867, has had an important effect upon the development of the press in the United States. It was responsible for establishing the world's first school of journalism at the University of Missouri and for founding the State Historical Society of Missouri, with the largest membership in the nation. There are numerous local newspapers and journals; newspapers of national distinction include the *St. Louis Post-Dispatch,* made famous by Joseph Pulitzer, and the *Kansas City Star* and *Kansas City Times.*

HISTORY

Before the coming of European explorers the land that was to become Missouri was the home of a diverse group of Indian tribes whose mounds and other remains dot the state. One of the tribes was called the Missouri.

French settlements

Exploration and settlement. The recorded history of the region dates from the settlement of some French lead miners and hunters at Sainte Genevieve, on the western bank of the Mississippi, in about 1735. At some distance from its original site, Sainte Genevieve remains the oldest continuously inhabited white settlement in Missouri. Some 30 years later, Pierre Laclede, a French fur trader from New Orleans, founded St. Louis. At the time of the Louisiana Purchase in 1803, most of the 10,000 residents of the region were French settlers from the Illinois country, but some Americans had come from Kentucky and Tennessee, which, with Virginia, were the major immediate sources of settlers in following decades.

Statehood, controversy, and war. The "pull of the West" solidified Missouri's position as a land of passage after it achieved statehood as a slave state in 1821 under the Missouri Compromise of 1820. Migrants bound for Texas outfitted in Missouri, and later thousands of people heading west poured through Saint Joseph, Independence, Westport Landing, and the City of Kansas (Kansas City).

Although slavery had become well established in the Missouri Territory, both on the plantations of transplanted Southerners and in French lead-mining ventures, the Abolition movement drew increasing support, particularly after statehood was granted. In challenging the traditional Southern institution, new arrivals from the North and from Europe challenged also the principle of states' rights. The emotional nature of the controversy gave rise in some instances to mob violence against abolitionists, including the Mormons who gathered in Missouri in the 1830s. Laws were enacted to prohibit teaching any black to read or write and to prevent any free black from entering the state. The case of the Missouri slave Dred Scott, who sued for his freedom on the grounds that his master had for a time moved him into the free state of Illinois and the free territory of Wisconsin, resulted in a Supreme Court decision (1857) that made slavery legal in all the territories.

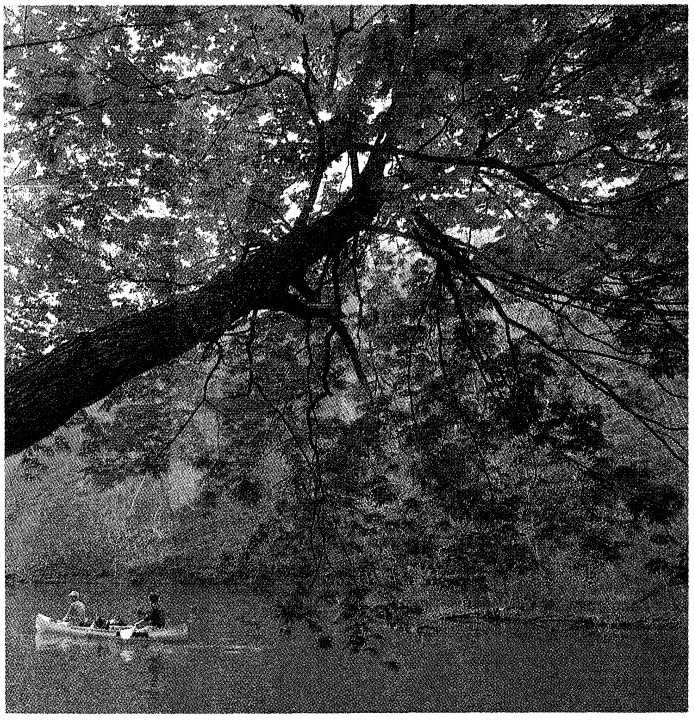
The 1850s were years of increasing dissension, worsened by the Kansas–Nebraska Act of 1854 that set slave- and free-state advocates at one another's throats for control of those adjoining territories. Missouri already was moving toward a free-state economy, however, and the state stayed within the Union during the Civil War. Missourians fought on both sides, but those in the Union army outnumbered those in the Confederate army by nearly 4 to 1. Conflict occurred in the state, much of it guerrilla warfare along the Kansas border. After the war, Confederate sympathizers were dealt with harshly; overall, however, Reconstruction was not as severe as in the Deep South.

Role in the Civil War

The 20th century. The continued growth of Missouri in the late 19th and early 20th centuries was celebrated in the famous St. Louis Exposition in 1904. The state remained heavily rural and agricultural, however, until the Great Depression of the 1930s and World War II brought about vast movements of people into the cities.

Three important developments have shaped the economy of Missouri since World War II: the shift from agriculture, mining, and lumbering to manufacturing, particularly of durable goods, and services; large investments in public and social services, highways, and rural electrification; and population growth, particularly near the large reservoirs and in the peripheries of St. Louis, Kansas City, and Springfield. Rural areas of Missouri have attracted many new plants that employ a small number of workers, but two-thirds of the manufacturing employment remains concentrated in the St. Louis and Kansas City metropolitan areas. Small towns have changed rapidly, their relative success depending to a large degree on geography and transportation. Some towns near large cities have grown as they have been brought into the commuting zone by improved highways, but most villages have suffered economic stagnation as the rural population has declined and the remaining residents shifted their commercial support to the larger towns. (E.J.W./M.D.R.)

Matt Bradley



Canoeing on the Eleven Point Scenic River, in Mark Twain National Forest, on the Ozark Plateau in southern Missouri.

Nebraska

One of the west central states of the United States, Nebraska during the first 60 years of the 19th century was primarily a water and land route to the rich trapping country to the north and west and the settlement and mining frontiers of the mountain and Pacific regions. With the development of railroads after the Civil War and the consequent immigrations, however, the excellent soils of Nebraska were plowed, and its grasslands gave rise to a range cattle industry. As a result, the state has been, almost since its admission to the Union on March 1, 1867, as the 37th state, a major producer of food commodities.

Rivers were important to Nebraska's geography and settlement. The Missouri River, a major highway to the trans-Mississippi West in the early 19th century, forms the eastern boundary with Iowa and Missouri and about a fourth of the northern boundary with South Dakota. The name Nebraska is derived from an Indian word meaning "flat water," a reference to the Platte River, which served as a magnet for urban clusters across the state. The river is formed by the confluence of the North and

South Platte rivers, both of which rise in Colorado on the southwest, although the North Platte swings northward through Wyoming, on the west, before entering Nebraska. The southern boundary with Kansas was established when the two territories were created by the Kansas–Nebraska Act in 1854, legislation that heightened the sectional hostilities that exploded into the Civil War. A majority of Nebraskans today live close to the Missouri and Platte, leaving much of the state's 77,355 square miles (200,350 square kilometres) lightly populated. Lincoln, in the southeastern part of the state, is the capital.

Creation of Nebraska territory

Agriculture is basic to Nebraska's economy, but only one-tenth of its labour force is employed directly in farming or ranching. Economic conditions have had a direct bearing on the state's political life, including a brief period of protest through the agrarian-oriented People's (Populist) Party during the 1890s. Although Nebraska traditionally has been a Republican stronghold, the Democrats also have been an important political force.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Nebraska comprises parts of two physiographic regions, the till plains of the Central Lowland (in the eastern third) and the Great Plains.

The elevation rises from a minimum of 840 feet (256 metres) above sea level in the southeast to a maximum of 5,426 feet (1,654 metres) near the Colorado–Wyoming boundaries. Much of the land is gently rolling prairie, although the river valleys, much of south central Nebraska, and a large portion of the panhandle district are flatlands. The Sand Hills country of north central and northwestern Nebraska is a vast, treeless, grass-covered region that comprises almost one-fourth of the area of the state.

Drainage. Nebraska lies within the Missouri River drainage system; the Platte, the major Nebraska tributary, joins the Missouri south of Omaha. Although shallow and unnavigable, the Platte is vital to the state's irrigation. The Elkhorn River enters the Platte west of Omaha, and the Loup River, formed by three tributaries flowing out of the Sand Hills, also discharges into the Platte. The Republican and Big Blue rivers flow through southern Nebraska, emptying into the Missouri in Kansas via the Kansas (Kaw) River. The Niobrara, a swift-moving stream that rises in the high country just west of the Wyoming border, flows across extreme northern Nebraska. The state also has a vast supply of groundwater that has made possible the extensive development of well irrigation.

Soils. All of Nebraska's soils are excellent for agriculture. The prairie soils of the southeast and the humous soils of central and northeastern Nebraska are highly suitable for general farming. South of the Platte and west of the prairie soil area, the soil is best suited to small-grain production. Winter wheat adapts to the soil and marginal precipitation of western Nebraska. The wind-deposited soil of the Sand Hills, because of limited precipitation and the danger of erosion, is suited solely to cattle grazing. With many small lakes and luxuriant grasses, the Sand Hills area is a superb rangeland. The alluvial soils of the Missouri and Platte river valleys and the valleys of smaller streams are outstanding for raising corn (maize) and other crops.

Soil types

Climate. Hot winds from the southwest often push summer temperatures in Nebraska above 90° F (32° C) and sometimes above 100° F (38° C). Average July temperatures range from 73° F (23° C) in the panhandle to 78° F (26° C) in the southeast. In the winter northwestern winds often bring in Arctic air masses from Canada, and temperatures commonly fall well below 0° F (−18° C). Low-pressure systems, moving out of the southwestern states, sometimes bring great blizzards to the state and pose a danger to travelers and stock raisers. Average January temperatures vary from 24° F (−4° C) in the panhandle to about 20° F (−7° C) in the northeast. The average growing season is 168 days in the southeast and 133 days in the panhandle. The average annual precipitation varies from 33 inches (840 millimetres) in the southeast to less than 16 inches in the extreme west. Since a minimum of 20 inches is usually considered necessary for normal crop production, approximately one-half of Nebraska may be considered semiarid. Irrigation is used extensively in eastern and central Nebraska and is essential to certain types of agriculture in the western part of the state.

Plant and animal life. A wide variety of grasses originally covered Nebraska's prairies, and the slopes of the river valleys were well covered with deciduous trees. Cottonwood, elm, and some oak and walnut are found along the bluffs of eastern Nebraska, while conifers grow in the Wild Cat and Pine Ridge highlands and the Niobrara valley. The Nebraska National Forest in west central Nebraska resulted from a human effort to bring trees to the barren plains. Until their near extermination at the hands of whites, bison roamed widely over the Nebraska plains. Some of these animals remain in their natural habitat on the Fort Niobrara National Wildlife Refuge, near Valentine. Antelope and deer are also native to the state, as are prairie dogs, coyotes, jackrabbits, skunks, and squirrels. Migratory birds and pheasants are common.

Settlement patterns. According to federal law, the land was surveyed into townships, six miles (10 kilometres) square and containing 36 sections; the section, comprising 640 acres (259 hectares), was the basic unit of land. This gridlike survey system remains a basic feature of Nebraska's landscape. Most of the towns and villages were located close to rivers, streams, and timber. A number of them developed as railroad terminals, but changing patterns of transportation brought about growth for some communities and stagnation or oblivion for many others.

Section land unit

Economics, geography, and politics created certain sectional distinctions within Nebraska. The placement of the capital of Nebraska Territory at Omaha so enraged the people south of the Platte that they sought annexation to Kansas, and in 1867 the state capital was moved to Lincoln, south of the Platte. Both Lincoln and Omaha emerged as major regional hubs, but, because these cities are located at the eastern end of the state, western Nebraskans often regard Denver, Colo., as their major service centre.

The people. In addition to the Americans who came to Nebraska, large numbers of European immigrants settled in the state during the late 19th century. From 1866 to 1877 the Nebraska State Board of Immigration employed an agent in Europe to recruit settlers. The Burlington and Union Pacific railroads made much greater efforts than did the state government in this direction. The largest immigrant group was the Germans, who in 1890 numbered 72,000; immigrants from the Scandinavian countries (particularly Sweden), Bohemia, and the British Isles also made important contributions to the settlement of Nebraska.

As a result of the European influx, Nebraska was not molded in the traditional Anglo-American form that had characterized much of the nation's earlier development. Large numbers of Roman Catholics from Bohemia, Germany, and Ireland and Lutherans from Germany and Scandinavia gave diversity to the religious and secular life of Nebraska. Although the linguistic identity of the non-English-speaking groups faded away, other aspects of their diverse cultural heritage have survived.

With the opening of the territory to settlement in 1854, the federal government created a reservation for the Omaha Indians in northeastern Nebraska, part of which subsequently was made into a reservation for the Winnebago people, recently displaced from Wisconsin. Some of Minnesota's Santee Sioux also went to a reservation in northeastern Nebraska. (The Omaha-Winnebago and the Santee Sioux reservations still exist.) In the 1870s the Oto-Missouri, Pawnee, and Ponca peoples, after living on reservations in Nebraska, were removed to Indian Territory (now Oklahoma). By 1878 the Dakota (Sioux) had given up their agencies in northwestern Nebraska and were located on reservations just over the border in Dakota Territory (now South Dakota). In the early 20th century the implementation of federal legislation permitting reservation Indians to hold land titles, rather than fostering the benefits of private ownership by native peoples, often resulted in quick and unseemly transfers of real estate to land-hungry whites. About one-third of Nebraska's native people now live in Omaha and Lincoln. Other Indians are concentrated on the reservations and in areas of the Nebraska panhandle.

The black community

Blacks came to Nebraska early in the history of the state. Most settled in Omaha, which by 1900 had a black population of more than 3,400, a figure that by the late 20th century had increased more than 10-fold. Blacks were concentrated north of downtown Omaha in an area that increasingly became characterized by the social and economic problems common to the ghettos of other large cities. This core of the black community has declined markedly in population, however, as many blacks have moved to adjacent neighbourhoods. Racial disturbances in Omaha in the 1960s emphasized the need for improved economic opportunities and better police–community relations. Omaha's black community has long had representation in the legislature.

The most striking trend in Nebraska's demography has been the steady decline of the population of the rural areas and the marked growth of the cities and their suburbs. Urban growth has been stimulated by the mechanization of agriculture, which brought about the working of more land by fewer persons, decreases in the number of farms, and increases in average farm sizes. Similarly, most small towns, reliant upon the local farm trade, have continued to lose population, a condition undoubtedly hastened by a modern highway system that has enlarged the trade areas of the cities. Rural schools, hospitals, and other institutions have been forced to adjust accordingly. Although employment opportunities have been diminishing in rural Nebraska, there has been an increase in the number of work opportunities in manufacturing, notably in the Platte valley with its excellent highway transportation, as well as in the state's major cities.

The economy. Nebraska's economic development is heavily dependent upon private investment from outside its borders. The state Department of Economic Development was established in 1967 to bring new industry to Nebraska. In addition, a law passed in 1987 provides tax incentives for the development of business and industry. Wholesale and retail trade and other services, manufacturing, transportation, and agriculture are the major sources of income. The state generally has been conservative in labour matters and ranks low nationally in the percentage of unionized nonagricultural workers. Nebraska has a right-to-work law that forbids compulsory union membership.

Farm production

Agriculture. Nebraska remains a leading agricultural producer. The principal crops are corn, sorghum, soybeans, hay, and wheat. Nebraska ranks high among the states in the production of corn for grain and in wheat, sorghum, dry edible beans, and sugar beets, as well as in the number of cattle, hogs, and pigs. Potatoes are also a significant crop.

Industry. Food processing is the most important industrial activity of the state in terms of value added by manufacture. Other leading industrial activities include the manufacture of machinery and chemicals and allied products, printing and publishing, and the production of primary and fabricated metals and electrical, electronic, and transportation equipment.

Crude petroleum accounts for more than half of the value of the state's mineral extraction. Nebraska also produces some natural gas, as well as significant amounts of cement, lime, sand, gravel, crushed stone, and clay. Additional quantities of natural gas, however, are imported to serve the commercial, industrial, and residential needs of the state. All electrical utilities are publicly owned, and consumer rates are among the lowest in the nation.

Nebraska, and Omaha in particular, is known as a major centre of the American insurance industry. Tourism is essential to the livelihood of the state and ranks third behind agriculture and manufacturing in economic significance.

Transportation. Nebraska is located on some of the most important arteries linking the East and West. Within the state traffic in the east tends to flow toward Omaha, Lincoln, and Sioux City, Iowa, and toward the cities in the Platte valley. Much of western Iowa lies within the trading area of metropolitan Omaha.

Nebraska has a good network of modern highways. The most important route is Interstate Highway 80, which carries heavy traffic east–west across the state. Several railroads also operate in the state, and both Omaha and Lincoln are served by Amtrak. Omaha is an important port for commercial barge traffic on the Missouri. Air carriers serving Nebraska include both major national lines and those that provide "feeder" service to the smaller communities of the state.

Administration and social conditions. *Government.* Nebraska functions under a frequently amended constitution dating from 1875. Since 1937 it has had a one-house legislature whose members are elected without political-party affiliation—the only such legislative body in the nation. The 49 members of the legislature, or "Unicameral," are popularly elected for four-year terms following primary and runoff elections in their districts, which are equally proportioned by population. The Unicameral meets for sessions of 90 legislative days in odd-numbered years and 60 days in even-numbered years.

The unicameral, nonpartisan legislature

The nonpartisan feature of the legislature has many critics, who charge that the lack of political parties in the Unicameral results in a lack of leadership in that body. Indeed, nonpartisanship may have enhanced the importance of lobbyists in the legislative process.

The governor, the chief executive officer, is elected for four years on a partisan ballot and is limited to two consecutive terms. The governor is responsible for the operation of some 20 administrative departments and is an ex officio member of various boards and commissions. The governor must present a detailed budget to the state legislature, which needs an affirmative three-fifths vote to appropriate more funds than recommended by the governor or to override a gubernatorial veto. Other elected state officers also run on partisan ballots.

Nebraska's court system, reorganized in 1972, comprises the Supreme Court, with seven justices, and 48 judges of district courts. In addition, there are conciliation courts, county courts, municipal courts in Omaha and Lincoln, and juvenile courts, as well as a Workers' Compensation Court. Nebraska has adopted the merit system for selecting judges. Judicial nominating commissions, chosen by the governor and the Nebraska State Bar Association, compose lists of nominees to fill vacancies on the bench. The governor then appoints one of the nominees to fill a particular position. After three years judges run for retention on a nonpartisan ballot in a general election and must run in similar elections every six years thereafter.

County government is vested in boards of supervisors or commissioners of from three to seven members, who like other county officials are elected on partisan ballots. The city manager and mayor–council forms of government are used in Nebraska's cities, and governmental authority in villages is vested in elective boards of trustees.

Nebraska Territory was the creation of a Democratic administration in Washington, D.C., and Democrats dominated Nebraska politics until the Civil War. The 30 years after 1860 were marked by Republican preeminence in Nebraska, but the political ferment during and after the 1890s brought an end to one-party rule. Although a slight majority of Nebraska's voters are registered Republicans, Democrats often are elected to office.

Education. Since the 1960s, state aid for education to local governments has increased greatly, and the number of school districts has been cut drastically in order to make more efficient use of educational facilities and programs.

There are more than 30 institutions of higher education in Nebraska; about one-half are private schools, and the rest are state-operated four-year colleges and publicly supported technical community (junior) colleges. The University of Nebraska (established in 1869) is the largest educational institution in the state and is composed of three semiautonomous campuses—the original campus in Lincoln, a campus in Omaha, and the medical school, with facilities in Omaha and Lincoln. Graduate degree programs are offered at the University of Nebraska and Creighton University (Omaha), both of which offer programs in medicine, law, and dentistry, and at state colleges in Chadron, Kearney, Peru, and Wayne.

Higher education

Health and welfare. Nebraska's programs of public assistance include medical aid and financial assistance for dependent children, the aged, and the blind. Federal fund-

ing provides more than half of Nebraska's public-assistance expenditures. Although welfare funding has steadily increased, the average monthly public-assistance payments in Nebraska are below the national average.

The state maintains a system of mental hospitals and other specialized health, correctional, and care facilities. Omaha ranks as a medical centre of national significance.

Cultural life. In less than two generations Nebraska was converted from a wilderness inhabited by a small number of Indians to a settled commonwealth of more than 1,000,000 residents. This conquest was an important achievement of the 19th century, and it is natural that the cultural contributions of Nebraska, like those of other Western states, are centred on this frontier experience.

Such Nebraska authors as Willa Cather, Mari Sandoz, and Bess Streeter Aldrich were among those who wrote perceptively of life on the plains and won national audiences. The relationship of the pioneers to a capricious natural environment, the life-styles and interaction of settlers of diverse social and ethnic backgrounds, and the plight of the Indians were among the important themes of these writers. The poet John G. Neihardt wrote with feeling of Plains Indian life; he also re-created the adventures of the explorers of the 19th-century West. In the early 1970s his narrative *Black Elk Speaks* achieved national recognition some 40 years after it had originally been published.

The Nebraska State Historical Society, organized in 1878, continues to make important contributions to an understanding of life in Nebraska and the West. In 1960 the University of Nebraska Press launched the paperback Bison Series, reprints of early and modern works on the American West, including histories, collections of lore from Indian and white settlers, and other important documents, many of which had been out of print.

The Joslyn Art Museum in Omaha and the University of Nebraska's Sheldon Memorial Art Gallery in Lincoln contain the state's major collections in the visual arts. The performing arts have flourished in Nebraska, both in the development of local musical, theatre, and dance groups and through performances by touring artists of national stature.

Folk and historical festivals

Various folk observances, such as the Czech Festival at Wilber, are reminders of the diverse origins of the people of Nebraska. Ogallala, a roaring cow town during the 1870s and '80s, relives its colourful past with its Front Street festivities held each summer. Each October the Knights of Ak-Sar-Ben (Nebraska spelled in reverse), an Omaha civic organization founded in 1895, crown a king and a queen of Quivira. This event commemorates the search through the plains in 1541 of the Spanish explorer Francisco Vázquez de Coronado for the legendary Seven Golden Cities of Cibola and the Kingdom of Quivira. The University of Nebraska football team has attained national prominence, and few subjects hold the attention of Nebraskans as do the fortunes of the Cornhuskers. The benefits derived from the citizens' enthusiasm for football are important in the state's economy. Recreational areas include several state parks, the Nebraska National Forest, and the Oglala National Grassland.

HISTORY

Various prehistoric peoples inhabited Nebraska as early as 8000 BC. In the 19th century semisedentary Indian tribes, most notably the Ponca, Omaha, Oto, and Pawnee, lived in eastern and central Nebraska. The west was the domain of the nomadic Brulé and Oglala Sioux, but other tribes, such as the Arapaho, Comanche, and Cheyenne, also roved the area.

Exploration and settlement. Nebraska was on the periphery of the North American empires of France and Spain, but in 1763 Spain won title to the trans-Mississippi region. Spanish efforts to develop Indian trade in upper Missouri brought little success, and international politics led to the transfer of the region, including Nebraska, to France in 1800. Three years later the United States acquired this vast area as part of the Louisiana Purchase.

In 1804 the Lewis and Clark Expedition visited the Nebraska side of the Missouri River and conducted the first systematic exploration of the area. Shortly thereafter a vigorous fur trade developed along the Missouri, but Nebraska was primarily a highway to richer fur-trapping areas to the north and west. During the 1840s the Platte valley became another highway as thousands of settlers moved westward.

Much interest soon developed in Nebraska and in the Platte valley as a potential railroad route to the Pacific.

Roles of speculators and railroads

Frontier land speculators in western Missouri and Iowa anticipated great financial gains if the Nebraska country, part of the large Indian domain between the Missouri River and the Rocky Mountains, were opened for settlement. With the adoption of the Kansas–Nebraska Act in 1854, the federal government extended political organization to the trans-Missouri region. Originally the Nebraska Territory comprised 351,558 square miles, but by 1863 the organization of the Colorado, Dakota, and Idaho (including the states of Montana and Wyoming) territories had reduced Nebraska almost to its present dimensions.

Much of the economy of the early Nebraska settlements along the Missouri River was based on land speculation. Agriculture soon began to develop, however, and some river towns became important transfer points for freight and passengers going west. The completion of the Union Pacific Railroad in 1869 and the railroad construction that followed contributed to the development of the state.

Statehood. After Nebraska's admission to the Union in 1867, and despite an economic depression and a grasshopper plague, the population increased from about 120,000 to more than 1,000,000 by 1890. The Indian resistance on the frontier was broken during these years, and settlement extended westward into the panhandle of the state. During the 1880s Omaha became an important industrial and meat-packing centre, and Lincoln became prominent as the state capital and as the seat of the University of Nebraska. In the late 19th and early 20th centuries Nebraskans struggled over the issues of woman suffrage and prohibition of alcoholic beverages, revealing cultural differences between old-stock Americans and recent European immigrants.

Impact of economic cycles

In the 1890s Nebraska's farmers, afflicted with poor crop prices, high transportation costs, and economic depression, expressed their protest through the People's Party. Although the Populist movement was short-lived, it invigorated the political life of the state. Prosperity returned by 1900 and continued for two decades. Through the 1920s Nebraska's agriculture again was beset with mediocre marketing conditions, which were in part responsible for the failure of 40 percent of the state-chartered banks from 1921 through 1930. With the advent of the Great Depression of the 1930s the state's economy deteriorated further, necessitating massive federal assistance.

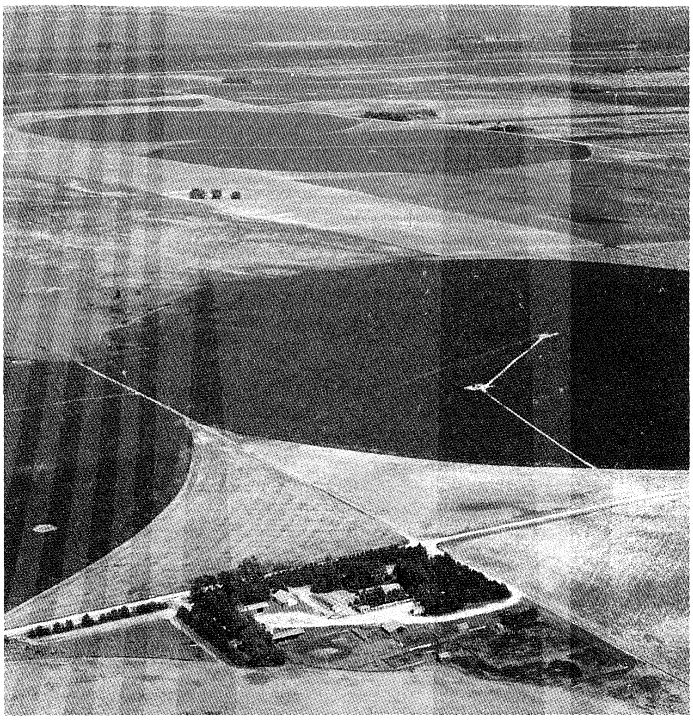
In 1933 the state legislature authorized the creation of public power and irrigation districts. Loans from the U.S. government enabled these districts to construct hydroelectric and irrigation projects in the Platte and Loup river valleys. A public agency later purchased the private electric power companies outside the Omaha area, and in 1946 the Omaha Public Power District acquired the local private power company. Nebraska thus became the first state with complete public ownership of electrical generating and distribution facilities, an ironic fact in view of its reputation for political conservatism.

World War II brought economic recovery and other changes. Fort Crook, south of Omaha, became the site of a huge aircraft plant. In 1948 this location, renamed Offutt Air Force Base, became the headquarters of the Strategic Air Command, which stimulated the growth of the greater Omaha area.

Centre-pivot irrigation

One of Nebraska's chief resources is a vast supply of groundwater. Tapping this resource for irrigation rose dramatically in the mid-1950s. The introduction of centre-pivot sprinkler devices in the 1970s constituted a fundamental change in the history of Nebraska agriculture because it made possible the cultivation of lands that previously could not be irrigated. The impact of centre-pivot irrigation is evident in the circular pattern now overlaid on much of the traditional "checkerboard" landscape.

The heavy utilization of groundwater plus the possibility that more water could be diverted from the upper reaches



Farm situated in the corners of the circles produced by centre-pivot irrigation in the cornfields southwest of North Platte, Neb.
Chester Peterson, Jr.

of the Platte system in Colorado and Wyoming have made Nebraskans more sensitive to the need to conserve their water resources. In addition, they have become more aware of the danger of groundwater contamination by chemical fertilizers and related products.

In the 1960s, changes in the meat-processing industry caused a sharp decline in Omaha's status as a meat-packing centre. However, economic diversification brought continuing prosperity to Omaha, which has remained Nebraska's principal industrial centre. Since the 1950s other Nebraska communities have risen in prominence in manufacturing activities. Lincoln, with its strong emphasis upon education and government, has also diversified economically and has become one of the most dynamic urban areas of the north central region. (H.A.D.)

North Dakota

Officially classed as one of the seven western north central states, North Dakota was admitted to the Union as the 39th state on Nov. 2, 1889. It is a land of generally clear skies, seemingly endless grain farms, and vast cattle ranches. The state is rural, agricultural, and sparsely populated. Its terrain rises through three regions from east to west, incorporating parts of the two major physiographic provinces that separate the Appalachian and the Rocky Mountain systems. The state's name derives from the Dakota division of the Sioux Indians who inhabited the plains before the arrival of Europeans.

North Dakota, bounded by Canada on the north, Minnesota on the east, South Dakota on the south, and Montana on the west, has an area of 70,702 square miles (183,119 square kilometres). The largest city is Fargo, and Bismarck is the centrally located capital.

Among the last regions of the American frontier to be settled, the area that became the state of North Dakota experienced comparatively little of the fighting, lawlessness, and gold-rush excitement that give other frontier areas a colourful and sometimes lurid history. Instead, the region developed first as the home of hunting and farming Indian peoples, later as a trading area for white fur traders and for steamboats working the upper Missouri River from St. Louis, and then as a rich farming land for white settlers. The cool, subhumid climate of its location made it ideal for spring wheat and for cattle ranching. The area subsequently developed a way of life dependent on outside centres of population, industry, and economic power. With adaptation to the environment, however, North Dakotans also developed constructive reactions to those conditions that underlie their state's dependency.

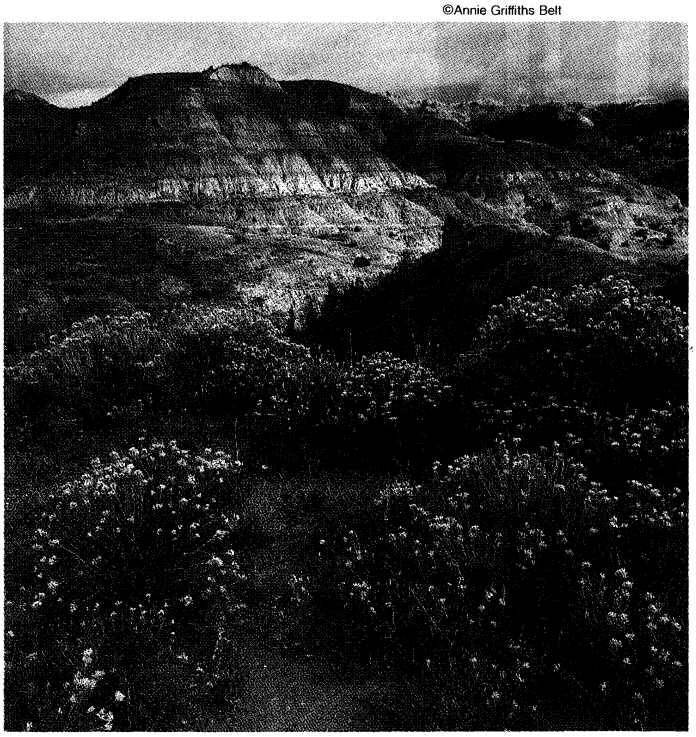
PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The eastern half of North Dakota lies in the Central Lowland, which stretches westward from the Appalachians, while the western half lies in the Great Plains, which extend to the Rocky Mountains. The state is like three broad steps rising westward: the Red River valley lies 800 to 1,000 feet (250 to 300 metres) above sea level, the Drift Prairie from 1,300 to 1,600 feet, and the Missouri Plateau from 1,800 to 2,500 feet. The highest point in the state is White Butte, at 3,506 feet (1,069 metres). The Central Lowland portion comprises the Red River valley, a flat, glacial lake bed extending from 10 to 40 miles (16 to 64 kilometres) on either side of the Red River of the North, and the Drift Prairie, a rolling country covered with glacial drift. On the west the Missouri Escarpment separates the Drift Prairie from the Great Plains. The North Dakota portion of the Great Plains is known as the Missouri Plateau. East and north of the Missouri River it is covered with a thick layer of glacial drift. The Altamont Moraine in this area, which is one of the principal flyways for migrating wildfowl, has numerous potholes, lakes, and sloughs. Saline Devils Lake, remnant of a shallow glacial sea and the largest natural body of water in the state, has undergone remarkable and not altogether explained variations in level since the 1880s. It is the subject of many local, particularly Indian, legends. Like the Drift Prairie, this region has a young drainage system, there being few rivers in areas once covered by the great ice sheets of the geologically recent pact.

Three regions

About two-fifths of the state is drained by the systems of the Red and Souris rivers, whose waters flow eventually into Hudson Bay. The Missouri Plateau and the James River system form a part of the drainage of the Missouri, which drains almost two-fifths of the state and flows into the Mississippi and thence into the Gulf of Mexico. West of the Missouri River the landscape has been shaped by running water that has carried away as much as 1,000 feet of sedimentary deposits. In some places, especially along the Little Missouri River, it has carved spectacular cliffs,

©Annie Griffiths Belt



The Badlands of the Little Missouri River in Theodore Roosevelt National Memorial Park, southwestern North Dakota.

buttes, and valleys that form a landscape known as the North Dakota Badlands.

Climate. North Dakota's location at the centre of the North American continent gives the state a continental climate: hot summers and cold winters, warm days and cool nights in summer, low humidity and low precipitation, and much wind and sunshine. The western part of the state has lower humidity, lower precipitation, and milder winters than the eastern half. For the state as a whole, the average annual precipitation is about 17 inches (430 millimetres). In July the average daytime high temperatures range from 88° F (31° C) in the south to 82° F (28° C) in the north. In January the average highs range from 26° F (−3° C) in the southwest to 10° F (−12° C) in the northeast. The growing season in North Dakota varies considerably, from 134 days at Williston, in the northwest, to 104 days at Langdon, in the northeast.

Original grasslands

Plant and animal life. Before settlement, 95 percent of the state was covered by grass, low precipitation, drought, and grass fires having inhibited tree growth. Long-lived perennial grasses begin to grow early in the spring, produce seed quickly, and go into a dormant state in drought. They protect the soil from erosion and provide food for grazing animals. The heavy grass cover of the Red River valley and the Drift Prairie formed black soils, while the lighter grass cover of the Missouri Plateau formed lighter, thinner, dark brown soils. The grassland was a natural habitat for great herds of buffalo and antelope. Belts of timber and brush along the rivers provided homes for animals such as white-tailed deer, elk, and bear. The remaining small buffalo herds are protected in parks, and, in arable parts of the state, croplands have replaced the virgin prairie.

Settlement patterns. The regions are reflected to some degree in the character of the people. The inhabitants of the Missouri Plateau tend to be more Western in their manners and dress, whereas those of the Red River valley are more Eastern. The Drift Prairie is a transition zone in this respect, a function it serves also in relation to climatic patterns and plant and animal distribution.

North Dakota is a land of large farms and ranches; its vast, open country has few fences. There is beauty in the great fields and pastures, the big sky, the endless view of flat or rolling prairie with the black earth of the plowed land, the green blanket of a new crop, or the yellow cover of ripened grain. The clean, dry air and the bright sun give a wholesome look to the land, but the large holdings, which average more than 1,200 acres (480 hectares), make the countryside seem lonely and almost uninhabited.

With the diminishing of farm population, characteristic of the second half of the 20th century, many small towns have also disappeared, while in others businesses and houses stand empty. The larger cities provide a sharp contrast, with their new stores, public buildings, and housing developments and their air of vigour and prosperity. The sparsity of population affects not only the state's economy but also the character of the people, who tend to be friendly, helpful, and straightforward. Distances create isolation, but the electronic media and improved transportation have reduced many of its effects.

The people. When white traders reached what was to become North Dakota, several Indian peoples lived in the region: Mandan, Hidatsa, and Arikara along the Missouri River, Chippewa and Cree in the northeast, Assiniboin in the north, Yanktonia Sioux and Wahpeton Dakota in the southeast, and Teton Sioux and Crow in the west. The fur trade brought the French, Scots, English, Canadians, and Americans, and, by 1800 the Métis, of mixed white and Indian ancestry, were an established element.

The earliest white settlers included many Norwegians, Canadians, and Germans who had earlier migrated to Russia. By 1890 the foreign-born constituted about 43 percent of the population, a higher percentage than in any other state; and in the census of 1920, when settlement had been completed, only 32 percent of the white population was of native-born American parentage. By 1980, however, large-scale immigration had ceased, and less than 5 percent of the population was foreign-born.

American Indians are the largest minority group in the state. They constitute about 3 percent of the total population. Unemployment and ill health still occur at a somewhat higher rate among North Dakota Indians than in the non-Indian population. Many Indians, however, are successful farmers, ranchers, professionals, athletes, and politicians. A two-year college is maintained by each of four of the state's reservations.

Most North Dakotans have religious affiliations. Almost one-half are Lutherans, more than one-third are Roman Catholics, and most of the remainder are divided among other Christian denominations. Jewish congregations have existed from before statehood.

The economy. North Dakota's cool, subhumid climate and its location far from the nation's markets have helped to shape its economy. Among the western north central states, North Dakota has one of the lowest farm incomes, the lowest average rainfall and temperature, the shortest growing season, and the least manufacturing.

Major crops

Agriculture. The state produces beef cattle, wheat, rye, and oats and ranks first in the nation in the production of barley, sunflower, and flaxseed. It also sends dairy products, sugar beets, potatoes, and other agricultural commodities to outside markets, from which it buys its farm machinery, building materials, trucks, automobiles, and most of its consumer goods. Wheat is the most important source of farm income.

Although agricultural production largely pays for the goods North Dakotans buy in outside markets, it employs less than one-fifth of the labour force. Since World War II, rapid improvements in farming efficiency have led to larger farms, fewer in number and supporting directly less of the population. Farmers' capital wealth, in land holdings and machinery, is often great, but annual income on that wealth is not proportionate.

Industry. The discovery of oil at Tioga in 1951 led to North Dakota's becoming one of the largest producers of crude petroleum in the nation, and production of electrical power grew greatly after the mid-20th century. In that period also the economy was stimulated by construction of the Garrison Dam, air force bases, and highways and by rural electrification. Manufacturing accounts for only about 10 percent of the state's income, and its lignite, the largest supply of solid fuel in the United States, plays a relatively minor role in the state's economy.

Transportation. Intrastate and interstate traffic moves primarily over east–west and southeast–northwest routes in North Dakota and secondarily over north–south routes. Fargo is the main centre for intrastate traffic; interstate traffic moves between it and other trading centres in North Dakota and Minneapolis–St. Paul, the nearest metropolis, and the Pacific Northwest. North Dakota has a well-developed system of rail lines, and airlines provide scheduled service to a number of cities.

Administration and social conditions. *Government.* The constitution of 1889 provides for a government consisting of a governor elected for a four-year term, 13 elected heads of executive departments, a bicameral legislature of 53 senators and 106 representatives, and several levels of state courts. There are almost 150 departments, boards, and agencies, as well as two state-owned industries. North Dakota ranks in the lower half of states in terms of property tax rates and all taxes per capita.

Local government

North Dakota's 53 counties, with populations ranging from less than 1,500 to more than 80,000, all elect commissions and certain officers. Most of the counties are further divided into townships totaling more than 1,300, all of which elect administrative officers. Of the more than 360 municipalities (designated by law as cities regardless of size), the vast majority have mayor–council governments. With about 500 school districts and more than 700 special-purpose districts, North Dakota has some 3,000 units of local government, more per capita than any other state.

The unified court system comprises the five-justice Supreme Court, 26 district courts in seven judicial districts, 26 county court judges, some serving several counties, and about 140 municipal courts. Judges are elected.

Relatively few North Dakotans vote a party ticket. While Republican candidates have predominated in presidential and state-legislature elections, Democratic governors and U.S. congressmen have not been uncommon.

Education. The great majority of North Dakotans finish high school, and most of these go on to postsecondary education within the state. Many rural elementary and high schools, however, have always been too small to provide full programs. In the late 1980s, for example, 12 high schools enrolled 20 or fewer students.

Access to publicly supported higher education has been greatly prized, with the consequence that in spite of its small population the state finances two full-program universities—the University of North Dakota at Grand Forks, founded in 1883, and North Dakota State at Fargo, founded in 1890. Both offer graduate and professional work and enroll between them well over half of the state's postsecondary students. Private institutions (most notably Jamestown College, founded in 1883, and the University of Mary, at Bismarck, founded in 1959) account for a small percentage of higher-education students.

Health services

Health and welfare. North Dakotans receive excellent medical care despite the state's low population density. Although some towns of less than 1,000 population have a doctor, medical practice is concentrated in the four larger cities—Fargo, Bismarck, Grand Forks, and Minot—often in group practice in well-equipped clinics. Few people live more than a two-hour drive from one of the centres. The state has more than 50 general hospitals, a rehabilitation centre, several regional mental health centres, and a state hospital for the mentally ill. The state health department and smaller health districts provide public health services. Colleges of medicine and nursing at the University of North Dakota educate practitioners, and this university and others train additional health-field personnel.

Economic assistance and a variety of social services are provided by the state human services department, county social services boards, and private welfare agencies, especially denominational groups. The state department provides aid to the aged, blind, and disabled and to dependent children; it also directs regional human-services centres. The county boards administer general assistance and medical aid for the aged. Federal sources provide more than two-thirds of the funds for welfare recipients. Relatively few able-bodied adults require or request assistance.

Cultural life. The traditional North Dakota spirit of self-reliance and voluntary cooperation is reflected in the cultural life of the state. Without a large metropolitan centre, the cities and towns with universities or colleges provide the main cultural leadership. Symphony orchestras have headquarters in Fargo, Minot, and Grand Forks, though they make appearances throughout the state. The North Dakota Ballet is located in Grand Forks, where in 1971 the University of North Dakota established the state's first College of Fine Arts. Most of the community art associations, public concert associations, and theatre groups are also located in college or university towns. A summer School of Fine Arts is held at the International Peace Garden, a park located on the border between North Dakota and Manitoba near the Turtle Mountain area.

There is some federal-assistance funding for arts projects in the state, but most other funds for the arts, apart from those expended by educational institutions, have had to come from public subscription. In 1971, however, a small state appropriation made to the North Dakota Council on the Arts and Humanities, the agency through which federal funds for the arts are dispensed, was considered the beginning of a long-term state commitment to the arts.

Among the weakest aspects of North Dakota's cultural life is library service. Because the larger part of the population lives in the country or in small villages, about one-fifth of the people have virtually no contact with library facilities. The libraries of towns with populations of 5,000 or more vary widely in their adequacy. Civic leaders have sought to meet the needs of the rural population with county and regional libraries and with bookmobiles.

Acclaimed North Dakotan writers have included Lois Phillips Hudson, Larry Woiwode, and Louise Erdrich. Elizabeth Hampsten, in *Read This Only to Yourself: The Private Writings of Midwestern Women, 1880–1910* (1982) and other works, has provided collections of previously unpublished writings of North Dakota women in the settlement years.

Indian arts and other ethnic activity

Indigenous folk traditions continue within the state among the Sioux peoples of Fort Totten and Standing Rock Indian Reservation, among the Plains Ojibwa (locally called Chippewa) people of the Turtle Mountain Reservation and area, and among the people of the Three Tribes—the Arikara, the Hidatsa, and the Mandan—of Fort Berthold. Traditional music and dances, together with beadwork and other crafts, attract many art lovers to the state. The durable and attractive pottery of the Three Tribes is particularly sought after.

Scandinavian cultural traditions remain vigorous. Although none of the 50 Norwegian-language newspapers published between 1878 and 1955 survives, Norwegian language and literature are taught at the University of North Dakota and in several elementary schools. The Sons of Norway have some 8,000 members in the state. Norwegian costumes, customs, and cookery are observed on many occasions but especially on Norwegian Constitution Day, May 17. North Dakotans of Icelandic, Czech, and German ancestry also retain some ethnic customs, and in many families the ancestral languages are still spoken.

North Dakota has several state parks, which draw about a million visitors each year, and the Theodore Roosevelt National Park. Of the state's historic sites, several are also in the National Register of Historic Places. The North Dakota Heritage Center in Bismarck is the most comprehensive of the state's museums, but many smaller museums of interest are to be found throughout the state, sometimes in very small centres of population.

The individualistic character of North Dakotans is reflected in their sports and pastimes, which include fishing, hunting, and trapping. Snowmobiling, ice skating, skiing, and ice hockey are popular winter sports.

HISTORY

The United States acquired the lands drained by the Red and Souris river systems (from 1670 parts of Rupert's Land) by the Rush–Bagot Agreement of 1817, and the remainder of what became North Dakota from France by the Louisiana Purchase in 1803. The recorded history of the state falls into three periods: the period of Indian trade, from about 1738 to 1871; of white settlement, from 1871 to 1915; and of adaptation, since 1915.

Early European visitors

Explorers and traders. Although European goods were traded among the Indian peoples before his arrival, the first known white visitor to North Dakota was Pierre Gaultier de Varennes, Lord de La Vérendrye, a native of Canada who visited a cluster of earthen-lodge villages near present-day Bismarck in 1738. Traders from Hudson Bay and Montreal began to go to the area on a regular basis in the 1790s. The best-known visitors of the early years were Meriwether Lewis and William Clark, whose expedition made winter camp in 1804–05 near present-day Stanton.

In the 1820s and '30s American traders made the upper Missouri country a hinterland to St. Louis. They brought in guns, kettles, blankets, and axes, as well as liquor and disease. The white man's goods made the Indians dependent on the traders, his liquor demoralized them, and his diseases killed them. In 1837 smallpox, carried up the Missouri by passengers aboard a steamboat, reduced the Mandan population from about 1,800 to 125 in a few months. Indian hostility grew when steamboat traffic increased after the discovery of gold in Montana in 1862 and when the U.S. Army built forts along the rivers. In 1876 Lieutenant Colonel George A. Custer and the 7th Cavalry set out from Fort Abraham Lincoln, south of present-day Mandan, for their fateful encounter with the Sioux and Cheyenne on the Little Bighorn River in Montana.

Pioneering and statehood. The fur trade declined in the 1860s, and white settlement began in earnest in 1871, when railroads reached the Red River from St. Paul and Duluth, Minn. A flood of pioneers acquired land under the Homestead Act and turned to wheat farming. During the period known as the Dakota Boom (from 1878 to 1886), the many giant farms publicized the new country, and North Dakota wheat made Minneapolis, Minn., the milling centre of the nation in the 1880s. The Northern Pacific and Great Northern railroads vied with one another to reach the richest grain centres. Dependence on

Populist strength

wheat unified the farmers and strengthened the populist revolt against eastern monopolistic practices. The Dakota Territory was divided in 1889, and both North and South Dakota were admitted to the Union on Nov. 2, 1889.

The modern state. Revolt against outside exploitation reached a climax soon after the period of pioneer settlement ended in 1915. Controlling the state government after the 1918 election, the Nonpartisan League enacted a socialist program that included a state-owned bank and a flour mill and grain elevator. The league soon lost political control, but the North Dakota Farmers Union (founded in 1927) launched a strong cooperative movement to control the selling of grain and the purchase of farm supplies. Such radical farm movements made many North Dakotans oppose American intervention in both world wars, because they identified participation with war profits for Wall Street.

Since 1915 North Dakota's history has been marked by continuing adaptations to the cool, subhumid grassland environment, the most important of these being the increasing mechanization of agriculture, the enlargement of farms, the loss of rural population, and the widespread use of the automobile. After World War II came rural electrification, soil conservation, and highway construction. In the 1950s North Dakota became an oil-producing state, and in the 1960s air bases, missile sites, and antiballistic-missile installations were built there.

In the decades following, the state's economy and its population levels and distribution reacted sensitively to external forces, especially variations in the pricing of both fossil fuels and agricultural products, in changes in weather, in federal support programs, and in international currencies. In comparison with states having larger populations and broader tax bases, in North Dakota such influences affect disproportionately even the small percentage of wealth derived from manufacture and tourism. North Dakotans generally retain a basic stability, balancing realism with long-range optimism and seeking new forms of economic development while preserving their love of the land and what it can produce. (E.B.R./B.O'K.)

Ohio

A representative state

The first state to be carved from the Northwest Territory when it became the 17th member of the Union on March 1, 1803, Ohio, in the 20th century, reflects the urbanized, industrialized, and ethnically mixed United States that developed from an earlier agrarian period. The pattern of its life is so representative of the nation as a whole that it is often used to test attitudes, ideas, and programs in education, politics, and industry. Significantly, Ohio has supplied by birth or residence eight U.S. presidents—William H. Harrison, Ulysses S. Grant, Rutherford B. Hayes, James A. Garfield, Benjamin Harrison, William McKinley, William H. Taft, and Warren G. Harding.

The state's accessibility has been perhaps the key factor in its growth. Its location between the Eastern Seaboard and the growing Midwest and its lack of natural barriers to movement made it a corridor for east–west travel. With Lake Erie on the north, Pennsylvania on the east, West Virginia and Kentucky on the southeast and south, Indiana on the west, and Michigan on the northwest, Ohio lies in the heart of the nation's old industrial belt, close to major resources of raw material and labour and to the markets of the East, Midwest, and South.

Ohio's area of 41,330 square miles (107,044 square kilometres), excluding 3,457 square miles in Lake Erie, ranks only 35th in size among the states; it is one of the smallest states west of the Appalachian Mountains. The state ranks near the top, however, in population. Ohio's capital, after being located in Chillicothe and Zanesville during the early years of statehood, was finally established in newly founded and centrally located Columbus in 1816. The state takes its name from that of the river, an Iroquoian word meaning "great water."

PHYSICAL AND HUMAN GEOGRAPHY

The land. The topography, river systems, groundwater, and soils in most of Ohio are the products of glacial activity. These factors have strongly influenced the patterns of human settlement and land use.

Relief. Ohio straddles two major subregions of the Interior Lowlands physiographic region of the United States: the Appalachian Plateau on the east and the Central Lowlands on the west. These two subregions divide the state almost in half. The Appalachian Plateau reaches westward from Pennsylvania and West Virginia into the counties along Ohio's eastern border, from near Lake Erie to the Ohio River. The northeast is only partially glaciated, while the southeast is unglaciated terrain. Throughout the plateau the land is dissected by rivers winding among steep hills, and many elevations reach 1,300 feet (395 metres).

West of the Appalachian Plateau stretch the Central Lowlands. The eastern lake section, or Lake Plains, stretch along Lake Erie to the northwestern counties and the Michigan border and then extend irregularly to the south. These level to slightly rolling lands were once under water, and the swampiness of the northwest, around Toledo, posed obstacles to settlement before drainage made it more arable. The Central, or Till, Plains, which extend westward toward the Mississippi River, include parts of western and southwestern Ohio and provide a deep soil. This region contains the state's highest and lowest points: Campbell Hill, the highest point, at 1,550 feet (472 metres), is located near Bellefontaine; the lowest, at 433 feet (132 metres), lies at the confluence of the Miami and Ohio rivers, near Cincinnati.

Drainage. The principal water sources are rain-fed streams, lakes, and reservoirs. Floods, once prevalent, are controlled by state and federal dams and other conservation measures. Groundwater is used widely for public supplies, though the industrial and population centres have limited resources. Huge stores of these waters are buried in preglacial valleys in central and south central Ohio.

Lake Erie, with an average depth of only 62 feet (18.9 metres), is the shallowest of the Great Lakes. It is also the most tempestuous, with frontal storms often roaring across it from Canada, and the most liable to shoreline erosion, harbour silting, and filling of its bed. Its shallowness, coupled with the concentration of population and industrial plants in its watersheds, led to severe pollution. Programs in various areas deal with the problems of the lake, which continues to be the principal source of water for many lakeside cities. Attempts to abate pollution in Lake Erie have begun to show signs of success. Fish have returned to previously uninhabitable waters, and a revival of sport fishing and recreational activity has stimulated economic growth along the shoreline.

Lake Erie and the river systems

A low watershed separates the 20 percent of Ohio drained by the Maumee, Cuyahoga, and other rivers emptying into Lake Erie from the 80 percent drained by the Miami, Scioto, Muskingum, and others flowing into the Ohio–Mississippi system. The Ohio, only a tiny part of which is under state jurisdiction, is canalized and channeled for its entire length, as is the Muskingum from Zanesville to Marietta. More than 100 lakes and reservoirs supply recreational and industrial water.

Soils. Most of Ohio's soils are well suited to agriculture. The Till Plains soils are mainly rich glacial limestone, but the Lake Plains are the most productive. The sandstone soils of central and northeastern Ohio are best adapted to pasturelands, while the thin-soiled and heavily eroded hilly areas of the southeast support little productive farming except in river bottomlands.

Climate. Temperatures in Ohio are similar to those across the north central and eastern United States, with summer highs and winter lows seldom reaching 100° F (38° C) and −20° F (−29° C), respectively. The state is open to cold, dry fronts from Canada and to warm, moist fronts from the Gulf of Mexico. The frequent meeting of such fronts causes much of the state's precipitation, which totals some 38 inches (965 millimetres) annually, including an average annual snowfall of 28 inches.

Plant and animal life. The great hardwood forests that covered 95 percent of Ohio prior to European settlement have been reduced to less than 25 percent. The glaciated areas have stands of timber that include oak, ash, maple, walnut, basswood, hickory, and beech. Much cutover land

Forested areas

in the southeastern and south central regions has been reforested. Both wild and domestic flowers abound, though the clover, wild rye, and bluegrass of early Ohio are gone.

Of the 350 bird species found in Ohio, at least 180 are native. Among the 170 fish species are bass, trout, and perch, while the 60 or more species of mammals include deer, opossum, fox, skunk, groundhog, and rabbit.

Settlement patterns. Despite its many large urban areas, more than four-fifths of Ohio is cropland and forest. The urban areas of Ohio first exceeded the rural in population in 1910, and by the late 20th century the urban population made up about 75 percent of the total. Areas outside central cities contain more than half of the population, and Ohio's large cities are following the national pattern of losing population to surrounding suburban areas. The growth of Columbus proper is largely attributable to annexation of township lands.

It is possible to identify several regions throughout Ohio that have distinctive landforms, human and physical resources, and economic characteristics.

The Maumee valley region in the northwest is primarily agricultural. Corn, soybeans, and wheat, as well as hogs and dairy and poultry products, are important. Its largest city, Lima, is an industrial and market centre. The Lake Plains region on the southwestern shores of Lake Erie also has flat, fertile plains with highly productive soils. Toledo, the major city, is an important centre in the Great Lakes industrial belt and a major coal-handling port. It supplies glass and transportation equipment to nearby Detroit and processes the farm products of the region.

The Lakeshore and Uplands region in the north and northeast, with approximately one-fifth of the state's land, contains Ohio's largest industrial concentration and holds more than two-fifths of its population. Cleveland is the industrial, financial, and cultural centre. Akron is a centre of rubber and polymer industries and of trucking. Youngstown is a major metal producer and fabricator, and Canton specializes in the production of such items as roller bearings, bank vaults, and vacuum cleaners.

The Sandusky valley region in north central Ohio is basically agricultural, though the small cities of Marion, Galion, and Bucyrus have some manufacturing. The Scioto valley region of rolling plains in central Ohio has a diversified economic base. Columbus, its central city, is the home of the state government and of numerous educational institutions, including Ohio State University. About half of the working force is employed in government, education, finance, and other service occupations.

The Tuscarawas valley region of eastern Ohio and the Ohio valley region in the south and southeast are predominantly rural. Terrain limits agricultural productivity in both regions. In the southwestern part of the region wheat, corn, tobacco, and hogs are the principal products. Mining and lumbering provide the largest proportion of income in the southeastern part. Stone, clay products, chemicals, and metal fabrication are major industries.

The Miami valley region, in southwestern Ohio, centres on Cincinnati and Dayton. Cincinnati is important in the production of machine tools and other manufactures. Dayton produces business machines, computers, and automotive products. Nearby Fairborn is the home of Wright-Patterson Air Force Base, a major research centre.

The people. The people who laid the foundations of Ohio came from the older seaboard states. The first permanent white settlement in Ohio and the Northwest Territory was at Marietta in 1788 by a company of New Englanders who had fought in the Revolution. In the same year a group from New Jersey established a settlement near Cincinnati, and in the next few years other villages sprang up. In the south, particularly in the Virginia Military District between the Scioto and Little Miami rivers, many of the settlers came from Virginia and Kentucky. In 1796 the Western Reserve in northeastern Ohio was first settled, mainly by New Englanders from Connecticut.

Early European settlers

Many Protestant Scotch-Irish settlers came from the Middle Atlantic and Southern states. Prior to 1830, Pennsylvania Germans and Swiss came to the east central area. After 1830 settlers came directly from Germany and Ireland. Many Irish came to work on the Ohio canals and stayed on, and when the railroads were built, the Irish and German workers remained as permanent settlers. Germans who drained the Black Swamp country of the northwest stayed on to develop the resultant farmlands.

After 1830 Roman Catholic immigrants from southern Ireland settled in such cities as Cleveland, Columbus, and Cincinnati, where by 1850 they were second in number to the Germans among foreign-born residents. In Lancaster immigrants from Württemberg joined the Pennsylvania Germans who had founded the city.

In northeastern Ohio, Canton and Steubenville were established by Pennsylvania Germans. German settlers also were attracted to the rolling surface and fertile soil of Wayne county, which became one of the top agricultural counties in the nation. German-speaking Moravian missionaries under the leadership of John Heckewelder and David Zeisberger came to the Tuscarawas valley to Christianize the Indians in the early 1770s. In 1817 Joseph Bimeler founded an experimental communist settlement in Zoar that lasted until 1898. The Swiss settled around Dover and Sugar Creek in Tuscarawas county, as well as in Monroe county. In Holmes county, Amish immigrants from Germany and Switzerland established settlements that still remain. There are now more Mennonites in Ohio than in Switzerland, and the state's Old Order Amish community is the largest in the world. Quakers from the Middle Atlantic states and the South settled in eastern and southwestern Ohio early in the 19th century.

The Welsh arrived in the early 19th century to develop the mineral resources in several regions of Ohio. They were especially numerous in Jackson county, and for a long period Welsh was the only language that was spoken there. The Eisteddfod, a festival of Welsh bards, and other elements of Welsh culture and music flourished. The language persisted to the third generation in many communities, with old Welsh songs passed on from one generation to the next.

In 1850 the principal racial stock was Scotch-Irish, although the Germans and the English also were important elements. In 1870 nearly 14 percent of Ohio's population and 40 percent of Cleveland's were foreign-born. The New England character of northern Ohio's beginnings was changing. Each new group established its own newspapers, clubs, social life, and churches.

Increasing numbers of immigrants from eastern and southern Europe came to Ohio after 1880. By 1920 large numbers of Italians, Poles, Hungarians, Russians, and other groups had come to Cleveland, Toledo, Youngstown, and other industrial cities. Southern whites from Appalachia came in large numbers to Akron, Dayton, and Cincinnati. Cleveland, however, became Ohio's most ethnically diverse city. Its foreign-born population was supplemented between 1880 and 1890 by new arrivals from Austria, The Netherlands, Russia, Hungary, Portugal, Greece, China, Japan, Turkey, and Mexico. The city's culture eventually was enriched by some 50 groups with different languages and backgrounds. Many new Roman Catholic and Eastern Orthodox churches, as well as synagogues, were built. The Greeks brought their coffeehouses, and the Slovenes and Poles brought their social halls. The descendants of these ethnic groups have firmly established themselves in the social, economic, and political life of the state.

The diversity of Cleveland's population

The changing character of the state is shown as well in the growth of the black population. In 1850 the black population of Ohio was about 25,000. By 1870 it had risen to more than 63,000. Most blacks lived in southern Ohio, where Wilberforce University, one of the first permanent black educational institutions, was established. By 1980 Ohio's black population of more than 1,000,000, most of it in the cities, constituted 10 percent of the total.

The economy. A good location, a rich store of natural resources, productive soils, cheap energy, and ample transportation facilities have made Ohio one of the great industrial states. More than half of the nation's population is within 500 miles (800 kilometres) of its borders; and coal, oil, natural gas, clay, salt, limestone, sandstone, shales, and gypsum help supply local industries. About two-thirds of the raw materials processed in Ohio's factories come from its own resources. More than one-fourth of the labour

force is employed in manufacturing, although heavy basic industrial production has declined since the 1970s. Ohio's continuing activity in agriculture and mineral production provides economic balance and diversity.

Mining. Ohio's mineral resources are heavily exploited. Coal production accounts for the highest return, followed by industrial minerals, gas, and oil.

Coal was discovered in Ohio as early as 1808. It was adapted for use with iron ore and limestone in the pioneer iron-making enterprises that sprang up in the eastern and south central parts of the state. Later, the discovery of deposits of iron ore in the upper Midwest gave rise to important iron and steel centres in northern Ohio. Usable coal supplies are found throughout eastern and southeastern Ohio. Most coal is produced by strip mining. Enforcement of laws regulating strip mining and requiring restoration has eased environmental problems, but citizen groups still battle for stronger safeguards.

Limestone is used in many construction and manufacturing processes. Ohio is among the top states in sandstone and in sand and gravel production. The abundance and quality of surface clays, plastic fireclays, shales, and some gypsum and peat have made Ohio a leader in the manufacture of ceramic products. The majority of its extensive salt production comes from large rock-salt mines, with the remainder from brine. It is estimated that the state's salt deposits could supply the nation's need for centuries.

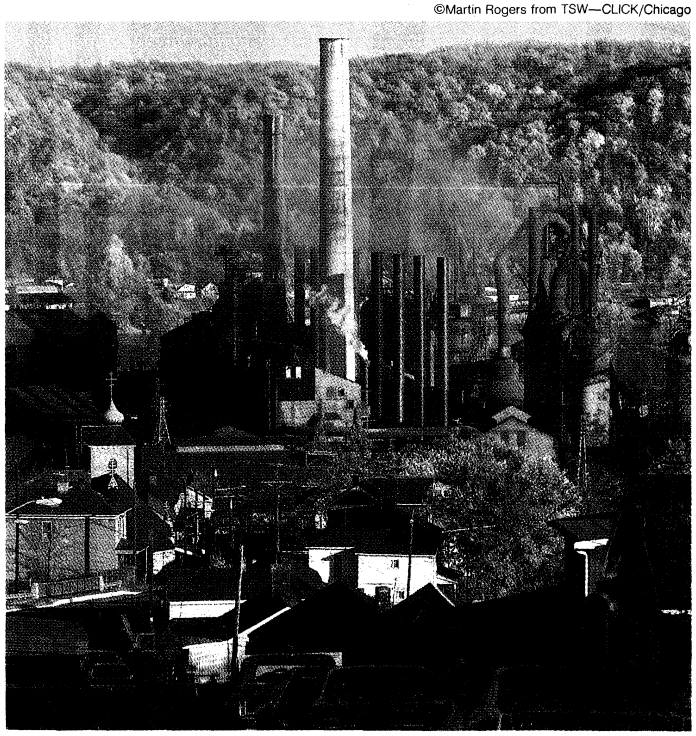
Petroleum and natural gas production

Ohio has been a producer of oil and natural gas since 1860, but by 1900 production in the state had declined. In the early 1960s, however, new oil and gas deposits were discovered, and the industry revived modestly. Ohio must import substantial amounts of oil and natural gas.

Agriculture. In 1850 Ohio ranked first among the states in agricultural production, and it has continued to rank near the top. Although farm acreage and the number of farms and farmers have decreased, nearly two-thirds of Ohio is still farmland. Commercial farming, or agribusiness, largely has supplanted the family farm in producing cash crops, but the Amish, using old-fashioned techniques, and others are still able to make the family farm profitable. Ohio produces corn (maize), wheat, oats, soybeans, and hay, and it maintains large marketing inventories of fruit, feed, and vegetables, as well as livestock and poultry.

Industry. Manufacturing is Ohio's most important economic activity and represents the largest single segment of the state's employment. Transportation equipment, nonelectrical machinery, and fabricated metal products are the largest manufacturing enterprises in terms of employment, despite the movement of much heavy manufacturing to other regions of the nation and the world. The steel and clay products industries are also significant.

©Martin Rogers from TSW—CLICK/Chicago



Steelworks at Mingo Junction, eastern Ohio, on the Ohio River.

Transportation. Ohio's chief transportation system in the first years of statehood, as in the territorial period, was its water routes. Lake Erie and the Ohio River provided east–west passage for Indian traders, pioneers, and settlers, and many rivers provided access to the interior. Shortly after statehood the development of transportation facilities began. Between 1825 and 1838 the federal government extended the Cumberland (National) Road across Ohio. In 1811 the first steamboats appeared on the Ohio River, and in the 1820s the era of canal building began and lasted for some 30 years. The first railroad was constructed in 1832, and in the 1850s the first great east–west rail lines were constructed across Ohio.

Ohio's transportation facilities play a major role in moving passengers and goods by highway, railroad, river, lake, and air. The shipping to and from its lake ports is worldwide, and the Ohio River carries more tonnage than the Panama Canal. The railroad mileage is among the nation's largest. The pioneering experiments of Dayton's Wright brothers, Orville and Wilbur, led to the first successful aircraft flight, at Kitty Hawk, N.C., in 1903, and Ohio is now both a testing and a commercial aviation centre.

Administration and social conditions. *Government.* Ohio's present constitution was adopted in 1851. The executive branch is composed of the governor, lieutenant governor, attorney general, secretary of state, auditor, and treasurer, all elected for four-year terms. The General Assembly consists of the Senate, with 33 members elected for four-year terms, and the House of Representatives, whose 99 members serve two-year terms. It has broad powers in policy formulation and monetary appropriation. The judiciary comprises the seven-member Supreme Court, 12 courts of appeals, courts of common pleas and of probate in each of the 88 counties, and such other lower courts as the legislature may establish. All judges are elected for six-year terms.

Local government

Each county except Summit, which has a home-rule charter, exists as a quasi-municipal corporation, an arm of the state government but without general authority of self-government in the legislative field. Most larger cities operate under home-rule charters that permit them to choose the form of government most suitable to their needs. The mayor–council type is most common, though Cincinnati and several other cities operate under a city-manager–council plan. The township, Ohio's oldest form of government, remains important, though the number is diminishing as townships are annexed into municipalities or as newly incorporated villages assume their functions.

State laws carefully prescribe the rules for forming and running political parties, conducting elections, and balloting. The two-party system has prevailed generally, but Ohio has produced such minor-party leaders as Norman Thomas, many times a presidential candidate on the Socialist Party ticket; Victoria Claflin Woodhull, in 1872 the first woman to run for president, with the Equal Rights Party; and Jacob Coxey, who led the march of "Coxey's Army" from Massillon, Ohio, to Washington, D.C., in 1894 to demand various economic reforms.

Since its inception the Republican Party has been slightly more successful than the Democratic in statewide elections. In national politics the parties are evenly matched. Although dynasties are rare in Ohio political life, the Tafts of Cincinnati may constitute one. William Howard Taft served as president and as chief justice of the U.S. Supreme Court. His father had been secretary of war and attorney general under President Ulysses S. Grant and was later U.S. minister to Russia and Austria-Hungary. His son, Robert A. Taft, served in the U.S. Senate from 1939 to 1953, and his grandson, Robert A. Taft, Jr., served in the U.S. Senate from 1971 to 1976.

Education. Authority for Ohio's primary and secondary public schools rests with the state legislature. A State Board of Education, comprising one elected member from each

congressional district of the state, appoints a superintendent of public instruction, who heads the Ohio Department of Education. City, county, local, and exempted school districts exist as subdivisions of the state organization. The direction, administration, and financing—shared by state and local government taxing units—of the public schools is delegated to the individual school districts. Private and parochial schools, though governed by the same laws that apply to public schools, receive no direct state support.

Higher education

Ohio often has been called a "land of schools and colleges," and it ranks among the top 10 states in the number of accredited colleges. Ohio University was established by Ohio's first legislature in 1804 as the first public institution of higher education west of the Alleghenies. In 1809 Miami University became the second. Ohio State University, founded in 1870, is the largest state-assisted university. A land-grant college and a major graduate and professional centre, it also has one of the largest undergraduate enrollments in the nation. Ohio has several comprehensive state universities and numerous branch campuses, technical colleges, and community colleges. Many of Ohio's small independent colleges have made distinguished contributions to the state and have pioneered in education in various ways. Oberlin College, founded in 1833, became the first coeducational college in the United States and one of the first to admit blacks. Antioch College (part of Antioch University), founded in 1852, is one of the nation's oldest experimental liberal arts colleges. Like some other Ohio institutions, it has implemented innovative programs for advancing minority students.

Health and Welfare. The aged and poor, the blind and disabled, and crippled and dependent children are among the groups that benefit from the welfare activities of several state agencies. Other state bodies oversee programs in the prevention and cure of illnesses. Ohio has six medical schools and an osteopathic college, as well as many strong regional hospitals. The Cleveland Clinic has an international reputation. A youth commission operates diagnostic and training centres, youth corps, and schools. State activities with labour and industry include programs in employment and unemployment services, industrial safety, and worker's compensation.

Cultural life. Early settlers of Ohio put the stamp of their former homes—New England, the Middle Atlantic states, Kentucky, and Virginia—on the state. Although there has not been a clearly identifiable Ohio school in any of the arts, there has been great activity in all of them.

When the log-cabin phase of early Ohio ended, most of the settlers followed the building styles that they had known in their former homes. In the Virginia Military District the red-brick and stone houses were built in the Southern Federal style. In the Western Reserve and the Marietta area the New England influence was manifested in the colonial and modified Georgian styles. Later developments tended to follow the fashions of American architecture in general, most of them revivals of earlier European modes such as Greek, Gothic, and Romanesque.

The state has produced such diverse writers as William Dean Howells, Ambrose Bierce, Paul Laurence Dunbar, Brand Whitlock, Charles F. Browne ("Artemus Ward"), David Ross Locke ("Petroleum V. Nasby"), Sherwood Anderson, Louis Bromfield, and James Thurber, many of whom drew upon their Ohio background.

Music and other performing arts

The Cleveland Orchestra is among the finest in the world, and the symphony orchestra of Cincinnati (once considered the musical centre of the inland United States) is also renowned. The Blossom Music Center, located between Cleveland and Akron, is the site of a summer festival. Programs in music, theatre, dance, and the visual arts abound in Ohio's colleges and universities. With community theatres and arts centres, they serve as the cultural hub for many cities and towns. The Cleveland Play House and the Karamu House, which attempts to bridge black and white cultures, also in Cleveland, have long had a national reputation. The Cincinnati Playhouse in the Park, noted for its experimentation, and the Cincinnati Opera are among major regional companies. The Ohio Arts Council, which was established in 1965 by the state legislature, aids communities and arts organizations.

The Cleveland Museum of Art ranks among the foremost art galleries in the nation, and those in Cincinnati, Toledo, Youngstown, and Columbus also hold major collections. In addition, many historical sites are maintained by state and local societies, including Indian mounds, old forts and battle sites, reconstructions of early settlements, and graves, homesteads, and memorials to Ohio's presidents and other leading citizens.

Ohio has a well-developed system of public libraries in addition to college and university facilities and specialized libraries in many fields. The State Library of Ohio, in Columbus, serves the entire state. Bookmobile service is a feature of rural areas.

The state has a number of laboratories maintained by specialized institutes, industries, educational institutions, and government agencies. Reflecting industrial concentrations, Akron is a world centre for rubber research, and Cleveland is known for research in lighting. Battelle Memorial Institute, in Columbus, is one of the largest private research organizations in the world. A number of federal centres are devoted to aviation medicine, aeronautics and space, atomic energy, agriculture, and forestry.

Reflecting Old World origins are the Welsh Eisteddfod festivities in Cleveland, Steubenville, Lima, Columbus, and Jackson, and a German Saengerfest (Song Festival). More than 40 other nationality groups present folk music and dances at festivals throughout the state. Ohio boasts one of the nation's largest state fairs, and each county has an annual fair. Other gatherings include the Apple Festival in Jackson, the River Days Festival in Portsmouth, the Ohio Hills Folk Festival in Quaker City, and the Pumpkin Show in Circleville.

Recreational facilities include extensive state park facilities in addition to numerous municipal recreational areas. The Cuyahoga Valley National Recreation Area lies between Cleveland and Akron. Public gardens, zoos, caves and caverns, and privately run amusement parks add to Ohio's recreational repertory.

HISTORY

Prehistory and settlement. Remains of ancient peoples dating to 9000 BC have been found in Ohio. The later Adena and Hopewell cultures built elaborate burial and ceremonial mounds and fashioned exquisite artifacts. Both cultures had disappeared from the area by about AD 300–400. Present-day Ohio was largely unoccupied when Europeans first saw it in the 17th century. Villages of historic Indians—the Miami, Wyandot, Shawnee, Delaware, Mingo (Iroquois), and Ottawa—appeared in the 18th century.

Indian cultures

The long Anglo-French struggle to control the trans-Appalachian west culminated in British victory in 1763. The United States then won this region during the American Revolution. Following the Treaty of Paris (1783), Congress created the Northwest Territory north of the Ohio River and enacted the Ordinance of 1785, which established an orderly settlement pattern, and subsequently the Northwest Ordinance of 1787, which called for the creation of new states therein.

Statehood. Ohio achieved statehood in 1803, the first state to be formed entirely from the public domain. From the outset it was socially diversified. It was a theatre of war in 1812–13; Oliver Hazard Perry's victory over a British fleet on Lake Erie helped clear the area of its last threat from Indians and their British suppliers. Population swelled, aided by a newly developed network of canals, roads, and railroads. By 1850 Ohio was the third most populous state in the nation, with nearly 2,000,000 residents, and the leader in diversified agriculture.

Economic growth. Ohio's industrial structure was built between 1850 and 1880, when the value of its manufacturing grew to more than twice that of agriculture. A major stimulus was provided by the American Civil War (1861–65), in which Ohio supported the North, though there was strong antiwar sentiment in the state. After the war the growth continued, notably in the northeast and around Lake Erie. This growth led to considerable economic and social dislocation. After 1900 much attention was given to municipal reforms in Cleveland, Toledo, and other

Development as an industrial state

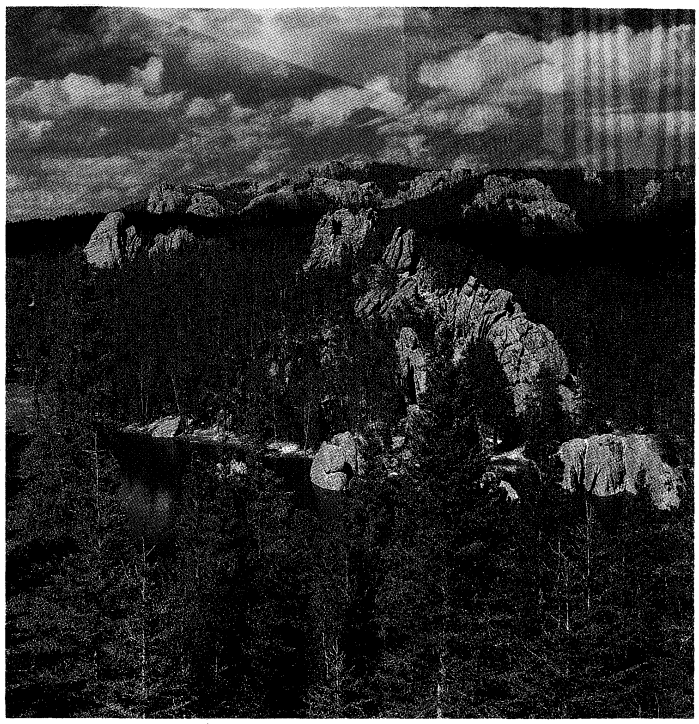
cities and to statewide programs that attempted to alleviate problems caused by industrialization. In 1920 two Ohioans, Warren G. Harding and James M. Cox, faced one another for the presidency, and Ohio has continued to play a pivotal role in national political life.

Ohio reflected the racial strife that was widespread in the United States in the 1960s, when disorders in the predominantly black Hough and Glenville districts of Cleveland took a number of lives. In 1968 Carl B. Stokes became Cleveland's mayor—the first black mayor of a large U.S. city. In May 1970 four students were killed by national guardsmen, who had been called out as a result of demonstrations at Kent State University, near Akron.

In the late 1970s Cleveland experienced severe economic problems that included the city's default on its debts. Although these difficulties were resolved, changing national and global economic conditions continued to hamper Cleveland and the state. Since the 1970s, therefore, Ohio has shifted away from manufacturing and more toward a service-oriented economy. (F.R.A./G.W.Kn.)

South Dakota

South Dakota is a Great Plains state bordered on the north by North Dakota, on the east by Minnesota and Iowa, on the south by Nebraska, and on the west by Wyoming and Montana. Its boundaries contain 77,116 square miles (199,730 square kilometres), which are split by the upper Missouri River valley into "east-river" and "west-river" regions. The state is named for the Dakota division of the Sioux Indians. Pierre, near the centre of South Dakota, is one of the smallest state capital cities; it is named for the 19th-century St. Louis, Mo., magnate Pierre Chouteau, Jr. South Dakota was admitted to the Union as the 40th state on Nov. 2, 1889.



Sylvan Lake and (centre) Harney Peak in Custer State Park, in the Black Hills region of South Dakota.
Craig Blacklock—Blacklock Nature Photography

Ethnic diversity

South Dakota remains a predominantly rural state. Slightly less than one-tenth of the population is American Indian, representing 13 tribes of the Sioux. The non-Indian populace contains more than 20 ethnic and religious groups that retain some Old Country ways—Norwegians, Swedes, Danes, Dutch, Irish, German-Russians, Mennonites, Hutterites, "Plain Germans," several subdivisions of Czechs, English, Welsh, and others. In a society of such diverse heritage there is no typical South Dakotan.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Eastern South Dakota lies within the glaciated physiographic province known as the Prairie Plains. West-river, except for the Black Hills in the southwestern corner, has the rolling topography of the unglaciated Great Plains, characterized by high buttes, rough canyons, and wide expanses of nearly level tablelands. It includes the Badlands, which extend along the White and Cheyenne rivers for more than 100 miles (160 kilometres). The eroded landscape of the Badlands has been a rich repository of fossilized prehistoric animals and the primary source of the siltation that has given the Missouri its nickname, Big Muddy.

The Black Hills

The Black Hills—two-thirds of which lie in South Dakota, with the remainder in Wyoming—constitute a dome-shaped uplift rising 3,500 feet (1,100 metres) above the surrounding terrain. Harney Peak, near the centre of the formation, at 7,242 feet (2,207 metres) above sea level, is the highest point in North America east of the Rocky Mountains.

Drainage. The Missouri River drains all of the state except the northeastern counties, from which runoff flows through Big Stone Lake and Lake Traverse, near the Hudson Bay–Gulf of Mexico divide, into the Minnesota River and the Red River of the North. In eastern South Dakota the principal tributaries of the Missouri are the Big Sioux, Vermillion, and James rivers, which flow southward. In the west-river region the Grand, Moreau, Cheyenne, Bad, and White flow eastward to drain the Black Hills and the Great Plains.

Soils. The chernozem, or black, soils, formed mostly from glacial drift and well adapted for wheat and corn (maize), cover the east-river area. Within the Great Plains province are the chestnut and Pierre, or gumbo, soils, distinguished by their heavy, sticky texture. They are well suited for grassland industries.

Climate. The climate is characterized by extremes in temperature, low precipitation, and relatively low humidity. The skies are generally clear. Cyclonic storms occur frequently in the east-river section during the spring and summer. A weather station at east central Huron reports average lows of 2° F (−17° C) and highs of 23° F (−5° C) in January, as well as average lows of 61° F (16° C) and highs of 87° F (31° C) in July. Extreme temperatures recorded are −39° F (−39° C) and 112° F (44° C). The average number of frost-free days ranges from 160 in the southeast to 110 in the Black Hills.

The average annual precipitation for the state is 19 inches (480 millimetres), but it ranges from about 24 inches along the eastern border to 14 inches or less in the northwestern corner. In the centre of the state, the transitional zone between the Prairie Plains and the Great Plains, precipitation drops from 20 inches to 18 inches. About three-fourths of the rain falls during the growing season, and snowfall ranges from about 22 to 60 inches. The Black Hills region receives more moisture, especially in winter, than the surrounding plains.

Virgin prairie grasslands

Plant and animal life. At the time of settlement by non-Indians the east-river region was a prairie covered by a thick matte of tall grasses growing to a height of three feet or more. These tall grasses have a deep root system adapted to subhumid conditions. The shortgrass species, chiefly the grama, buffalo, and western wheat grasses, are endemic to the Great Plains. Drought-resistant with a shallow root system, they mature quickly.

The wooded areas lie mainly in the Black Hills and along the buttes that rise in the northwestern part of the state. Most wooded acreage is incorporated into the Black Hills National Forest and the Custer National Forest. The western yellow, or ponderosa, pine is the chief commercial tree.

Custer State Park in the Black Hills has a large herd of bison. Other Black Hills species include antelope, deer, elk, beaver, bobcat, and porcupine. Coyotes, jackrabbits, and prairie dogs are plentiful elsewhere. The state has nearly 300 species of birds. Bald and golden eagles are found along the Missouri valley and in the Badlands. The Missouri is an important flyway for the north–south migration of waterfowl, mostly ducks and geese. South Dakota has long been a hunter's paradise because of its plentiful

supply of ring-necked pheasants, a game bird introduced into the state from Asia early in the 20th century.

Settlement patterns. During the second quarter of the 18th century, most of the Sioux, in a federation of 13 tribes, abandoned east central Minnesota and spread westward to settle their former hunting areas on the prairie and the Great Plains. The Yankton division claimed most of east-river, while the Teton tribes, together with Yanktonai, occupied west-river. Present-day South Dakota lies near the centre of some 80,000,000 acres (32,380,000 hectares) that once comprised Sioux country.

Settlement by Sioux

Between the American Revolution and the middle of the 19th century, some non-Indians settled along the Missouri valley. More immigrants, of northern and western European heritage, spread out across the east-river farmland from the mid-1850s to the late 1880s. Others, mainly of eastern European origin, joined migrants from settled U.S. communities to the east and took over the best arable and grazing areas west of the Missouri River by 1920.

Except for the Black Hills region, the west-river area has only about two persons per square mile, in contrast with the northeastern and southeastern regions, where the density is generally four times greater. Only about one-fifth of the rural residents in the state live west of the Missouri River.

The east-river section comprises two demographic areas, approximately north and south of a line running eastward from Pierre. Cash-crop farming prevails to the north, with reliance on wheat and other small-grain crops. To the south a more diversified farming economy exists, with feed grains and livestock production as its specialties.

The people. *Ethnic distribution.* The Indians of South Dakota maintain a way of life established by forebears as their reservations were founded to make room for white settlement during the 19th century. Some three-fourths of the Indians claim reservation residence. Most urban Indians in the state either live on the reservations part-time or make frequent visits to them to maintain cultural ties, while they take advantage of the job and educational opportunities offered in urban areas.

The earliest white settlers who followed British and French fur traders into what is now South Dakota came mainly from the upper Mississippi valley, many with a New England background. Immigrants arriving directly from northern and central Europe also played an important part in the colonizing process. In 1890 one-third of the white residents were foreign-born. Although today only a small percentage of the residents are foreign-born, nearly half the population are only two or three generations removed from Europe. The Americanization of the general population, white and Indian alike, is apparent. Yet ethnicity is featured in rural enclaves through the preservation of languages, religious affiliations, food preparations, festive celebrations, and ethnic educational efforts.

Immigrant pioneers

Those of Scandinavian descent make up about a third of the foreign stock. The Norwegians are located mostly in the east-river region, and Swedish and Danish communities are found in the southeastern part of the state. Persons of German descent make up about a sixth of the population, with a substantial number reportedly using German as a first or second language. They include German-Russians, who are heavily concentrated near Yankton in the southeast and in three north central counties. The Czechs, largely Bohemians, generally live in the south central counties. The Dutch, the Finns, and the Welsh are scattered across the eastern part of the state.

Religions. Immigrant groups established about half of the state's churches. Their religious institutions not only promoted social solidarity but also played an important part in the acculturation process. In 1890, Roman Catholics were the most numerous, with Lutherans a close second; today, however, Lutherans outnumber Roman Catholics. Other leading churches are the Methodist, United Church of Christ, Presbyterian, Episcopal, and Baptist.

German groups include Mennonites and Hutterian Brethren. The Hutterites live in isolated colonies, most of them in the James River basin, where they practice communal agriculture. For their refusal to support American involvement in World War I, the Hutterites were driven into temporary exile in Canada. In 1955 they were subjected to a state law prohibiting the expansion of their colonies or the formation of additional communities. They have circumvented this law by changing their legal status from that of church to corporation, but they remain under social pressure as anomalies in the general population.

Mennonite and Hutterite sects

Roman Catholics have stood out, too, as an important and somewhat persecuted minority. To preserve their faith and avoid harassment, those of Irish, German, and Czech descent formed rural enclaves and supported ecclesiastical schools and hospitals, some of which have survived as institutions vital to the entire population. During the 1920s Catholics were the primary victims of attack by Ku Klux Klan groups in the state. Shunned by the Protestant majority in organized sports, Catholics formed a separate high-school athletic conference early in the 1920s that was not allowed to merge with the public system until forced by the legislature to do so in 1966.

The most maligned of all groups for religious behaviour have been the Indians. Only their sun dance was ever declared illegal by federal officials, but Indian worship involving the sacred pipe and peyote had been driven underground through social pressure and economic discrimination by the 1920s. Tribal members fashioned a "Reservation Christianity," blending Indian traditions with the teachings of Roman Catholic, Presbyterian, Congregational, and Episcopal missionaries. The cross, the pipe, and peyote are focal points of a spiritual environment that is distinctively Indian and remote from the majority of white South Dakotans. In this isolated ecumenical environment, the sun dance has resurfaced.

The economy. Since the fur trade era, the economy of South Dakota has relied mainly on livestock production, farming, tourism, and forest and mineral industries. In addition, the state has benefited from federal installations—notably from operations of U.S. Indian agencies, facilities built by the U.S. Army Corps of Engineers along the Missouri basin, national parks and monuments, and Ellsworth Air Force Base, a part of the Strategic Air Command.

Agriculture. Cash income from livestock and livestock products is generally several times that from crops. South Dakotans are among the leading cattle producers in the United States, and they also rank among the top wool producers in the nation. The state ranks high in the production of rye and flaxseed as well as spring wheat. It is usually among the top 10 states in corn and alfalfa-seed production. Normally, some 300,000 acres are under irrigation, mostly in the west-river region.

Industry. Tourism is second to livestock production and farming in the state's economy. Major industrial goods are food products, light machinery, and lumber and wood materials. Meat processing is important, and there are increasing numbers of "clean" industries, such as a major bank credit-card centre in Sioux Falls.

Principal mining products are gold, cement, stone, and sand and gravel. Gold is extracted from the Homestake Mine at Deadwood-Lead, and cement is produced by the state government in Rapid City. Large lignite beds containing low-grade uranium deposits await development.

The multipurpose Oahe, Big Bend, Fort Randall, and Gavins Point dams on the Missouri River have made South Dakota a major producer of hydroelectric power, most of which is transmitted to consumers outside the state. South Dakotans rely heavily on coal- or oil-fired electrical plants, which are considerably more expensive to operate than hydroelectric plants. Nevertheless, nearly all South Dakota farms are electrified through rural cooperatives.

Hydroelectric power production

Transportation. The transformation of 19th-century surface trails into modern roads began early in the 20th century. In the 1920s concrete highways were built, and all routes to centres of population in excess of 750 were graveled. During the 1930s hinterland roads were improved through the use of work-relief and conservation funds. Federal allocations initiated in 1956 led to the completion of two interstate highways (north–south and east–west) across the state. Reliable crossings over the Missouri River were restricted mainly to ferries and periods of ice cover until the 1920s, when modern bridge construction began. The number of crossings also increased with the construc-

tion of dams on the Missouri River in 1954–66. Since then, the Missouri has not been navigable for commercial purposes upstream from Sioux City, Iowa. The last ferry in South Dakota, at Running Water near Yankton, closed in the mid-1980s.

Passenger rail traffic has disappeared, but freight train transportation revived in the 1980s through the use of state funds for track improvement. South Dakotans have had air service since World War II, when federal funds were used to build airports. Airlines offer regular service to the largest cities, while private planes operate out of more than 150 public and private airstrips.

Administration and social conditions. *Government.* The state constitution was adopted in 1889 and has been amended many times. The governor, lieutenant governor, and most other high-ranking administrative officials are chosen by the electorate, for four-year terms. The legislature comprises the Senate of 35 members and the House of Representatives of 70 members. In 1951 the Legislative Research Council was created to provide continuity between the annual sessions. The congressional delegation includes one U.S. representative and two U.S. senators.

The judicial system comprises the Supreme Court, consisting of five judges, and the Circuit Court, consisting of 36 judges operating within eight judicial circuits. Supreme Court judges are appointed by the governor; Circuit Court justices are chosen by nonpartisan ballot. In January 1975 the Supreme Court consolidated the system of county and local judicial officers into a unified system of magistrates. Law-trained magistrates are appointed by Circuit and approved by Supreme Court judges; lay magistrates are installed by presiding judges in their respective districts. The machinery for law enforcement includes the state's attorney and sheriff at the county level and the office of attorney general at the state capitol. Special enforcement agencies include a state highway patrol, a force of game wardens, and the tribal police cross-deputized with county sheriffs' departments.

Court system

South Dakota has more than 3,500 units of government below the state level. These units include 66 counties, more than 300 incorporated towns and cities, and more than 1,000 organized township governments. There are also more than 100 special districts, most of them concerned with soil conservation, drainage, and irrigation.

The state constitution forbids deficit spending. The major sources of income for the state government are a sales tax, revenue from licenses and other fees, and federal aid. Most of the state's expenditures support highways, education, and welfare.

A primary election is held in June during even-numbered years, followed by general elections in November. Special elections often decide local issues. The Republican has been the dominant party since territorial times.

Each of the nine Sioux reservation groups lives under an elected tribal government, six of which are authorized by the Indian Reorganization Act of 1934. The Yankton Sioux tribe sustains a constitution adopted in 1932.

Education. The public school system is administered by local and county boards subject to policies formulated by the Department of Education. Since 1943, legislative appropriations that support public schools have supplemented the general property tax and revenues from the common school lands. School district reorganization was voluntary until 1968, when all districts were compelled to offer a 12-year curriculum. Since then, the number of country schools has diminished considerably, while the number of consolidated high school districts has grown.

Among the state-supported institutions of higher learning are the University of South Dakota in Vermillion (opened in 1882); South Dakota State University in Brookings (1884); South Dakota School of Mines and Technology in Rapid City (1885); and regional colleges at Spearfish, Aberdeen, and Madison. A system of private liberal-arts colleges has been reduced by funding shortages. Private institutions are Augustana College and Sioux Falls College in Sioux Falls, Dakota Wesleyan University in Mitchell, Mount Marty College at Yankton, and Presentation Junior College at Aberdeen. In addition, there are several vocational schools and Indian community colleges.

Colleges and universities

Health and welfare services. The Department of Health is responsible for programs dealing with communicable diseases and sanitation, as well as with the inspection and licensing of hospitals and nursing homes and the collection of vital statistics. South Dakota's health care training and facilities are unusually strong for a rural state because of a tradition in hospital service and nurses' training by the Benedictine and Presentation orders of the Roman Catholic church. In addition, Sioux Valley Hospital in Sioux Falls has evolved into a sophisticated medical complex, while the University of South Dakota Medical School, originally founded as a two-year institution, has been elevated to four-year, degree-granting status. The addition of federal facilities on Indian reservations since 1917 has kept tribal members apace with rural health care standards. Small county, municipal, and private hospitals increasingly have closed as health care has become concentrated in larger, centrally located facilities.

The welfare needs of the state are the responsibility of the Department of Social Services. More than two-thirds of funds for public assistance are received from federal grants. Institutions maintained by the state include a school for the deaf at Sioux Falls, a centre for the mentally handicapped at Redfield, and a mental hospital at Yankton.

Cultural life. The early settlers attended musical and dramatic performances by traveling artists in local opera houses and heard national lectures sponsored by lyceum bureaus. Schoolhouses were centres of community social life, and local reading circles and library associations developed into modern libraries. Libraries are maintained by numerous state and private institutions, including the library of the South Dakota State Historical Society in Pierre. A traveling library system, operated under the Free Library Commission, was established in 1913.

Many native writers have dealt with the South Dakota scene, including Hamlin Garland and Norwegian-born Ole Edvart Rölvaag, who spent his early immigrant life near Sioux Falls, the locale portrayed in his *Giants in the Earth* (1927). Charles Eastman and Elaine Goodale Eastman preserved 19th-century Indian ways in several volumes. The *South Dakota Review,* a literary quarterly published by the University of South Dakota since 1963, affords an outlet for regional as well as local talent.

In the graphic arts two native-born artists, Harvey Dunn and Oscar Howe, gained wide recognition. Harvey Dunn, reared on a pioneer Dakota homestead and one of the nation's leading illustrators, became well known for his paintings of pioneer life. Oscar Howe, a Yanktonai Sioux, has made use of the motifs and symbolism of his Indian heritage. Gutzon Borglum's stone carvings of four U.S. presidents on Mount Rushmore in the Black Hills are a major tourist attraction. Traditional Indian crafts such as beadwork have undergone a revival.

Native artists

Among the numerous museums in South Dakota, several stand out for their extraordinary collections and exhibits. The State Historical Society controls the Robinson Museum (Pierre), W.H. Over Museum (Vermillion), Agricultural Heritage Museum (Brookings), and Smith-Zimmerman State Museum (Madison). The University of South Dakota supports the Shrine to Music Museum at Vermillion, and South Dakota State University has the South Dakota Memorial Art Center in Brookings. The large South Dakota State Archaeological Research Center is located in Rapid City, and the Prehistoric Indian Village with its Boehnen Memorial Museum, is under management by the city of Mitchell.

The dramatic arts manifest their appeal at community and summer theatres and the departments of dramatic art on college and university campuses. Notable among the summer theatres is the Black Hills Playhouse, which has operated in Custer State Park since 1946. More spectacular is the Black Hills Passion Play, which has been presented during the summer at Spearfish since 1939. The musical arts are represented by a number of symphony orchestras. Ethnic festivals include Nordland Fest, Czech Days, and numerous Indian powwows.

Among recreational areas are Custer and Bear Butte state parks, Black Hills National Forest, and Wind Cave National Park.

HISTORY

French exploration

Settlement and gold rush. A lead plate, discovered at Fort Pierre in 1913, records the presence of French explorers during 1742–43. Twenty years later Spain acquired sovereignty over the region, and in 1800 it reverted to France. The territory's inclusion in the Louisiana Purchase led to U.S. ownership in 1803. Trappers and fur traders were the principal European residents of the area until the mid-1850s, when land speculators appeared and the U.S. Army built Fort Randall on the Missouri River. Permanent settlements at Vermillion and Yankton sprang up in 1859. The Dakota Territory was created in 1861, but for several years settlement was confined to the southeast between the Big Sioux and Missouri rivers.

Wars between the Indians and white immigrants went on intermittently from the Grattan Affair in Nebraska in 1854 to the massacre at Wounded Knee, near Pine Ridge, in 1890. Fairly peaceful relations during the fur trade era thus had changed to cultural separation, out of fear if not hatred, from which Indian and white South Dakotans have never fully recovered.

The search for gold in the Black Hills in the early 1870s attracted non-Indians to the western part of Dakota Territory, despite the recognition of Indian ownership by federal treaties. In 1877 the Indians were forced by Congress to accept a reduction of their reservation and to cede the Black Hills. Rapid City emerged as the main gateway city to the region. Freight and stage lines connected the mining population with the East until railroads entered west-river provinces early in the 20th century.

Statehood and homesteading. The gold rush was followed by a flood of settlers into the east-river region, swelling its population from about 80,000 to 325,000 between 1878 and 1887. This rapid expansion led to calls for division of the territory at the 46th parallel and separate statehood for the southern half. In the north and in Congress a single state was favoured. The southern section held constitutional conventions in 1883 and 1885; at the latter the state of Dakota was established. Dual statehood based on a division below the 46th parallel received congressional approval in 1889, and both North and South Dakota were admitted to the Union simultaneously.

The Great Sioux Agreement of 1889 established six reservations for Teton and Yanktonai Sioux and opened more than 9,000,000 acres to white entry. The general complexion of life in South Dakota has not changed appreciably since 1920, when the majority of residents were positioned in enclaves on farms, ranches, small urban centres, and Indian reservations. Adverse climatic and economic conditions have caused some rural-to-urban migration among non-Indians, while limited reservation resources have forced some Indians to leave the reservations. Education, health care, social services, transportation, and tourist industries all have improved. Yet population size has remained fairly static and ethnic diversity has survived.

Conservatism and progressivism. South Dakota's political history is similar to that of its neighbouring states. During the 1890s the appeal of the Populist movement led it temporarily and briefly away from the Republican Party. A four-year fusion administration, however, produced only the nation's first initiative and referendum law.

Political traditions

At the turn of the 20th century the state was influenced by the wave of progressivism. An insurgent, progressive wing dominated the Republican Party in 1906 and enacted legislation that called for direct primary elections and railroad regulation. From 1917 to 1919 the state adopted a still more radical program, which established a rural credits plan, a system of state hail insurance, a state coal mine, and a cement plant. Of these state-operated enterprises, only the cement plant at Rapid City remains in operation. Since the 1920s the voters have shown a more conservative bent. A major exception to this conservatism has been the state's support of liberal to moderate candidates for Congress, who have been successful in competing for federal benefits. A noteworthy example was liberal Democrat George McGovern, who was the only Democrat elected to Congress from South Dakota between 1936 and 1970. McGovern, however, failed to carry South Dakota in his bid for the presidency in 1972.

The occupation of the village of Wounded Knee by members of the American Indian Movement in 1973 was an expression of the desire for attention to the needs of Indian peoples. It also proclaimed the survival of tribalism, with emphasis on Indian self-determination. The subsequent siege by federal marshals lasted for more than two months, until the Indians were promised that their grievances would be negotiated.

In general, white South Dakotans have been scarcely less dependent on federal support than have the Indians, whose lives have been governed by national policies for reservation Indians. Adversities of climate and economy have shaped a way of life in South Dakota that shows few signs of losing its diverse character. (H.S.Sc./H.T.H.)

Wisconsin

One of the north central states of the United States, Wisconsin became the 30th member of the Union on May 29, 1848. It is situated between Lake Michigan to the east and the upper Mississippi River on the west. On the north it touches the western portion of Lake Superior and the Upper Peninsula of Michigan. Minnesota and Iowa lie to the west and southwest, respectively, and on the south is Illinois.

The economy of Wisconsin is diversified, with individual elements spread generally throughout the state, though its three major facets have specific regions of concentration. Its southeastern industrial belt, extending across the state line along Lake Michigan from the Chicago area to and beyond Milwaukee, the state's largest city, is the primary factor in making Wisconsin one of the largest manufacturing states in the nation. In the southern two-thirds of the state, a combination of favourable physical factors of climate, soil, and topography makes possible a dairy agriculture that allows Wisconsin to be first in the nation in the production of milk, cheese, and butter. The sparsely settled, northern evergreen–hardwood forest and lake country hosts tourist and recreational activity.

The area of Wisconsin is 56,153 square miles (145,436 square kilometres). The name Wisconsin is an anglicized version of a French rendering of an Indian name said to mean "the place where we live." Madison, the territorial capital from 1836, became the state capitol and home of the University of Wisconsin in 1848.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief and drainage.* Wisconsin comprises five physical regions. The Northern Highland is a broad upland underlain by granitic bedrock. It contains the state's highest point, Timms Hill (1,951 feet [595 metres]) in Price county. The Lake Superior Lowland is a narrow plain to which the surface of the Northern Highland drops abruptly. The upland slopes gently southward to the Central Plain, a crescent-shaped region on a sandstone bed stretching across the centre of the state. In the southwest rises the Western Upland, a region etched into ridges and valleys by streams that have cut into the limestones and sandstones. The Southeastern Ridges and Lowlands are formed by three broad, parallel limestone ridges running north–south and separated by wide and shallow lowlands. The lowest elevation in the state is along the shoreline of Lake Michigan, approximately 581 feet above sea level. Except in the southwest, there are thick glacial deposits throughout the state. The surface of the glaciated areas commonly consists of broad expanses of undulating terrain, with interspersed marshy wetlands or with lakes occurring in clusters or singly; in places the glaciated areas consist of rough, boulder-strewn moraines.

The five physical regions

Wisconsin is one of the few states in which essentially all drainage is outflowing. The principal river, the scenic, island-studded Wisconsin, 430 miles long, originates on the Michigan boundary and flows southward to near Madison, where it skirts the Baraboo Range before turning west to cross the Western Upland and enter the Mississippi near Prairie du Chien. A system of reservoirs regulates its flow. Wild rivers include the upper St. Croix, the Namekagon, the upper Wolf, the Pine-Popple, the Brule, and the Pike. The lower St. Croix is designated a scenic river. All of

these are in northern Wisconsin. This region, with a section of neighbouring Minnesota, has one of the largest concentrations of lakes in the world. Wisconsin has nearly 15,000 inland lakes of more than 20 acres (eight hectares), for a total of more than 1,500 square miles. The largest is Lake Winnebago (215 square miles) in the Fox River valley. Included in Wisconsin's boundary waters and under its jurisdiction are 7,387 square miles of Lake Michigan and 2,675 square miles of Lake Superior. Wisconsin has approximately 400 miles of shoreline along Lake Michigan and 150 miles along Lake Superior.

Unusual surface features include the Apostle Islands in Lake Superior; the rocky Door Peninsula between Lake Michigan and Green Bay; the broad gorges of the Mississippi and lower Wisconsin rivers, cut 300 to 500 feet below the general surface; ancient mountain remnants such as the Baraboo Range, Rib Mountain, and the Gogebic Range; the irregular glacial "kettle moraine" west of Milwaukee; the narrow river gorge known as the Wisconsin Dells; and the sandy beaches of Lakes Michigan and Superior.

Major soil types

Soils. The best Wisconsin soils for agricultural use are the black prairie soils and gray-brown forest soils of the Eastern Ridges and Lowlands Region and the Western Upland; these coincide rather well with the areas having the warmer and longer growing seasons. Soils less favourable for agricultural use are found in the predominantly forested regions of the Northern Highland and the Central Plain. But through the use of irrigation, drainage, and fertilization, even some of these soils have been made highly productive for special crops of vegetables and cranberries. On the steep slopes of the Western Upland strict soil erosion measures must be practiced, and in the Central Plain the sandy soil must be protected from wind erosion. In some places fences such as those used to control drifting snow are also used to control drifting sand.

Climate. Wisconsin has long, cold winters and warm but relatively short summers. Average temperatures in January range from 10° F (−12° C) in the north to 22° F (−6° C) in the southeast; in July, from 66° F (19° C) in the north to 72° F (22° C) in the southwest. The Great Lakes ameliorate both summer and winter temperatures along their margins. The length of the growing season diminishes westward and northward, from about six months in the southeast—where the best soils are found—to about three months in parts of the Northern Highland.

Annual rainfall averages about 30 inches (760 millimetres), the bulk of it occurring between May and October. Snowfall varies from about 30 inches in the south, with an 85-day cover, to approximately 50 or 60 inches in the north, with a 140-day cover near Lake Superior. Streams and lakes in the state may be frozen from December to mid-April.

Plant and animal life. Forests once covered about 85 percent of the state, with the remainder in prairies and wetlands. Most of the forests were cleared for lumber and agriculture, but by natural regrowth and reforestation nearly 45 percent of Wisconsin is again forested, most heavily in the Northern Highland and Central Plain. Trees are second-growth hardwoods—*e.g.,* maple, birch, oak, aspen, elm, basswood, and ash—and evergreens—white, red, and jack pine, hemlock, balsam fir, black spruce, white cedar, and tamarack.

White-tailed deer, foxes, cottontail rabbits, skunks, woodchucks, squirrels, chipmunks, and gophers are common in all areas. Black bears, coyotes, wolves, porcupines, beavers, otter, snowshoe hare, and eagles live primarily in the north. Pheasants are stocked in southern farming areas. Waterfowl are abundant, and migratory Canada geese by the thousands visit refuges twice annually. The numerous fish types include panfish, as well as various trout species, bass, walleye, northern pike, muskellunge, and sturgeon.

Settlement patterns. The majority of the people live in clusters of cities in the southeast, the area first reached and settled by migrants from the East. There they found soils and climate favourable for agriculture. Those who moved farther on across the state, seeking farmland, in time spread themselves fairly evenly except in the southern part of the Central Plain and the Northern Highland, where infertile or wet soils and short growing seasons discouraged settlement.

Rural agricultural settlement consists of dispersed farmsteads. When viewed from the air, field shapes and road patterns give a checkerboard appearance to the land and reflect the township and range survey system from which they are derived, except where rough surfaces, wetlands, or contour farming patterns break the regularity. In many areas the traditional red dairy barn (considerably larger than the farmhouse), along with its milk house and cylindrical silos, still dominates the farmstead. The rectangularity of the open countryside moves into the small towns in which rectangular lots, each with its dwelling, line the crisscross streets oriented in cardinal directions.

Towns of less than 1,000 people dot the entire state, but most of the larger cities are in the southeast. Milwaukee, with its satellite cities, is one of the country's major manufacturing centres, specializing in machinery and electrical equipment, and is a port. Madison, scenically located between Lakes Mendota and Monona, is the location of the main campus of the University of Wisconsin. Racine and Kenosha, on Lake Michigan south of Milwaukee, are small ports and between them produce tractors and metal goods. Green Bay is a lake port at the mouth of the Fox River and a paper centre. As the home of the Green Bay Packers, it is the sole survivor among major-league franchises of the small Midwestern cities that gave birth to professional football. Appleton is the largest city of a major paper-manufacturing complex located where the Fox River leaves Lake Winnebago. Oshkosh, on the western shore of Lake Winnebago, is a woodworking centre. La Crosse, a Mississippi River port, manufactures varied products.

Principal cities

The people. Wisconsin's population is predominantly of northern European origin. German immigrants were most numerous, followed by Poles, Scandinavians (primarily Norwegians), and British. Persons of German ancestry are widely distributed today but more concentrated toward the east and in Milwaukee. Poles are numerous but live mainly in Milwaukee and the Stevens Point area. Norwegians are more numerous toward the west and south, with Swedes more toward the north and northwest. Few persons are now of foreign birth, and the majority have become a blend of different national ancestries. Blacks and Hispanics live primarily in the southeastern lakeshore cities. More than four-fifths of Wisconsin's blacks live in Milwaukee, where they constitute nearly one-third of the population. The American Indian population is concentrated largely in the city of Milwaukee and in Menominee county.

Wisconsin's immigrants brought diverse religious affiliations to the state. Norwegians were mostly Lutheran, Germans both Lutheran and Roman Catholic, Poles Roman Catholic, and Southerners and Easterners non-Lutheran Protestant. This diversity continues, and churchgoers are divided almost equally among these religious groups.

From the Great Depression of the 1930s to the late 1960s, northern Wisconsin generally lost population, but since that time the downward trend has reversed. Much of the Western Upland and the Central Plain show increases. Although the southeast continues to increase in population, the rate of increase has slowed, and the historic flow of migration from the north and the west to the southeast appears to have stopped. A migratory trend common throughout the nation is the continued movement from urban centres to suburban areas.

Migratory trends

The economy. Wisconsin's economy has shown stability. The three major economic enterprises are manufacturing, agriculture, and the tourist-recreation industry. Although the production of durable goods, the state's major type of manufacturing, fluctuates with the economy, this fluctuation tends to be balanced by the processing of agricultural and forest raw materials (largely papermaking), in which there is less fluctuation. The major markets for Wisconsin's products, the sources of most of its energy supplies, and a high proportion of its raw materials lie outside the state, creating inward and outward flows that keep the state in close harmony with the nation's overall economy. It ranks among the top one-fourth of the states in farm income and value added by manufacture.

The state has a comprehensive economic policy. Efforts have been directed in general toward retaining and creating jobs, upgrading the labour force through vocational and technical education and job training, improving the business climate by controlling state spending and taxation, aiding small and minority businesses, adding maximum value to raw materials before shipment out of state, promoting tourism development, and assisting international trade and investment.

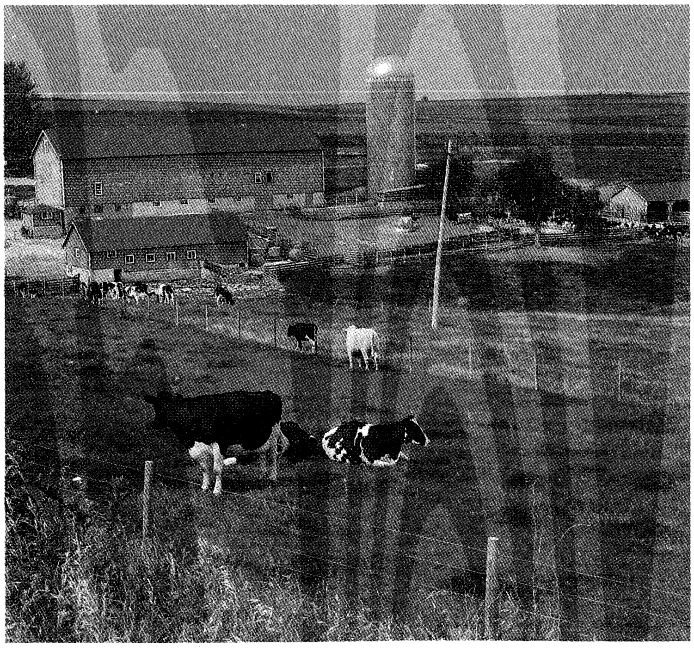
Resources. Iron is no longer mined in Wisconsin, but nonmetallic minerals such as sand and gravel, building and monumental stone, and agricultural lime are important to the state's economy. One of the 10 largest ore bodies of zinc-copper massive sulfide in North America, containing 5 percent zinc and other minerals, was discovered in northern Wisconsin in 1976. Although 75 percent of the forests are hardwoods, paper pulp is the major timber product. In commercial fishery, most of it in Lake Michigan, the lake trout and whitefish are prized. Most of the state's electrical power is generated in coal-burning plants, although a significant amount is produced in the state's three nuclear facilities. Only a minor percentage comes from hydroelectric plants.

Industry, agriculture, and tourism. The major producers of income in Wisconsin are manufacturing and processing, wholesale and retail trade, services (covering much of tourism), government, and construction. Manufacturing is concerned mainly with the production of metal goods and with the processing of agricultural and forest products. Agriculture is largely dairy farming, which requires a high degree of skill and efficiency. About four-fifths of farm incomes is derived from livestock and livestock products, three-fifths from dairy products alone. Among the features of Wisconsin highways and communities are stores specializing in cheeses of all kinds, wide varieties of wurst (sausage), and similar products, many of them manufactured locally. In addition, Wisconsin is one of the major beer-producing—and consuming—states in the nation, a reflection of its strong German heritage.

Manufacturing and agriculture

Tourism emerged in the mid-20th century and now ranks with manufacturing and agriculture as one of the major economic enterprises. It is promoted by the Division of Tourism Development through information centres in seven entry cities and Chicago and by the issuance of periodicals, brochures, films, and slides. Available for recreational use are some 117,000 acres in state parks and millions of acres in national, state, and county forests. Most of the public forests are in the north, with most of the parks in the southeast near the population concentrations, although there is a park within an hour's drive of any location.

©Robert Frerck—Odyssey Productions



Holstein cows on a dairy farm in south central Wisconsin.

Forestry and fisheries. Pulpwood production dominates the Wisconsin timber industry, accounting for more than half of the timber cut, mostly aspen and pine. Sawtimber is mostly from hardwoods, such as red oak, aspen, hard maple, and elm; the smaller softwood supply is most notably white pine. Fuelwood production in the state is also significant, having surged with the energy crisis of the late 1970s.

Commercial fishing has been restored to some degree in Wisconsin's portion of the Great Lakes after its near elimination by the sea lamprey from the 1940s to the '60s. Since that time there has been a vigorous restocking of lake trout; whitefish have also made a comeback, as have lake herring, yellow perch, and chub. Average commercial catches of these fish have come to compare favourably with those preceding the lamprey scourge. The introduction of Pacific coho, (chinook) salmon, and other game fish into Lake Michigan has met with surprising success, causing a boom in the sport fishing industry, which now surpasses commercial fishing in the economy.

Transportation. Transportation route patterns concentrate toward the southeast, with Milwaukee and Chicago as the focal points, reflecting the greater population density and industrial concentration of that area. Intercity bus service is widespread, and the larger cities have intracity service. Two interstate highways cross the state. Wisconsin ports handle more than one-fourth of the domestic freight tonnage on the Great Lakes; the largest are Duluth (Minn.)–Superior, Milwaukee, and Green Bay. Madison, Milwaukee, and several smaller cities are served by major airlines; regional and commuter lines provide service to outlying areas.

Administration and social conditions. *Government.* The first Wisconsin constitution of 1848 is still the basic law. Amendments are passed by both houses of two successive legislatures and approved by referendum. Constitutional officers, since 1970 elected for unlimited four-year terms, are the governor, lieutenant governor, attorney general, secretaries of state and treasury, and superintendent of public instruction. The legislature comprises the Senate of 33 members elected for four-year terms and the Assembly of 99 members serving two-year terms.

The Supreme Court, primarily an appellate court, consists of seven judges elected statewide for 10-year terms. The state also has a Court of Appeals and a Circuit Court of original jurisdiction. Appellate and circuit judges serve for six-year terms. There are some 200 municipal courts throughout the state.

Local government

Units of local government include counties, cities, villages, and towns. Counties, which are agents both of the state and the locality, are governed by elected boards of supervisors. Within a county all areas not part of a municipality are organized into towns, which usually coincide in boundary with the government townships. At annual town meetings qualified voters make policy decisions that are carried out by a three-member town board.

Nonpartisan office elections are held in the spring, partisan ones in the fall; both are preceded by primaries. Chosen in nonpartisan elections are judges, the state superintendent of public instruction, school-board members, county supervisors, and city, village, and town officers. All other officers are chosen on a partisan basis. Primaries are open; that is, a person may vote in the primary of any one party regardless of accustomed party affiliation or of how the voter plans to vote in the general election. Wisconsin political leaders who have gained national prominence embrace all reaches of the ideological spectrum, from the Progressive traditions of the La Follettes to the conservative senator Joseph McCarthy.

There are two systems of party organization, statutory and voluntary. Each voluntary party, which consists of dues-paying members, holds a state convention and develops a party platform, which is officially adopted at the statutory party platform convention.

Education. A kindergarten that opened in Watertown in 1856 is thought to have been the first in the United States. After the American Civil War, Milwaukee became

known as a kindergarten centre. Private academies proliferated in Wisconsin before the Free High School Law was passed in 1875. Overall responsibility for elementary and secondary education lies with the state's Department of Public Instruction, with local boards of education overseeing local districts.

Higher education

The major system of public higher education is the University of Wisconsin System, which in 1971 was combined with the Wisconsin State Universities System to create 13 four-year, degree-granting campuses, 13 two-year (Center System) campuses, and University of Wisconsin Extension. In addition there is a statewide vocational, technical, and adult education system. Among the major private degree-granting institutions are Marquette University (Milwaukee; founded 1857), Lawrence University (Appleton; 1847), and Beloit (Beloit; 1846), Carroll (Waukesha; 1846), Carthage (Kenosha; 1847), Ripon (Ripon; 1863), and St. Norbert (De Pere; 1898) colleges. The school of architecture located on the grounds of Taliesin, the home of Frank Lloyd Wright near Spring Green, is a mecca for both students and experienced architects.

Health and welfare. State and local human-service agency programs include public health and Medicaid; employment-oriented programs for the vocationally handicapped; integrated institutional and field/county service programs in corrections, mental health, and developmental disabilities; and economic assistance.

In 1982 the Department of Health and Social Services initiated the Community Options Program to establish and fund community-based noninstitutional alternatives to long-term care for the frail and disabled elderly. Persons who would otherwise be consigned to nursing homes are now afforded continued independence, dignity, and personal choice through a variety of local supportive services that enable them to remain in their own homes or other homelike settings.

In 1983 the legislature passed enabling acts for health maintenance organizations, bringing about a rapid growth statewide of these and other such groups. The competition among health care providers to lower medical costs resulted in reduced length of hospital stays, same-day surgery, fewer hospitals, and a springing up of 24-hour immediate care centres, ambulatory surgery centres, and day-care centres for the aged.

Cultural life. The settlers of Wisconsin represented extreme diversity in ethnicity, ranging from New Englanders and Southerners to immigrants of almost every European nationality. Those of like origin tended to settle in enclaves and to retain much of their transplanted cultural heritage, which gave the state a wide variety in arts and art forms. As the population became more homogenized, these unique modes of expression were retained to a considerable degree, producing both a rich diversity and a widespread appreciation of the arts in general among the populace. Among Wisconsin natives who achieved national recognition in the arts are the writers John Muir, Thornton Wilder, and Edna Ferber, the actors Alfred Lunt and Lynn Fontanne, and the architect Frank Lloyd Wright.

The cultural milieu

The University of Wisconsin has reflected and enhanced the statewide interest in the arts. It was the first university in the country to sponsor an artist in residence, the painter John Steuart Curry (1936), followed by the pianist-composer Gunnar Johansen (1939) and others. It supports the Fine Arts Quartet in Milwaukee and the Pro Arte String Quartet in Madison, groups with an esteemed international reputation. Through University of Wisconsin Extension it has over the years sponsored artists', writers', and theatrical and dance groups throughout the state. In summer it operates music clinics for high school students from throughout the country.

In 1957 University of Wisconsin Extension was instrumental also in bringing into being the Wisconsin Arts Foundation and Council, which in 1970 became an official state agency known as the Wisconsin Arts Council. In 1973 it was designated as the Wisconsin Arts Board, an agency designed to administer aid to groups and individuals in the arts.

Milwaukee is a major arts centre. The spacious 19th-century Pabst Theater has been restored. The Performing Arts Center is a multipurpose facility, where the Milwaukee Symphony Orchestra, the Florentine Opera (the state's oldest performing arts organization), the Milwaukee Ballet, and the Bel Canto Chorus perform; the Magin Art Gallery is also located there. The 100-year-old nationally recognized Milwaukee Art Museum has a collection of European and American masters and of contemporary art. The Golda Meir Library on the University of Wisconsin–Milwaukee campus contains the near-priceless map collection of the American Geographical Society.

Many of the national groups continue to hold annual festivals. The William Tell Pageant by the Swiss in New Glarus features the production of Schiller's play *Wilhelm Tell.* Norwegians hold the Syttende Mai festival in Stoughton and perform the Song of Norway at the Cave of the Mounds near Mount Horeb. Annual festivals in Milwaukee include Summerfest, German Fest, Polish Fest, and the Holiday Folk Fair, the oldest and largest multiethnic festival in the country. Numerous arts-and-crafts fairs are held in centres throughout the state. The Wisconsin State Fair is held in August in the Milwaukee suburb of West Allis.

Ethnic festivals

Among the many historical sites and museums in the state is the Circus World Museum in Baraboo, which collects and displays artifacts and other materials from circuses around the world. Many of its wagons and other paraphernalia are used in Milwaukee's annual circus parade.

The sparsely settled expanses of northern and western Wisconsin are popular areas. Among the more interesting vacation areas is the Door Peninsula, between Lake Michigan and Green Bay, with miles of rocky shoreline and sandy beaches and five state parks. It is largely forested but has cherry and apple orchards, summer cottages, small coastal villages, arts-and-crafts shops, and a summer theatre. Another is sparsely settled, heavily forested northern Wisconsin. With clear streams and hundreds of lakes, it is the epitome of the Northwoods for outdoors people. The Wisconsin Dells in the Central Plain is a mecca for tourists. Its initial attraction was the rocky canyons cut by the Wisconsin River, to which have been added many amusement park features. One of the least-heralded areas of the state but one deserving more attention is the scenic hill-and-valley country of the Western Upland, with its steep, wooded slopes, bare rock bluffs and towers, and tree-lined back roads winding through quiet pastoral scenes that include preserved homes of Cornish lead miners and a renowned summer Shakespearean theatre—an area described as like an unofficial national park.

Professional athletic teams in the state include the Green Bay Packers (football), the Milwaukee Brewers (baseball), and the Milwaukee Bucks (basketball). In addition there are several minor-league baseball teams and collegiate teams of various schools, including the Big Ten Conference teams of the University of Wisconsin–Madison.

HISTORY

Prehistoric populations. The so-called Paleo-Indians arrived in what is now Wisconsin during or after the retreat of the last continental glacier. The relationship between the prehistoric Indians and those present when Europeans arrived has not been clearly traced. Those encountered by the Europeans were the Ojibwa (Chippewa), Menominee, Winnebago, Potawatomi, Kickapoo, Sauk, Fox, Illinois, Miami, Mascouten, Huron, Ottawa, and the Santee Sioux. Four of these groups remain—the Ojibwa, Menominee, Winnebago, and Potawatomi—plus four others who migrated from the east in the 1820s—the Stockbridge (Mahican), Munsee, Brotherton, and Oneida.

Early settlement. Wisconsin was visited by a French explorer, Jean Nicolet, in 1634 and was under French control until 1763, when it was acquired by the British. It was subsequently ceded to the United States by the Treaty of Paris in 1783. It became part of the Wisconsin Territory in 1836 and in 1848 the 30th state.

French control

Unlike the French and British, who were interested primarily in the fur trade, the Americans made profound changes: clearing the land for farms; building houses, roads, and towns; and cutting the timber for lumber. They quickly dispossessed the Indians of their land by treaties

and overwhelming military defeats. They occupied the land, initially in the southwest, as lead miners and subsequently as pioneer farmers. An influx of immigrants from northern Europe began in the 1830s and grew in volume through the following decades. By 1880 the agriculturally suitable areas had been largely delineated by settlement.

Political and economic maturity. As the slavery issue intensified, a new political party, organized at Ripon in March 1854, became the Republican Party in Wisconsin. From then until 1932, Republicans dominated most state and presidential elections. After the Civil War, partly in reaction against the growing economic and political strength of the railroads and big business, there emerged a deeply rooted political unrest. It culminated around 1900 in the Progressive movement, which brought reformer Robert M. La Follette, Sr., to the forefront and resulted in the passage of bills that made the state a leader in social legislation. Among the bills was a corrupt practices act, a worker's compensation act, and the first state income tax law.

The Progressive movement

Although the Progressive movement was a strong political force in the state, it was part of the state Republican Party until 1934 when it separated to become the Wisconsin Progressive Party. In 1946 it rejoined the Republicans, but many adherents went instead to the resurgent Democratic Party. After more than 100 years of Republican near dominance, the Democratic Party then elected four out of six governors within 25 years and had a majority in the legislature much of the time. However, Republican presidential candidates have usually received greater support.

Another outgrowth of the Progressive movement was the "Wisconsin idea." Operating under the theme, "The boundaries of the university campus are the boundaries of the state," it was an effort to bring together the resources of state government, the university, and citizens' groups to solve social, political, and economic problems.

As agriculture developed through the southern two-thirds of the state, dairying with an emphasis on cheese production emerged. Since 1920 Wisconsin has ranked first in the country in dairying. In the 1870s commercial lumbering reached Wisconsin's northern forests. Timber exploitation continued for about 40 more years, leaving a devastated countryside that only since the mid-20th century has begun to recover, through the regrowth of timber and the establishment of a tourist-recreation industry.

Iron mining began in the north in the 1880s, and millions of tons of ore were shipped before the last mine of that area closed in the 1960s. Manufacturing, beginning with the small-scale processing of local raw materials, turned largely to metal fabrication and grew phenomenally in the southeast as population increased and markets expanded. The growing position of manufacturing reflects the gradual change from a rural to a predominantly urban society. The balance was reached and passed during the 1920s, and by 1980 approximately two-thirds of the population was urban.

Wisconsin, with its blend of human, social, political, and economic qualities within the context of a favourable natural setting, has proved to be viable as a state. There is an acceptable balance between agriculture and manufacturing, and forestry resources are being reestablished. The tourist-recreation industry, along with manufacturing and agriculture, has become a third major producer of income. Although state government conducts strategic planning and research on issues related to development, there is an increasing emphasis on community leadership as the catalyst for growth. (Ro.W.F.)

THE SOUTHWEST

Arizona

Arizona is a land of contradictions. Although widely reputed for its hot, low-elevation desert covered with cacti and mesquite, more than half the state lies 4,000 feet (1,200 metres) above sea level, and it possesses the largest stand of evergreen ponderosa pine trees in the world. Arizona is well known for its waterless tracts of desert, but, thanks to many large man-made lakes, it has more shoreline and more boats per capita than almost any other state in the United States. Such spectacular landforms as the Grand Canyon and Monument Valley have become international symbols of the region's ruggedness; yet Arizona's environment is so delicate that in many ways it is more threatened by pollution than are New York City or Los Angeles.

The environment

Arizona achieved statehood on Feb. 14, 1912, the last of the 48 coterminous United States to be admitted to the Union. It is located in the southwestern quadrant of the coterminous states, bordered by Utah on the north, New Mexico on the east, and the Mexican state of Sonora on the south. The Colorado River forms the boundary with California and part of Nevada on the west. Phoenix, situated in south central Arizona, is the capital. The state's name comes from *arizonac,* derived from two Papago Indian words meaning "place of the young spring."

Although the area of Arizona is the sixth largest in the Union—114,000 square miles (295,260 square kilometres)—the state has a relatively small population that is identified by demographers as urban rather than rural. Its romantic reputation as the last refuge of primitive society and old-fashioned, close-to-the-earth simplicity is at variance with the fact that, after the 1860s, the Arizona economy became industrial and technological long before it was pastoral or agrarian.

PHYSICAL AND HUMAN GEOGRAPHY

The land. Plate tectonics—the shifting of large, relatively thin segments of the Earth's crust—and stream erosion have done the most to create Arizona's spectacular topography. Specifically, the Pacific Plate and the North American Plate came into conflict and created the major tectonic forces that uplifted, wrinkled, and stretched Arizona's geologic crust, forming its mountain ranges, basins, and high plateaus. Over the centuries rivers and their tributaries have carved distinctive landforms on these surfaces.

Relief. To Arizona's two major physiographic divisions, the Colorado Plateau and the Basin and Range Province, local authorities add the Transition Zone (or Central Highlands). The northeastern two-fifths of Arizona is part of the scenic Colorado Plateau. Far less rugged than adjacent portions in Utah, these tablelands in Arizona consist mainly of plains interrupted by steplike escarpments. Although they are labeled mesas and plateaus, their ruggedness and inaccessibility have been exaggerated. The incomparable Grand Canyon of the Colorado River provides the major exception to what has proven to be an area easily traversed. Forest-clad volcanic mountains atop the plateaus provide the state's highest points: Humphreys Peak, 12,633 feet (3,851 metres), in the San Francisco Mountains, and Baldy Peak, 11,590 feet, in the White Mountains.

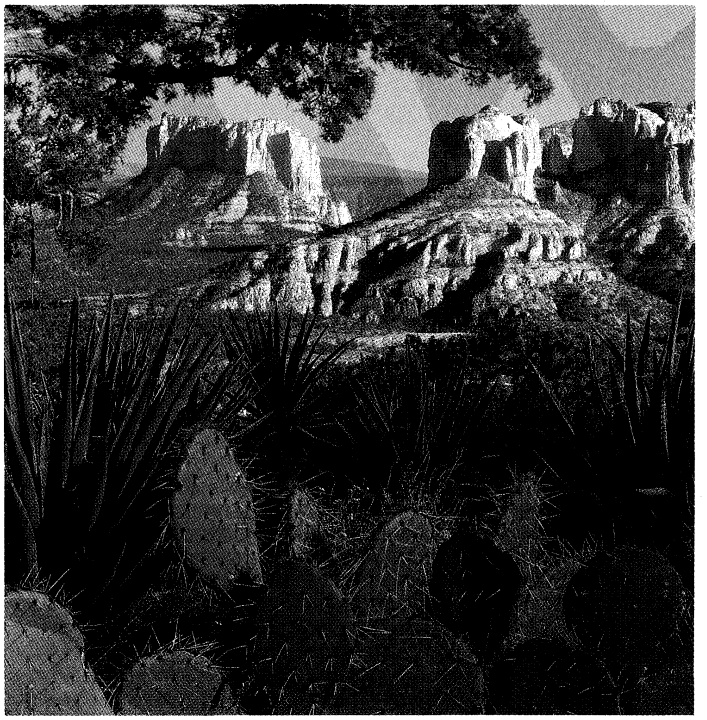
Landforms

More than 200 miles (320 kilometres) of the southern border of the Colorado Plateau is marked by a series of giant escarpments known collectively as the Mogollon Rim. West and south of the rim, a number of streams follow narrow canyons or broad valleys south through the Transition Zone and into the Basin and Range Province. The Transition Zone bordering the plateaus comprises separated plateau blocks, rugged peaks, and isolated rolling uplands so forbidding that they remained mostly unexplored until the late 19th century.

The Basin and Range region of the southern and western third of the state, containing the bulk of the population but none of the large canyons and mesas for which Arizona is famous, consists largely of broad, open-ended basins or valleys of gentle slope. Isolated mountains rise like islands in the desert plain.

Contrary to desert stereotypes, sand dunes cover less than 1 percent of the state, and stony desert surfaces are seldom visible. The younger soils of river floodplains provide the more desirable soils for agriculture.

Drainage. Virtually all of Arizona lies within the Col-



Rock formations in Oak Creek Canyon in the Coconino National Forest, Arizona.
© David Muench 1989

orado River drainage system. The Gila River, with its major feeder streams—the Salt and the Verde—is, by far, the Colorado's main Arizona tributary.

The Salt and Gila rivers

The Black, White, and Verde rivers are the primary perennial tributaries of the Salt River, which enters the Gila River southwest of Phoenix. Only during the infrequent—and occasionally devastating—flood periods does runoff water advance downstream past the numerous dams built on the Salt's system. The Gila River rises in that part of the Mogollon Rim located in western New Mexico, and it includes another and smaller Mogollon Rim tributary, the San Francisco River. Two intermittent southern Arizona streams, the Santa Cruz and San Pedro rivers, flow northward into the Gila, while two other intermittent streams, the Agua Fria and Hassayampa rivers, drain central Arizona southward into the Gila. Dams and irrigation systems, except on rare occasions, leave the Gila River dry for most of its length.

The Little Colorado River—which drains the Mogollon Rim's lee side and flows from southeast to northwest into the Grand Canyon and the Colorado River—draws and transports little water from its large watershed. Because of the rain shadow effect on the Mogollon Rim's lee side, the Little Colorado usually is no more than a trickle and often is dry. Several other small and intermittent streams, such as the Bill Williams River, drain a large but arid part of western Arizona.

Climate. About half of Arizona is semiarid, one-third is arid, and the remainder is humid. The Basin and Range region has the arid and semiarid subtropical climate that attracts most winter visitors and new residents. Receiving more than 80 percent of the possible sunshine, January days have a mean maximum temperature of 65° F (18° C) in Phoenix. Occasional light frosts occur at most locations in the Basin and Range region in winter, and light rains interrupt the exceedingly dry springs and mildly dry falls. Daily maximum readings average 105° F (41° C) in Phoenix in July, and nighttime temperatures drop to an average of 78° F (26° C).

Moisture-laden air from the Gulf of Mexico appears in July, bringing more than two months of irregular but sometimes heavy thundershowers that are locally referred to as the "summer monsoon." Phoenix and Tucson receive about one inch (25 millimetres) of rain in July. Winter rains come from the Pacific.

The Colorado Plateau has cool to cold winters and a semiarid climate. Average mile-high elevations and direct exposure to polar air masses can produce January mean high and low temperatures as divergent as the 46° F (8° C) and 19° F (−7° C) in Winslow. Year-round temperatures in Flagstaff are generally 30° F (17° C) cooler than those of Phoenix. Most of the region receives from 10 to 15 inches of rain annually, with the Mogollon Rim and White Mountains receiving the maximum for the state of 25 inches.

Because of the great amount of relief, climatic conditions within the Transition Zone vary widely over small areas. Despite its desert image, about one-sixth of Arizona is humid, much of it lying in the Transition Zone and adjacent high southern edge of the Colorado Plateau.

Plant and animal life. Considering the variety in relief and climate, it is not surprising to find similar diversity in the state's vegetation. About 10 percent of Arizona is forested, 25 percent woodland, 25 percent grassland, and 40 percent desert shrub. Elevations above 6,000 to 7,000 feet host forests of ponderosa pine, topped in the highest areas by Douglas and other firs, spruces, and aspen. From 4,500 to 7,500 feet in the northern half of the state, piñon pine and juniper predominate, while evergreen oak and chaparral grow between 4,000 and 6,000 feet in the central mountains. Plains grasses cover about one-third of the Colorado Plateau, and Sonoran or desert grass carpets the higher elevations of the basins. Mesquite trees have invaded many former grasslands in the south. Cacti grow throughout the state, with the greatest variety below 2,000 feet. Foothills in the Tucson–Phoenix area carry giant saguaro cacti of the Sonoran Desert, matched in areas of the northwest Basin and Range by dramatic stands of Joshua trees. Shrubs dominate the lowest portions of all areas: big sagebrush and saltbush in the Colorado Plateau, creosote bush in the Basin and Range.

Variety of natural life

Animal life is even more varied, with representatives of the Rocky Mountain, Great Plains, and Mexican ecological communities. Important larger animals are black bear, deer, desert bighorn sheep, antelope, and wapiti, or elk. The tropical coatimundi, a raccoonlike mammal, has spread northward into Arizona, while the javelina, or peccary (wild pig), is a favourite game animal in the south. Among the several cats, the bobcat and the mountain lion are most characteristic of Arizona. Coyotes, skunks, and porcupines abound, as do cottontails, jackrabbits, and several varieties of foxes. Lying along a major flyway, the state's southern border area is rich in birds, which attract thousands of watchers. Game birds include turkeys and a variety of quails and doves. Among native fish are the Arizona trout and the Colorado squawfish. Poisonous animals include the rattlesnake, scorpion, and Gila monster.

Settlement patterns. Despite Arizona's romantic image as a land of picturesque ghost towns and mining camps, isolated ranches, Indian reservations, and bucolic cotton and citrus farms, most of its population is concentrated in urban settlements. More than half of the state's people live in just one of the state's 15 counties—Maricopa, where Phoenix is located. Of the 15 counties, six collectively contain more than 85 percent of the state's population. Only a small number of people live on farms and ranches.

Most towns and cities have low population densities. Buildings of adobe can be seen in the older inhabited areas of southern Arizona, while Flagstaff and Prescott, northern Arizona cities settled by New Englanders in the 1860s and '70s, have Victorian homes that reflect the traditions and preferences of their first inhabitants.

Phoenix is the primary trade centre of the state. Its central location, extensive agricultural economy, and attractive vacation and retirement amenities have caused it to become one of the largest and fastest-growing urban areas in the Southwest. Tucson, while older and smaller, has acted as a doorway to Mexico and maintains well-developed commercial and medical ties with Sonora and other northern states of Mexico. Since 1970, its population growth rate has rivaled that of Phoenix.

The people. Arizona has no broadly observable cultural traditions of its own. Rather, the state has always been a cultural outpost that has reflected tastes, fashions, speech,

religious preferences, political attitudes, and life-styles that have come from such diverse localities as Boston, New York City, Washington, D.C., San Francisco, and Los Angeles.

Early inhabitants

Until the last half of the 19th century, except for very small and scattered groups of indigenous Indians, almost all of central and northern Arizona remained uninhabited. Most of the Spanish occupation of the state was tentative at best and remained confined to a few intermittently occupied missions, presidios, and ranches in the Santa Cruz valley, south of Tucson. In fact, few enduring remnants of Spanish occupation existed in the region after 1859.

Not until the 20th century did the number of Hispanic residents in Arizona soar. Today most Arizonans of Hispanic heritage are Mexicans or descendants of Mexicans who have arrived since 1900. Relations between Mexicans and Anglos (the Hispanic term for English-speaking whites) have at times been strained in Arizona, but in general the two ethnic groups have a history of cordiality that has often been absent in other border states. While some communities have Mexican barrios (ethnic quarters, often characterized by severe poverty), most Mexican-Americans live in a variety of neighbourhoods and participate fully in Arizona's business, political, and social life. Intermarriage with Anglos is common. Although Mexican food, building styles, home furnishings, clothing, social customs, and music have been incorporated into the Arizona lifestyle, the great majority of people have been affected by Mexican culture in only a superficial way. If anything, the Mexican-American population has been attracted to mainstream American culture.

In no state in the Union is the presence of the American Indian more conspicuous. Although the Indians of Arizona, from the advent of the Spanish conquistadores to the 20th century, have been subjugated, badly exploited, and abused—much like American Indians elsewhere—total annihilation and permanent displacement did not result. Arizona ranks in the top three states in the number of American Indians, who also constitute one of the highest percentages of the total population. The Indians are grouped into 15 tribes on 17 reservations that range in size from the 85-acre (34-hectare) Tonto Apache reserve to the almost 15,000,000-acre reserve (nearly three-fifths of which lies in Arizona) of the Navajo. The latter tribe,

The Navajo

numbering more than 92,000 in Arizona, presses vigorously to direct the development of its land and people, and the tribal government assumes complete responsibility in many areas of Navajo social and economic life. Among the remaining tribes the best known are the legendary Apache and the much-studied Hopi. Less well known are the Havasupai, who live at the bottom of the Grand Canyon; the Hualapai; the Yavapai; the Papago (Tohono O'odham); and the Pima tribes.

Arizona's black population constitutes less than 4 percent of the state's total. Although housing remains largely segregated, Arizona voluntarily desegregated the states's schools in the early 1940s.

The economy. Before World War II the focus of Arizona's economy was primary production—mineral extraction, lumbering, cattle raising, and growing crops. Since the late 1940s the focus has shifted toward manufacturing industry and service, the economy becoming one that better represents America's growing affluence and technology.

Resources. Metallic ores such as copper, zinc, and, to a modest degree, silver and gold traditionally have brought revenue to the state. Coal from the Black Mesa area of the Indian reservations in northeastern Arizona has become important, since coal-fired stations generate much of the electricity for the southwestern United States; the northeastern area also produces a small amount of petroleum.

Since the 1880s, northern Arizona's massive stands of ponderosa pine have made the lumber and pulp-paper industry a steady industry in the state. Rich alluvial soils, particularly in Yuma, Pinal, Pima, and Maricopa counties, have supported large and profitable agricultural operations. Since the late 19th century, people throughout the world have perceived the state's climate and landscape to be among its most valuable resources.

The natural geographic corridor created by the Colorado Uplift together with its Mogollon Rim escarpment has made possible Arizona's irrigation projects and most of the state's hydroelectric power, including that generated by the Roosevelt, Hoover, and Glen Canyon dams. Altogether, 11 dams control the Mogollon Rim's runoff, impounding and diverting the water to provide flood control and lakes for water storage. This hydrologic pattern has been a source of much 20th-century political and legal trouble for Arizona, including years of litigation with California over rights for water that falls in Arizona. The state's internal sharing of water is also a major problem because groundwater has been largely depleted, particularly around Phoenix and Tucson, and there are no new sources of surface water. Cities have found it necessary to buy water rights from distant areas.

Irrigation and hydroelectric projects

Agriculture and cattle raising. Good soil, plenty of irrigation water, and a long growing season enable Arizona to produce cotton, alfalfa, and a variety of grains, vegetables, fruits, and nuts. For many years citrus growing has remained an important and expanding part of the Arizona economy and, more recently, wine producers have enjoyed success with a number of varietal grapes. Livestock products include beef, dairy goods, and poultry and eggs. The average size of farms in Arizona is larger than that in any other state, and farmers use almost 90 percent of the state's precious water.

Industry. Between 1880 and 1950 the production of copper remained, by far, the most important industry in Arizona. Since then, however, manufacturing has grown to become the state's most important basic industry, notably in the electronics, communications, aeronautical, and aluminum industries. Although this growth has brought one of the most dynamic and affluent economies in the nation, many of Arizona's outlying counties, particularly those with large American Indian populations, remain among the poorest areas of the United States.

Tourism and retirement. Urban and industrial expansion have so polluted major areas of Arizona that it no longer serves as the refuge it once did for sick people seeking pure air. The climate, scenery, and casual life-style, however, still attract millions of visitors each year, and the state has become a popular retirement centre, particularly in the lower desert areas. Large "senior citizen" communities such as Sun City, near Phoenix, and Green Valley, near Tucson, have continued to grow.

Administration and social conditions. *Government.* The constitution of Arizona reflects the ideals of the progressive movement, which was influential at the time of the constitutional convention in 1910. It provides for maximum citizen participation through initiative and referendum on legislation and recall of all elected officials including judges. A reorganization of the state government in 1968 strengthened the power of the governor and streamlined the executive branch. The governor is elected for a four-year term. The secretary of state, who succeeds to the governorship in case of a vacancy, holds the second most highly contested elective office in the state. Other members of the executive branch include the attorney general, the state treasurer, the superintendent of public instruction, the state mine inspector, and the three-member corporation commission, which oversees public service corporations.

The legislature meets annually and comprises a 60-member House of Representatives and a 30-member Senate. The massive growth of Phoenix and Tucson, combined with reapportionment, has given urbanized Maricopa and Pima counties some three-fourths of the seats in both houses.

A constitutional amendment in 1960 restructured the judicial branch into the Supreme Court, the Court of Appeals, Superior Court, and local justice and other courts; there are no special courts. Judges of the Supreme Court and the Court of Appeals are appointed by the governor from nominees chosen by a commission. Other judges are appointed or elected.

The 15 counties, acting as agents of the state, constitute the basic units of local government. State law prescribes the town type of government for settlements of less than 3,000 people and the city form for larger communities.

Metropolitan centres have considerably more freedom in organization and operation.

Since the 1950s, Arizona has changed from a traditional one-party state dominated by the Democrats to a two-party system. Political dynamics, however, reflect the conflict between Maricopa county and the rest of the state more than any party differences. Republican strength is centred in the Phoenix area, but the party also receives support from rural, conservative "Pinto" (*i.e.,* "Spotted") Democrats. Democratic factions continue to receive support in Flagstaff, Tucson, some mining communities, and among traditionally Democratic Mexican-Americans and blacks.

Education. Public education has struggled to meet the rapid increase in students accompanying the population boom. Children must attend school between the ages of eight and 16 or until graduation from the eighth grade. Elementary, secondary, and consolidated districts operate with the assistance of county and state superintendents and an appointed state Board of Education.

Higher education

Higher education in Arizona, as in most western states, is dominated by large public universities. The Arizona Board of Regents assumes responsibility for the University of Arizona (founded 1885) in Tucson, Arizona State University (1885) in Tempe, and Northern Arizona University (1899) in Flagstaff. The State Board of Directors for Community Colleges oversees nine county community-college districts, which operate more than 30 campuses, branch centres, and skill centres. There are few private colleges in the state.

Health and welfare. The Arizona Department of Health Services, together with appointed boards, commissions, and councils, provides aid and inspection services, including a number of public health centres and hospitals. Private medical care in the metropolitan areas is excellent, but residents of rural areas and the reservations receive substandard medical services. Despite its attractiveness to the ill and the aged, Arizona has no more than its per capita share of the nation's hospitals and nursing homes. The College of Medicine at the University of Arizona and university-affiliated nursing programs work toward expanding the supply of medical personnel.

In 1981 the legislature created the controversial Arizona Health Care Cost Containment System Administration as an alternative to the federal Medicaid program. The system includes an insurance program designed to provide health care for those citizens who are indigent or who cannot otherwise afford adequate medical care. At the same time, it attempts to contain hospital and other medical costs. Arizona is the only state in the Union that does not participate in the Medicaid program, but it does receive some federal funding for its own system.

A generally modern approach to public welfare by the legislature has been coupled with financial restraints that limit the effectiveness of the programs. The Department of Economic Security works mainly through county agencies with a variety of programs for children and for the aged, blind, and disabled.

Cultural life. Although traditionally a centre for Indian folk arts and crafts, Arizona has had no circles of painters and writers comparable to those of neighbouring New Mexico. Interest in painting, crafts, drama, music, and publishing, however, has increased with population growth. Architecture and the graphic arts have been particularly influenced by Southwestern regional themes. Almost none of this, however, reflects the tastes and attitudes of the people native to the state; for the most part, the artistic and literary designs and imagery have been developed and imposed upon the region by outsiders and immigrants such as the writers Zane Grey, Oliver La Farge, Mary Austin, and Edward Abbey. A few native-born writers of the 20th century, such as Marguerite Noble and Eva Antonia Wilbur-Cruce, have contributed to a genre that emphasizes the real hardships of the region.

Indian arts and crafts

Contemporary Indian arts and crafts, executed within the traditions of the tribes, receive worldwide praise. In particular, Hopi and Navajo painters, silver and jewelry craftsmen, weavers, basketmakers, and potters produce a high volume of much-desired and expensive work.

No city dominates as an art centre, although Scottsdale, Tucson, Sedona, and Tubac have colonies of working artists; Flagstaff is home to a number of well-respected photographers. Paintings can be viewed in the Phoenix Art Museum. The Arizona State Museum at the University of Arizona in Tucson, the Heard Museum in Phoenix, and the Museum of Northern Arizona in Flagstaff feature archaeological and traditional collections of Indian arts and crafts. Symphony orchestras, theatres, ballets, and opera are well supported in Phoenix and Tucson.

More than any other art form, architecture embodies the relationship between the regional traditions of the Southwest and modern international trends. Several fine examples of Frank Lloyd Wright's work—including his home at Taliesin West—and the futuristic city of Arcosanti designed by Paolo Soleri are found in Arizona. Among the many structures in the Spanish style, the Heard Museum is outstanding, and the Nogales Public Library synthesizes the Spanish Southwestern and contemporary styles. Probably the most photographed building in all of Arizona is the San Xavier del Bac Mission (popularly called the "White Dove of the Desert") located near Tucson and completed by the Franciscans in 1797.

The state's leading book publisher, the University of Arizona Press, releases a variety of scholarly and popular titles, most with a Southwestern focus. The state's most widely known publishing venture, *Arizona Highways,* brings varied features of Arizona to a worldwide audience.

A variety of sports and recreational activities provide entertainment and leisure. Varied desert and forest terrains and many man-made lakes attract thousands of hunting and fishing enthusiasts, campers, hikers, and amateur prospectors and historians throughout the year. Arizona has more national parks and monuments than any other state; among the best known are the Grand Canyon and Petrified Forest national parks and the Chiricahua, Montezuma Castle, and Saguaro national monuments. The Arizona–Sonoran Desert Museum in Tucson has received worldwide attention as a living museum dedicated to the natural world of the Sonoran Desert. Rodeos revive the spirit of the Old West in all of the cities and on the larger Indian reservations. (M.E.H./J.W.B.)

HISTORY

Early settlement. *Prehistoric peoples.* Although the region's physical traits may appear inhospitable, even repulsive, to habitation and subsistence, Arizona contains some of the oldest records of human occupation. Relics of material culture are evidence that humans most likely have lived in Arizona for more than 25,000 years. For most of this prehistoric period, these people lived in caves and hunted animals, many species of which no longer exist. Scholars believe that the Cochise culture, made up of people living in what is now southeastern Arizona, began more than 10,000 years ago and lasted until 500 BC or later.

Hohokam, Anasazi, and other cultures

During the past 2,000 years, prehistoric societies developed within Arizona that were highly organized and advanced. Many of these Indians lived in durable masonry villages called pueblos (from the Spanish word meaning "town" or "village"). Arizona has become one of the most intensively excavated parts of the New World for archaeological research on this period. This group of prehistoric cultures, which are better known than their predecessors, includes the Hohokam, Anasazi, Mogollon, Sinagua, and Patayan. The nomadic Apache and Navajo Indians arrived in the region in the 16th century.

The Spanish period. The documented record of the European explorers and settlers of this region began in Mexico during the second quarter of the 16th century with Spaniards who wrote about the legend of El Dorado and the Seven Cities of Gold. In 1539 Fray Marcos de Niza, a Franciscan priest, entered Arizona in search of riches and Indians to convert to Christianity. Finding the Indians to be hostile, Fray Marcos returned to Mexico. One year later Don Francisco Vásquez de Coronado led a large, well-armed expedition to Arizona in an effort to claim for Spain what is known today as the American Southwest. Members of Coronado's expedition visited the

Grand Canyon and the Hopi pueblos. In 1583, Hopi Indians guided the Spanish explorer Antonio de Espejo to the site of present-day Jerome. He was disappointed to find copper and other nonprecious metal ores instead of the gold he sought. By 1675 several Franciscan missionaries had established themselves at the Hopi villages, but five years later the Hopi revolted and drove the Spaniards out. In the early 1700s Roman Catholic missionaries established churches in the upper Santa Cruz valley in southern Arizona. During this period other Hispanics also settled there. Later in the 18th century priests visited various parts of northern Arizona, including the Hopi villages, but made no serious attempt at religious conversion.

After the successful revolution that brought Mexico's independence from Spain in 1821, the new government ordered the missions in Arizona to close. Arizona was ceded to the United States as part of New Mexico in 1848. Following the Gadsden Purchase in 1853, when Mexico sold Arizona's southernmost region to the United States, only a few scattered and isolated Mexican-American ranches, all located near the Mexican border, remained.

The Gadsden Purchase

Modern Arizona. Until the Mexican War (1846–48) only a few Americans—explorers, soldiers, trappers, sheep drivers—had visited Arizona. In 1851 the U.S. Army Corps of Engineers sent several expeditions into Arizona to find a suitable route on which to build a wagon road to California. To protect travelers, miners, and other settlers from Indians, the U.S. government began to locate army posts at key sites. In 1883, workers completed the Atchison, Topeka and Santa Fe Railway across northern Arizona, thereby linking St. Louis, Mo., with California; that same year the Southern Pacific Railroad completed a line from New Orleans, La., to Los Angeles by way of Tucson and Yuma.

The copper era. Copper, Arizona's premier industry until the 1950s, was first mined in Arizona at Ajo in 1854. The Planet mines opened on the banks of the Colorado River about the same time. By 1876 the Clifton-Morenci district in eastern Arizona had two large-scale mining operations. Copper mines in Globe and Jerome, both in central Arizona, also developed rapidly. The richest copper find of all occurred in 1877 in Bisbee, in southeastern Arizona near the Mexican border. By 1880, national and international advances in electrical engineering and the availability of the necessary investment capital had created a vigorous demand for copper that had not existed earlier. Arizona began to satisfy a rapidly growing copper market; the state still mines and processes more than half of the copper produced in the United States.

Commercial agriculture. During the 1870s a few homesteaders, including a number of Mormon immigrants from Utah, attempted to develop farming economies along Arizona's few streams and rivers. Droughts, floods, and the need for heavy capitalization made it clear that for commercial farming to succeed in the state it would have to be practiced on a large scale, be highly organized, and be at a high technological level. To do this, central Arizona agricultural interests developed plans for large water-storage and flood-control systems that included expensive dams and extensive canal systems. The Salt River Project, completed in 1911, delivered water to farmers in the Phoenix area, which since has become the state's agricultural heartland. Water shortages continued to plague the state, however, and in 1963, after a long and bitter fight with California, the U.S. Supreme Court affirmed Arizona's right to 2,800,000 acre-feet of water annually from the Colorado River, as well as the entire flow of the Gila River. In 1968, after a lengthy and heated debate, the U.S. Congress authorized the Central Arizona Project, a massive system of pumps and canals to conduct water from the Colorado River to the Phoenix and Tucson areas.

Salt River and Central Arizona projects

Range cattle constituted a major source of income for Arizona from the 1860s until World War II, when large feeder lots became prominent. These lots have begun to disappear, however, and today Arizona's range cattle constitute only a small percentage of the nation's edible beef.

Population growth. After Arizona achieved statehood in 1912, it soon began to tout itself as the place of the "five Cs": copper, cattle, cotton, citrus, and climate. Health seekers from the rest of the nation discovered that the clear, clean, dry air of Arizona brought relief from various respiratory ailments, and the foundation was laid for the large number of immigrants and visitors that continues to infuse money into the state. During the 1920s "motor courts," dude ranches, and resorts grew up to accommodate increasing numbers of tourists, winter residents, and retirees. Because of its climate and its distance from the nation's coasts, Arizona became the site of several World War II flight schools and other U.S. government military bases. Following the war, a surge of immigration from other states—particularly from the Midwest—changed Phoenix into one of the fastest growing urban areas in the United States. It is a persuasive argument that the development and use of refrigeration and air conditioning did more than anything else to make Arizona attractive and habitable.

Many people think of the state as a romantic getaway fraught with images of noble Indians, heroic cowboys, and a Mexican "mañana" atmosphere. Arizona, however, has never been free from the curses of an urban and industrial society. Were it not for such technological developments as railroads, copper smelters, atomic reactors, automobiles, refrigeration, computers, and hydroelectric turbines, few people would be living in the state today.

Although many newcomers anticipate a land of personal fulfillment, the state's rates of divorce and crime, for example, are consistently among the nation's highest. Arizona is attempting to build a modern society in an arid land and, at the same time, to preserve as much as possible a beautiful natural landscape and a desired quality of life that easily could be sacrificed in the process of demographic growth and development. Unquestionably, Arizona's major problem is the limit of the state's ability to support a growing population in a way that does not bring deterioration to those characteristics that made Arizona attractive in the first place. (J.W.B.)

New Mexico

A state of the American Southwest, New Mexico is part of the "Old West" of cattle drives, cowboys, and clashes between pioneers and Apache Indians. In the vast flatness of its Great Plains and the rough, weather-scored peaks of its mountain ranges, it still retains much of its frontier flavour. Severe tensions and increasingly frequent confrontations between its Spanish-American (Hispano), Indian, and Anglo (*i.e.,* English-speaking white) populations are a continuing reminder of the bitter antagonisms that characterized its long history and were still unresolved when it became the 47th state in the Union in 1912.

Frontier flavour

The 121,593 square miles (314,925 square kilometres) of New Mexico make it the fifth largest of the U.S. states; it has only 258 square miles of water. Rectangular in shape except for a small panhandle in the southwestern corner, New Mexico is bounded on the north by Colorado, on the east by Oklahoma and Texas, on the south by Texas and the Mexican state of Chihuahua, and on the west by Arizona, which was part of the Territory of New Mexico from 1850 to 1863. At its northwestern corner it joins Arizona, Utah, and Colorado in the only four-way meeting of states in the nation.

Despite the traditionally agrarian nature of the state, augmented by successful irrigation methods, New Mexico has become urbanized. Large numbers live in Albuquerque and surrounding Bernalillo county. The capital, Santa Fe, is a much smaller city, but its founding in 1610 preceded that of Albuquerque by 96 years, and it is the oldest continuously used seat of government in North America. It was also the southwestern terminus of the Santa Fe Trail, a wagon trail that was a major commercial and migration route from Missouri to the Southwest from 1821 to 1880, when the railroad was completed.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* New Mexico has some of the flattest land in the world and also some of the most rugged mountains. Some portions have pine forests, rich meadows, and fish-laden mountain streams, while other areas are devoid

of streams, and even cacti struggle to survive. The eastern third of the state is an extension of the Great Plains that includes the Llano Estacado, or Staked Plains. The Rocky Mountains extend into the northern centre, the ranges interspersed with valleys and running in a north–south direction. The rest of New Mexico is a high plateau, but it also contains many plains and short mountain ranges.

Surface features

The average elevation ranges from 5,000 to 8,000 feet (1,500 to 2,500 metres) above sea level in the northwest to less than 4,000 feet in the southeast, with 85 percent of the state more than 4,000 feet above sea level. The highest mountain peaks, Wheeler Peak (13,161 feet [4,011 metres]) and Truchas Peak (13,102 feet), are in the Sangre de Cristo range in the north central part of the state. The lowest elevation, of 2,817 feet, lies in the southeastern corner of the state. The numerous valleys between the ranges are indispensable to agriculture and grazing. Unique volcanic formations abound as reminders of past lava flows. The caverns near Carlsbad are among the most spectacular natural rock formations in the world, while wind and water erosion have created one of the world's most extensive gypsum sand dunes at White Sands National Monument.

The Rio Grande

Drainage. Five major river systems—the Rio Grande, Pecos, Canadian, San Juan, and Gila—drain the state. The Rio Grande, which has played an influential role in New Mexico's history, virtually bisects the state from north to south. Agriculture in its floodplain has been significant since prehistoric times; European settlers initially lived exclusively in its valleys and those of its tributaries. The Pecos, east of the Rio Grande and approximately parallel to it, was also a popular route for explorers. The Canadian River, rising in the Sangre de Cristo range and flowing east across the arid plains, was a useful avenue for explorers despite its deep canyons. The San Juan and Gila rivers lie west of the Continental Divide, in the northwest and southwest, respectively. All but the Gila, which is not dammed in New Mexico, provide water for irrigation, recreation, and flood control.

Climate. New Mexico's pleasant climate has long been one of its greatest attractions, especially for those seeking a comfortable retirement or relief from respiratory and other ailments. Although New Mexico's average annual temperature is 54° F (12° C), extremes range from 116° F (47° C) to −50° F (−46° C). Variations are caused more by altitude than latitude, with temperatures falling by about 5° F (3° C) with every 1,000-foot increase in elevation. Nighttime temperatures tend to fall sharply. The average annual rainfall is 13 inches (330 millimetres), though precipitation tends to increase with elevation. About 40 inches of rain falls in the higher mountains, whereas lower areas may get no more than eight to 10 inches. Generally, precipitation is greatest in the eastern third of the state and least in the western third.

Plant and animal life. New Mexico has six vegetation zones, which are determined mainly by altitude. The Lower Sonoran Zone, in the southern sections of the Rio Grande and Pecos valleys and in the state's southwestern corner, usually occurs at altitudes below 4,500 feet. It includes nearly 20,000 square miles of New Mexico's best grazing area and irrigated farmland. The Upper Sonoran Zone, comprising about three-fourths of the state and including most of the plains, foothills, and valleys above 4,500 feet, is a region of prairie grasses, low piñon pines, and juniper shrubs. At higher altitudes, better stands reflect the more abundant rainfall. The Transition Zone, covering some 19,000 square miles, is identified chiefly by the ponderosa pine. The Canadian Zone, covering 4,000 square miles at elevations of 8,500 to 9,500 feet, contains blue spruce and Douglas fir. The Hudsonian and Arctic-Alpine zones, above 9,500 feet, are too small in area and too sparsely covered to be of great importance.

Wildlife

The diversity of natural vegetation and elevation is reflected in the state's wildlife. Mule deer, brown bear, bighorn sheep, mink, muskrat, fox, mountain lion, and bobcat live in the mountain and forest areas above 7,000 feet, while at lower elevations antelope, coyote, and jackrabbit are found. Barbary sheep from North Africa have been introduced into several mountain areas. Many species of trout are common in the mountain streams, and warm-water fish abound in lower streams. Approximately 300 species of birds can be found year-round, including various game birds. Rattlesnakes and black widow spiders are common. Despite the establishment of federal and state parks, forests, and refuges, many of New Mexico's animal species and subspecies are endangered.

Soap-tree yucca (*Yucca elata*) growing in the gypsum sand of White Sands National Monument, New Mexico.
Tom Algire

Settlement patterns. The first Spanish settlers—in the central valley of the Rio Grande and its tributaries—were awarded land grants by Spain and Mexico that have been contested throughout New Mexico's history. The Spanish-speaking inhabitants are still concentrated in the north central portion of the state. The eastern third of the state, often referred to as the "East Side," is an extension of the High Plains of West Texas and was originally settled as a cattle frontier after the Civil War. It continues to attract Protestant Anglos from Texas as ranchers, farmers, or oil-field workers, and they are often at odds with the Spanish-American Roman Catholics of Albuquerque and Santa Fe. The southwestern corner of the state, settled by Anglo miners after the coming of the railroads, also has little in common with the central area. The northwest corner received Mormon settlers from Colorado, but the greatest growth of this area resulted from oil and natural gas discoveries after World War II.

Spanish villages

Early settlers remained along streams because of the scarcity of water elsewhere. In a typical community adobe houses opened onto a plaza from which four streets ran outward, and the entire enclave was enclosed by a wall for defense. Nearby were small agricultural plots and orchards owned by individuals. Just beyond was the ejido—land for communal grazing, recreation, or firewood. Despite fear of Indian attack, ranches away from settlements often were established. At the time of the American conquest, New Mexico was a self-sufficient agrarian community, with most people residing in small villages.

After the Civil War, vast cattle ranches appeared on the East Side, their size limited only by the availability of water. The coming of the railroads in 1879 brought several waves of Anglo farmers, but frequent droughts ruined many who tried to till the soil as they had in their more humid homelands. Dry farming—tilling that uses drought-resistant crops or otherwise conserves soil moisture—saved many who remained, but today irrigated farming is the most important form of agriculture.

The people. The people of New Mexico are primarily

Anglos, Spanish-Americans, or Indians, with blacks numbering only about 2 percent of the population. The original Spanish settlers intermarried with the Indians, and their descendants are designated as Spanish-Americans or Hispanos rather than Mexican-Americans, as elsewhere in the Southwest. Spanish-Americans were in the majority until the 1940s, and people of Hispanic heritage still make up more than one-third of the population. After World War II an influx of Anglos accompanied a widespread desertion of small agricultural villages by their Spanish-speaking residents, who moved to urban centres in the state or to California. Many such villages became ghost towns.

Indian population

Indians constitute less than 10 percent of the state's population. The large Navajo reservation extends over the northwestern corner into Arizona, and nearby Gallup is known as "the Indian capital of America." There are also reservations for the Ute and for the Jicarilla and Mescalero Apache; Pueblo Indians live on nearly 1,900,000 acres (769,000 hectares) of scattered land grants. The Indians preserve many of their ancient ways, tending flocks of sheep and producing handicraft items. But dissatisfaction with their low income, inadequate housing, poor health standards, and lack of educational opportunity has led to a growing militancy and an increasing exodus from their reservations or pueblos to urban centres.

New Mexico, traditionally rural, has joined the national trend toward urbanization. Nearly one-half of the population now lives in the metropolitan areas of Albuquerque, Santa Fe, and Las Cruces, with some one-third in the Albuquerque area alone. Urbanization has involved a number of factors: the movement of Hispanos away from their rural homes, the consolidation of farms, and the increasing inclination of many farmers to abandon their isolation for the larger towns and commute to their fields and flocks.

The economy. New Mexico is a comparatively poor state. Its economy is similar to that of a developing nation in that it is largely at the mercy of forces over which it has little control. Relying heavily on the export of raw materials and on federal expenditures for programs of no certain permanence, New Mexico is subject to shifting demands from outside the state. Government spending accounts for nearly one-fourth of the state's economy.

Agriculture. In 1540 Francisco Vásquez de Coronado, the Spanish explorer of what would become New Mexico, ably described the greatest concern of the local population as the need for water, and the modern farmer is no less dependent on the water supply than was the Indian of ancient times. The Indians and Hispanos were self-sufficient farmers, growing beans, corn (maize), cotton, and squash on the alluvial plain of the Rio Grande. Efforts to raise crops with New Mexico's rainfall (a technique called dry farming) have usually failed, and most fields are irrigated. The arid land was best used for pasture, and sheep thrived until well into the 20th century. The Anglos brought cattle raising from Texas, and the sale of the cattle and calves accounts for more than one-half of the marketing receipts from agricultural products. Wholesale milk is second in value, followed by hay, chile, and wheat.

Gold and silver mining

Mining. Spanish exploration and settlement of New Mexico were prompted in part by a quest for precious metals. Even the naming of the land reflected Spanish hopes that it would be as rich in minerals as Mexico. Some mining was carried on under the Spanish and Mexican colonial governments, but only after the Civil War did gold and silver extraction become important. The mining brought many settlers and attracted capital in territorial days but never produced the riches expected. Gold and silver continue to be significant but are mainly recovered as by-products of copper smelting. Copper mining also began in the 19th century; copper is found in many parts of the state, although most of the ore reserves and production are in Grant county, in the southwestern corner.

In an area lacking wood, coal has long been an important fuel. Coal production soared with the coming of the railroads but declined after the fuel source shifted from coal to diesel oil. Production almost ceased for a time but has again become important as a fuel for thermoelectric power generation.

New Mexico produces about 85 percent of the nation's potash. Since the discovery of uranium deposits in 1950, it generally has led the nation in uranium production. Iron ore, lead, zinc, manganese, and molybdenum are also mined, but oil and natural gas account for about 50 percent of the state's mineral income. Natural gas is mainly produced in the southeastern corner and in the San Juan basin in the northwest.

The ups and downs of past mineral production have been repeated in New Mexico several times. Boom times in the mines, oil fields, or natural gas fields have been often succeeded by depression.

Industry. Manufacturing in New Mexico was originally limited to the production of consumer goods, but it has increased since World War II. Food processing, petroleum refining, smelting, and the manufacture of electronic components, communication equipment, and construction materials are leading industrial activities. Nuclear weapon and energy research is carried on at the Los Alamos National Laboratory and the Sandia National Laboratories in Albuquerque. An offshoot of this is the private manufacturing of such products as ordnance, electronic equipment, and precision instruments. Unions are not widespread and are confined largely to the mining, smelting, and petroleum industries.

Tourism and recreation. New Mexico attracts millions of visitors annually. For many years Texans, fleeing hot, humid summer weather, have been drawn to the crisp, cool mountain resorts of the state. Many people come to New Mexico in the summer to fish, camp, admire the magnificent scenery, or attend the various festivals and rodeos. Indian ceremonials are major attractions. State and national parks, monuments, and historic sites also draw countless tourists. Hunters come in the fall seeking the great variety of game birds and animals. During the winter the state's ski runs attract enthusiasts. Tourism has flourished so extensively that some suggest it has become, in terms of cash receipts, New Mexico's leading industry.

Transportation. Geographic isolation was a basic cause of New Mexico's slow economic development. In the Spanish and Mexican periods, it took about six months to travel the distance between Mexico City and Santa Fe. The Santa Fe Trail route was much shorter and faster, and American consumer goods helped prepare the way for conquest. This isolation ended when the railroads reached Albuquerque and Santa Fe in 1880. Today an extensive rail network unites the state. Highways link New Mexico's major population centres; three of these highways are part of the federal interstate system. Mountainous terrain makes road construction expensive, but secondary roads are adequate. Air transportation provides a vital link with other parts of the nation.

The Santa Fe Trail

Administration and social conditions. *Government.* New Mexico's constitution was adopted in 1911. In most instances it can be amended by a majority vote of the legislature and by a majority vote of the electorate. A public referendum on major issues is permitted, but public initiative on legislative matters is not. Nomination to office is by closed primary.

The governor has more authority than those of most states. The governor has the usual powers of pardon, reprieve, and veto, and in addition appoints most of the state boards, departments, agencies, and commissions. Consequently, the governor is the virtual master of patronage and the political organization. Like the lieutenant governor and other executive officials, the governor is elected for one four-year term. Officials are ineligible for state elective positions for four years thereafter, with the exception of the lieutenant governor, who may run for governor.

A legislature composed of 42 members of the Senate, elected to four-year terms, and 70 members of the House, elected to two-year terms, meets for a 60-day session in odd-numbered years and a 30-day session in even-numbered years. Heading the judiciary are five Supreme Court justices elected for eight years, with overlapping terms. Judges of the 13 judicial districts are elected for six years and serve ex officio as judges of juvenile courts.

Most of New Mexico's 33 counties are administered by an elected board of commissioners. Other county elective officers are assessor, clerk, sheriff, surveyor, treasurer,

and probate judge. In the territorial era citizens usually favoured Republicans, but since statehood Democrats have tended to dominate. Each of New Mexico's Indian groups elects a tribal council, which administers tribal affairs such as local government and represents the tribe in negotiations with the federal and state governments.

Education. A public school system was established in 1891, and today the educational standards in the urban centres are comparable with or superior to those in the other Western states. Many rural and small-town schools remain substandard, however. One reason for this is that many school districts cover large, sparsely populated areas, and students often live far from their schools. In addition, many country schools are located in the poorer areas of the state, where learning may be secondary to survival. Because Hispanos are dominant in such areas, education is poorer for their children. Legalized segregation for the Hispanic minorities ended in the 1950s, but de facto segregation often remains, primarily on the elementary-school level.

Colleges and universities

The state's largest institution of higher education is the University of New Mexico, in Albuquerque, established in 1889. Other state-supported institutions include New Mexico State University (1888), in University Park, and the New Mexico Institute of Mining and Technology (1889), in Socorro. Northern New Mexico Community College at El Rito, originally established in 1909 to train Hispanic teachers, has branches at Española and Santa Fe. The state universities have established branch campuses, while some cities have organized junior colleges. There are also several private colleges.

Health and welfare. The state's department of health, created in 1919, administers an extensive social service program, often in collaboration with federal agencies. Most of New Mexico's hospitals are in urban areas, and medical services in rural areas are generally inadequate. The Emergency Health Communication network links emergency vehicles, including helicopters, with medical facilities throughout the state. Other state institutions include a penitentiary, an industrial school for boys, a girls' welfare home, schools for the blind and the deaf, a development centre for mentally retarded children, and several special state-supported schools.

Cultural life. Writers and architects have been influenced by New Mexico's Indian and Hispanic heritages, which in turn have been influenced by Anglo culture. The appearance of cowboys and miners, and the conflicts of the frontier territory in the 19th century, have also been dominant cultural themes. Painters have been concerned especially with the unique landscape, since few other areas of the United States present such a variety of scenery or so many modes of life existing side by side. New Mexico's cities have attracted artists from many parts of the nation and of the world. Taos was the first to have an important art community, but it is now rivaled by Santa Fe and Albuquerque. The state institutions of higher education, through their libraries and their departments of art, music, dance, and theatre, have played a key role in the dissemination of cultural knowledge. This has been accomplished directly and through the training of public school teachers. The success of the Santa Fe Opera, which had its first season in 1957, reflects the growth of musical appreciation. The company performs in an outdoor theatre in the Sangre de Cristo Mountains near the city, presenting a repertory that has won the group worldwide acclaim.

The historical atmosphere of New Mexico and its fusion of three cultures is represented by its unique architecture. Indian pueblo buildings were modified by Spanish settlers when they built Santa Fe, and many of these original structures have been restored. The statehouse, most public buildings, and many private ones have been constructed in the modified Spanish mission style.

Indian arts

Local Indians produce beautiful pottery of high quality. Each village has its own design to identify the work of its people. Navajo blankets are famous throughout the world. Many Indians make buttons, beads, pins, rings, necklaces, earrings, and belts, mainly for sale to the growing number of tourists. The United States Indian Arts and Crafts Board has attempted to preserve the authenticity of Indian jewelry by establishing standards in handworked silver. Individual pueblos preserve traditional dances by performing at numerous fiestas, the most important being the Intertribal Indian Ceremonial, which draws thousands of visitors to Gallup every summer.

Spanish folk art has been preserved largely by the Penitentes, a religious group within the Roman Catholic church. In rural areas medieval Spanish music and art also have been preserved.

In most counties there are museums stressing history, Indian arts, or subjects of local interest. The Palace of the Governors in Santa Fe, part of the Museum of New Mexico, helps preserve archaeological artifacts, mementos, and folk arts of the past. The state archives also contain important relics. The Wheelwright Museum of Navajo art is in Santa Fe. In Taos is the Kit Carson Home and Museum, a history and art museum featuring Indian and Spanish art and 19th-century Western furniture.

HISTORY

New Mexico's first inhabitants were various groups of Indians, who farmed and hunted on the land for at least 10,000 years before white explorers appeared. The more peaceful agriculturists included the later groups, whose pueblo ruins dot the state. These groups had well-developed irrigation systems by the time the more aggressive and nomadic Navajo and Apache arrived from the north, probably in the 15th century.

Spanish and Mexican rule. Reports of the fabled seven cities of gold brought the first European explorers into New Mexico in 1540, led by the Spanish adventurer Francisco Vásquez de Coronado. The journey was fruitless, and they returned to Mexico. After several decades of desultory exploration by soldiers and friars, Juan de Oñate was given contracts for colonization in 1595 and made the first permanent white settlements during the following years, founding Santa Fe in 1610.

Colonial period

For the next century missionary work predominated, but attempts to eradicate Indian religion and culture brought on an uprising and massacre in 1680 that cleared out the Europeans for a time. By 1700, however, the Spanish had reasserted themselves, and for the next century there was considerable settlement. Albuquerque, founded in 1706, was the focal point in the south, and Santa Fe was the centre of the north.

Subsistence agriculture in the valley of the Rio Grande and its tributaries was supplemented by the raising of sheep and horses. Trade with the Comanche to the east brought consumer goods (probably from French traders) in exchange for wool, furs, and horses. The Spanish population increased rapidly, possibly to 25,000 by 1800, making New Mexico several times more populous than the colonies of Texas and California. Although there was substantial trade with Chihuahua, Mex., Spanish authorities usually neglected this important frontier province.

Although French traders from New Orleans, La., made inroads into the economy of Santa Fe, a greater threat to Spanish New Mexico came from attacks by the Apache and Comanche. Some 100 soldiers garrisoned at Santa Fe were powerless to halt the Indian forays. In 1806–07 Lieutenant Zebulon Pike led a small detachment of U.S. Army troops into New Mexican territory. After his capture and imprisonment for illegal entry into Mexico, Pike wrote a report about the Mexican southwest that brought American fur trappers and traders into the area. When New Mexico became a part of the Republic of Mexico, founded in 1821, it already had begun to trade with the United States over the Santa Fe Trail, and this trade led to still another allegiance 25 years later.

Territory and state. During the Mexican War, which began in 1846, New Mexico was taken by the Army of the West under General Stephen Kearny. All residents were granted amnesty and citizenship in return for an oath of allegiance to the United States. The Territory of New Mexico was established by Congress in 1850. During the Civil War an invading Confederate force was driven out by the Colorado Volunteers.

The Navajo tribes were quelled and in 1868 were given a large reservation; but the Apache, settled on two reser-

vations in 1880, continued their struggles until 1886. The burgeoning cattle industry was the main development of these decades, and the territory often was bloodied by battles between cattlemen and sheepmen, large landowners and homesteaders. The legendary Billy the Kid and his lawman-nemesis, Pat Garrett, were products of this struggle. The Apache leaders Geronimo, Cochise, and Victorio, though mainly active in Arizona, made forays into southwestern New Mexico. The Atchison, Topeka and Santa Fe Railway, which reached Albuquerque in 1880, brought new immigration, and farming grew rapidly with the development of new irrigation methods and resources.

Following admission as a state on Jan. 6, 1912, New Mexico retained its agricultural basis and its frontier image. In some isolated areas, stagecoaches still made connections with trains and cowboys herded cattle on the ranches, some of them vast enterprises. The Hispanic communities were little touched by the changes brought by statehood, and Indian culture was little altered by the events of the new century. There were, however, forces at work that were to materially change the state and its people.

Effects of the automobile

The automobile soon ended the isolation of even the most remote Hispanic village or Indian pueblo. Young people moved to the city, and farm products were more easily marketed by truck. The disruption brought by the motor car was continued in the 1930s by the New Deal, the many relief programs of which brought most rural New Mexicans into contact with government for the first time.

World War II acted as a catalyst to speed the changes already under way. Young Hispanic and Anglo men were conscripted into the military, and others found employment at government installations in New Mexico or in the war plants in distant states. Research facilities included that at Los Alamos, the centre of the project that created the first atomic bomb in 1945. Since then, many of the military activities have been continued. There also has been an increase in the exploitation of oil, natural gas, and other mineral resources and an expansion of agriculture through improved irrigation. (W.A.B.)

Oklahoma

In its land and its people, Oklahoma is a state of contrast and of the unexpected. The terrain varies from the rolling, timbered hills of the east, where the state borders Missouri and Arkansas, to the treeless high plains that extend into Texas and New Mexico to the west. Oklahoma's east central region is dominated by the lowlands of the Arkansas River, sweeping in from Colorado and Kansas on the north, and by the Red River, which forms nearly all of its southern border with Texas. Once basically agricultural—and the dust-bowl locale of John Steinbeck's famous novel *The Grapes of Wrath*—Oklahoma now has hundreds of lakes and a diversified economy.

The word Oklahoma is derived from two Choctaw Indian words: *okla,* "people," and *humma,* "red." During the 19th century the future state was a symbol of one of the least glorious chapters in American history, becoming known as Indian Territory, the dumping ground for Indian tribes displaced by white settlers' ever-increasing hunger for land. Since its admission in 1907 as the 46th state of the Union, however, Oklahoma has achieved an integration of its Indian citizens into modern economic and social life that probably is unmatched by any other state. There is no reservation in the usual sense for the Indian population. Though numbers of "blanket Indians" may possess no more than their bedrolls, others have risen to positions of distinction. Many share in the great wealth that oil resources have brought to the state.

Indian Territory

Oklahoma covers an area of 69,956 square miles (181,186 square kilometres). The customs of the Deep South are maintained in the habits and attitudes of southern Oklahoma—"Little Dixie"—despite the decline in cotton production. The customs of the wheat growers in the north, however, reflect their largely Kansan origins.

PHYSICAL AND HUMAN GEOGRAPHY

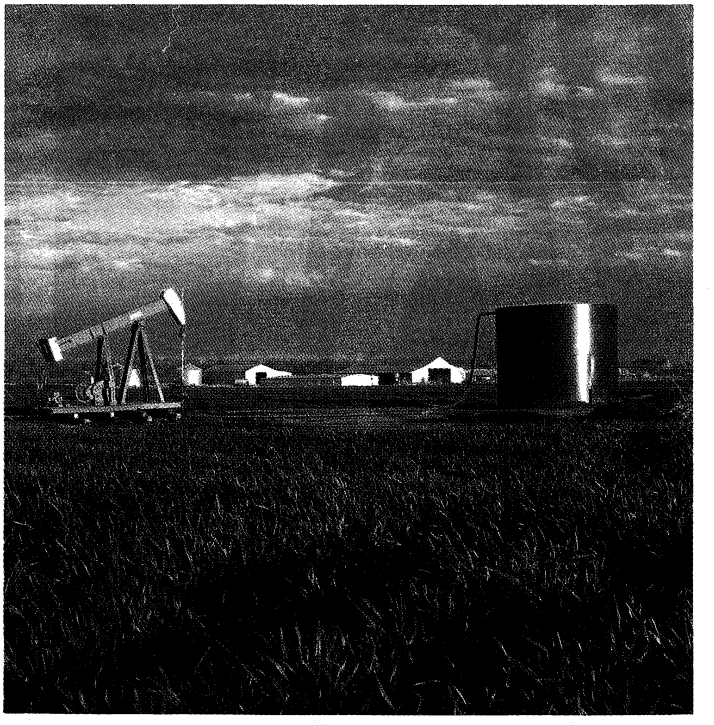
The land. Lying in a transitional zone in topography, climate, and other features, both east to west and north to south, Oklahoma comprises a jumble of environments.

Relief. Three of the nation's large physical regions extend into or across the state. The Interior Highlands is in the east; the Coastal Plain, extending through Texas to the Gulf of Mexico, is in the south; and the Interior Plains, including the Central Lowland and Great Plains, cover the remainder. Ten subregions lie within Oklahoma. Three are mountainous and in the south—the Ouachita, Arbuckle, and Wichita mountains—and are characterized by rough topography and thin soils; lumbering, grazing, some farming, and mining are their principal economic activities. The northeastern Ozark Plateau, most of which lies in Missouri and Arkansas, has rough terrain and small fields devoted primarily to growing fruits and vegetables. Once important as a lead and zinc producer, the plateau has a Cherokee heritage and beautiful rivers that make it a major recreation and tourist attraction.

The Sandstone Hills, a wide band stretching through the east central portion between the Red River and the Kansas border, is poor in agriculture and timber but important for its oil, gas, and coal deposits. The region is sprinkled with deserted or dying oil-boom towns, with Tulsa a prosperous exception. The sparsely populated Gypsum Hills section of western Oklahoma is devoted largely to grazing and farming, with large wheat acreages in the north and smaller cotton farms in the south.

The remaining four areas are flat to rolling and are agricultural. The Red River Plains, once the area of the best farmlands in the state, has been depleted by cotton. Its agriculture has been diversified by the addition of peanuts (groundnuts), melons, and vegetables grown on medium-sized plots. Its population is relatively dense, with many small towns serving as trade centres. The Prairie Plains region in the northeast is marked by grazing in its rougher portions and vegetable farms in the river valleys. Oil and gas fields are common, as is strip-mining for coal. It contains a number of middle-sized towns, some of which have small manufacturing plants. The Red Beds region is the largest, running through the middle of the state. The greatest population density is located there, as are most of the larger towns. Oil provides much of the income. Although cotton rules in the south and wheat in the north, corn (maize), watermelons, sorghum, alfalfa, vegetables, and livestock are common. The High Plains region of the northwest and the Panhandle offers a marked contrast.

© Don L. Crawford from TSW—CLICK/Chicago



Oil rig in a wheat field near Okmulgee in east central Oklahoma.

With the highest elevation and the least moisture, the eastern portion is dominated by wheat and natural gas production, and the western by grazing.

Oklahoma's drainage pattern, consisting of the Arkansas and Red rivers and their tributaries, slopes from an elevation of nearly 5,000 feet (1,500 metres) in the northwest to about 300 feet in the southeast.

Moisture and ecology

Climate. Rainfall varies from more than 45 inches (1,140 millimetres) annually in the Ouachitas to less than 20 inches in the Panhandle. Wheat and sorghum predominate in the drier western sections, peanuts thrive in the middle areas, and corn, soybeans, vegetables, and berries grow in the damper east. Irrigation has made corn a successful crop in the dry Panhandle. Virtually all of the regions have enough water for grass; hence, ranching is common.

Oklahoma has a southern humid belt merging with a colder northern continental one and humid eastern and dry western zones that cut through the state. The result is normally pleasant weather and an average annual temperature of about 60° F (15.5° C), increasing from northwest to southeast. No region is free from wind; and, as the collision point for warm and cold air masses, with sudden rises and falls in temperature, the state has heavy thunderstorms, blizzards, and tornadoes.

Plant and animal life. Oklahoma is a transitional area for plant and animal life. More than 130 species of trees are native: the eastern forests of maple, sweet gum, hickory, oak, and pine change into the cottonwood, elm, hackberry, and blackjack and post oaks of the grasslands. The arid-zone plants are chiefly mesquite, sage, and cacti. Among animals are deer, elk, antelope, rabbits, coyotes, wolves, foxes, prairie dogs, and American bison. Native fish include bass, perch, catfish, and buffalo, and virtually every bird common to the land between the Mississippi and the Rockies is found. Horned toads, lizards, many varieties of nonpoisonous snakes, and rattlesnakes and cottonmouth moccasins are native.

Settlement patterns. The outlines of roads and farms generally produce a pattern of unusual symmetry in the landscape, revealing the original survey divisions into townships, sections, and quarter sections. Small squares predominate where small-scale farming is common and very large ones where wheat and ranching prevail. As elsewhere in the nation, however, the trend has been toward urbanization. The Red Beds in the centre of the state grew most rapidly, and three of the state's four largest cities are found there, the exception being Tulsa.

Major cities

Oklahoma City, the capital, is near the centre of the state and in area is one of the larger cities areas in the nation. Banking, insurance, manufacturing, trade and transportation, state and federal installations, and educational facilities have made it the commercial and industrial heart of the state. Tinker Air Force Base, located in nearby Midwest City, is the metropolitan area's largest employer. Lawton is a centre for the Fort Sill Military Reservation, the Wichita wildlife recreational centre, and the rural population of the area. Norman, seat of the University of Oklahoma and site of the major state mental hospital, is also a bedroom city for Oklahoma City and Midwest City commuters. Tulsa, a former Creek Indian village in the Sandstone Hills region, grew slowly until the discovery of oil nearby. Refineries and facilities for manufacturing and distributing oil-field supplies have made it the headquarters for many oil companies, and it has many other financial and industrial functions. Enid is the home of Vance Air Force Base and is the marketing centre for a prosperous agricultural community.

The people. Most of Oklahoma's Indians live in the former Indian Territory in the eastern part of the state. The Plains tribes remain in western Oklahoma. Some Indians live on tribal landholdings that are informally called reservations. Most blacks in the state are descended from slaves of the Five Civilized Tribes (the Cherokee, Choctaw, Chickasaw, Creek, and Seminole), although some migrated from the South after 1865 and others came during the land runs that began in 1889. The majority live in urban centres or in the southern and eastern parts of the state, and several towns have entirely black populations.

Ethnic and religious groups

A wide variety of other racial and ethnic strains have contributed to Oklahoma's population. The original French claimants left their names and bloodlines, usually in conjunction with Indian families, and a mining boom in the 1870s brought Europeans into the Choctaw Nation. Descendants of these Italian, Slavic, Greek, Welsh, Polish, and Russian miners still live in Little Dixie. The land runs brought homesteaders from China, Japan, Mexico, England, France, and Canada, and the spread of wheat farming attracted German Mennonites and Czechs to the northwest. By the 1980s, sizable groups of Mexicans and Vietnamese had arrived. Nearly all of Oklahoma's residents, however, reflect a typically Midwestern American culture.

The state's religious sects bear out this trend toward conformity. Of the Protestant majority, the Southern Baptists and the United Methodists predominate, and the resulting conservatism has placed Oklahoma in the Bible Belt. (This fundamentalism was a primary cause for Oklahoma's retention of the prohibition of liquor sales until 1959.) Other leading denominations include the Disciples of Christ, Assemblies of God, Churches of Christ, and Episcopalians. Roman Catholics and Greek Orthodox are represented throughout the state, but Jewish congregations are limited to the cities. Most Indians have adopted some form of European religion, although the Native American church—in which use of the drug peyote is a part of the worship—is recognized by state charter. The sun and ghost dances of the western tribes reflect earlier religious practices and reactions to white settlement.

The economy. Oklahoma is one of the nation's younger states, and its economy is not as balanced as those of older, more prosperous areas of the country. There has in the past been overdependence on agriculture and petroleum, but the efforts of state and local officials to attract new forms of industry have shown some success.

Agriculture. Traditionally, agriculture has furnished an important part of Oklahoma's income, though Oklahoma's farms, which are slightly larger than the nation's average in size, have slightly less value per acre. In line with national trends, the averages are likely to remain the same, but the number of units will probably continue to decline. In commercial agriculture, livestock ranks first, followed by wheat, dairy products, peanuts, cotton, and other field crops and general produce.

Industry. Oklahoma remains somewhat of an economic satellite of the industrial North and East, furnishing food, raw materials, and fuels. Despite great efforts to diversify—for example, the manufacture of transportation equipment has become important—the state still has far to go. Only about 14 percent of its workers are in manufacturing, lower than the national average. Wholesale and retail trade employ the greatest number of people, followed by services, manufacturing, transportation and public utilities, finance, insurance and real estate, mining, and construction.

Manufacturing and labour organization

Oklahoma ranks high nationally in the value of mineral production, which includes petroleum, natural gas, natural gas liquids, coal, and stone. Commercially exploitable timber primarily consists of softwoods. The first major commercial pulp and paper plant in the state was established in 1970. Oil and gas production historically has been the major component of Oklahoma's economy. Fluctuations in oil prices—such as those in the 1980s—have tended to reduce the importance of oil and gas and cause widespread economic depression, characterized by a large number of bank failures in Oklahoma.

Transportation. Oklahoma's transportation facilities help account for its favourable record in attracting new industry. The state has well-developed networks of roads and highways and of railroads. Tulsa and Oklahoma City act as the major collection and distribution points. Several airlines provide direct flights for passengers and freight to most cities. Intricate networks of pipelines move the petroleum products, and a barge system links Tulsa to the Gulf of Mexico by way of locks and dams on the Arkansas River.

Administration and social conditions. *Government.* The general structure of the state constitution (1907) is similar

to that of other states, but Oklahomans strengthened the legislature by limiting the governor's appointive powers and ability to succeed himself, although the latter prohibition was removed in 1966, and by making the judiciary elective. Also unusual is the right of initiating legislation by popular initiative and referendum. The governor is elected for four years. The 48 senators serve staggered four-year terms, and the 101 house members serve two years. The original 75 counties, later increased by two, are represented through house districts. Constitutional provision was also made for township and city governments, though the former was abolished in 1913. This constitution, often amended, is still in force.

Oklahoma politics

Since statehood Oklahomans have favoured the Democrats. Even when the state supports Republican presidential nominees, normally that party can hope for only one or two congressional seats, and it was not until 1962 that it won the governorship. Oklahomans have a history of giving strong support to third parties; in 1914 the Socialists received 52,703 votes, and in 1968 the Southern states' rights candidate, George Wallace, received more than 20 percent of the total vote.

A major governmental change was the revision of the state's court system in 1967, which abolished justices of the peace and established selection of major judgeships according to what has become known as the Missouri Plan. Under this plan judges are nominated by a joint commission chosen by the governor and the state bar association rather than by the political parties. The state Supreme Court has exclusive appellate jurisdiction in civil cases, while the Court of Criminal Appeals has exclusive appellate jurisdiction in criminal cases; in both courts judges are elected for a term of six years. The Court of Appeals, with a judge elected from each congressional district, hears only cases assigned to it by the Supreme Court, and there is no appeal from its decisions to other state courts. Lower courts include 24 district courts, with judges elected on nonpartisan ballots. In 1966 county attorneys were replaced by district attorneys.

For financial support of its functions, Oklahoma relies basically on taxes on petroleum, natural gas, gasoline, income, and sales. Property taxes are used largely for the support of county, municipal, and school needs. A major check on spending since 1941 has been Oklahoma's "budget-balancing" amendment, by which the legislature is forbidden to appropriate more money than in the previous year plus estimated additional revenues.

Education. Supervision of public schools is conducted by elected state and county superintendents, and higher education is coordinated by the regents for higher education, appointed by the governor. The state university system is often regarded as inadequate in relation to the state's needs and resources. Exceptions are the University of Oklahoma (founded 1890) in Norman, and Oklahoma State University (1890) in Stillwater. Both have a large number of graduate departments ranked above average in achievement. Private institutions enroll only about one-fifth of the college population.

Higher education

Health and welfare. The Department of Mental Health has general charge of mental hospitals in Norman, Vinita, and Fort Supply. The Department of Human Services and the Department of Corrections administer social welfare and penal programs. In spite of a generally conservative attitude toward federal intervention in local social questions, most federal welfare programs operate in Oklahoma. In addition, more than 100 recognized agencies or groups within the state have resulted from minority initiative by Indians, blacks, and Hispanics.

Cultural life. Oklahoma is a blend of the old and new. Cowboys and Indians may be seen at numerous rodeos and at annual performances of Red Earth or at the American Indian Exposition. As host of the annual exposition and the site of Indian City U.S.A., the National Hall of Fame for Famous American Indians, and the Southern Plains Indian Museum, Anadarko is a major tourist attraction. Among the features are full-sized reproductions of the homes of various tribes, pictures and busts of their leaders, and extensive displays of their artifacts. Western historical collections are maintained by the University of Oklahoma and by the Oklahoma Historical Society in Oklahoma City. The National Cowboy Hall of Fame and Western Heritage Center, at Oklahoma City, is noted for its Western art and its exhibits of cowboy paraphernalia. The Will Rogers Memorial museum at Claremore features exhibits depicting early Oklahoma and Rogers' career as a cowboy and entertainer.

Oklahoma's best-known graphic artists are Indian, and Indian works as well as those of European masters are represented in many museums. Oil has made the state influential on the international petroleum landscape, but those that it has enriched have contributed much to the artistic scene. The Thomas Gilcrease Institute of American History and Art and the Philbrook Art Center, both in Tulsa, and the Woolaroc Museum, in Bartlesville, originally reflected individual tastes, but they have joined other art museums (notably the Oklahoma Art Center in Oklahoma City) in offering wide-ranging displays.

The performing arts

Symphony orchestras are supported in Tulsa, Lawton, Enid, and Norman. A public-school music program culminates each spring in the Tri-State Music Festival. Several ballerinas of international fame are of Oklahoman Indian descent, the most noted of whom are Yvonne Chouteau, Rosella Hightower, and the sisters Maria and Marjorie Tallchief. Theatres have been sources of entertainment since frontier days. Universities and civic groups continue to provide a wide variety of dramatic experiences and professional training. Several towns feature annual folk plays or pageants, while Tulsa boasts an opera company with a regional reputation. The Tulsa Little Theater has given more than 50 years of uninterrupted productions. The state is unusually active in literature, with numerous writers' clubs, poetry societies, and folklore groups.

Oklahoma has a wide variety of recreational opportunities and actively seeks tourists from other states. Parks range from mountainous to arid land. Among popular natural features are the Little Sahara Recreation Area and Great Salt Plains and Quartz Mountain state parks. Not the exclusively arid state that some imagine, Oklahoma has an active program of water impoundment and now boasts canoe trails, fishing tournaments, and more shoreline than the Atlantic coast. There are also many ethnic celebrations, as well as nationally known college teams in football, basketball, and wrestling.

HISTORY

Early habitation and European exploration. Of the newer states, Oklahoma is one of the oldest in terms of human occupation. The abundant game of its plains attracted hunters of the Clovis and Folsom cultures 15,000 to 10,000 years ago. Others followed, producing between AD 500 and 1300 a golden age of exquisite pottery, textiles, sculpture, and metalware. Evidence indicates a widespread system of trade and communication. This high culture apparently fell before the onslaught of primitive peoples from the western plains, and until the expedition of Francisco Vázquez de Coronado in 1541 the region's population included representatives of at least three major Indian language groups.

Coronado claimed the area for Spain, but it became little more than a highway for wide-ranging Spanish explorers. In 1714 Juchereau de Saint Denis visited Oklahoma, and other Frenchmen subsequently established a fur trade with the Indians. France and Spain struggled for control until 1763, leaving only the natives to contest Spanish authority until the return of the French flag in 1800. Three years later, through the Louisiana Purchase, Oklahoma was acquired by the United States.

American dominion. As one of the purchase's most attractive parts—because of trade opportunities—the area might well have become one of its first states; but it was, in fact, the last. Because of hostile Indians, Spanish intrigue, the mislabeling of its treeless plains as the American Desert, and the pressure for removal of the Indians from the settled East, the U.S. Congress in 1828 reserved Oklahoma for Indians and required all whites to withdraw. By 1880 more than 60 tribes had joined the local ones in Indian Territory. Some were sedentary, peaceful, agricultural, and semi-Europeanized; others were

Indian Territory

migratory and belligerent. Indian Territory consisted of five republics, or nations, with fixed boundaries, written constitutions, courts, and other governmental apparatus similar to those of the Eastern states. The major difference was that in each republic all land was held jointly, or in severalty, by an individual tribe. The first major threat to these governments came when, as former allies of the South during the American Civil War, they were placed under military rule during the Reconstruction period.

The Reconstruction treaties required, among other things, land cessions to the former slaves, the resettlement of additional outside tribes, and railroad rights-of-way. Although a scheme to colonize free blacks in Oklahoma never materialized, the weakness of the Indian governments encouraged both blacks and whites from adjoining states to trespass. Thus, the territory again became a dumping ground for Indians and an even greater cultural hodgepodge of red, white, and black people.

White settlement and statehood. Railroads seeking revenue and whites seeking property coveted the Indians' land. By 1879 organized bands, the Boomers, were moving in despite federal law. Although most were ejected, pressure continued until Congress opened some 2,000,000 acres (800,000 hectares) of western Indian Territory, bringing on the famous land run beginning at noon on April 22, 1889. Known as Oklahoma Territory, the new area came to include, through further land runs, about half of the former Indian domain. Then its settlers, many called Sooners for entering the area before official permission, sought union of the two territories in statehood. The remaining Indian Territory was dissolved by assignment of lands to the various tribes, and the Indians joined in approving the constitution of the proposed state in 1907.

The drought years of the 1930s blighted many rural areas of Oklahoma, driving thousands of farmers into long migrations in search of some form of livelihood. The economic boom of World War II, however, allowed the economy to diversify. This diversification was marked by the growth of the oil and natural gas industry, which suffered setbacks in the 1980s. The major political development of the postwar period has been the growing strength and assertiveness of Oklahoma's Indian population. (J.S.E.)

Texas

The vastness and diversity of Texas, the largest state in the Union except for Alaska, are evident in nearly all aspects of its physical character, its history, and the economic and social life of its people. As an example, January temperatures in the Rio Grande valley have been known to register well over 90° F (32° C), while at the same time, nearly halfway to Canada, blizzards were blocking highways in the Panhandle section of the state. The image of Texas was that of a raw and lawless frontier when, in 1845, it surrendered its status as an independent republic to become the 28th state of the United States. This picture has altered drastically in the 20th century and now combines great agricultural wealth, major oil and natural gas production, high national rankings in industry and finance, huge urban centres that foster a cosmopolitan cultural life, and seemingly unending stretches of high prairie and range devoted to cattle and cotton.

Image and heritage of Texas

Texas, with the fourth longest seacoast among the 48 coterminous states and a large shipping industry to match, occupies the south central segment of the nation. Its 266,807 square miles (691,030 square kilometres) make it larger than any nation in Europe with the exception of the Soviet Union. Water delineates many of its borders: the Rio Grande carves a shallow channel that separates Texas from Mexico on the southwest; the Gulf of Mexico laps its crescent-shaped coast on the southeast; the Sabine River forms most of the eastern boundary with Louisiana, where by land it is bounded by Arkansas as well; and the wriggling course of the Red River on the north makes up two-thirds of the state's boundary with Oklahoma. The Panhandle section juts northward, forming a counterpart in the western part of Oklahoma, and New Mexico lies to the west.

The name of the state derives from the Spanish name (from an Indian word meaning "allies" or "friends") for an Indian group. Austin is the state capital. Texas is commonly divided into East and West, although the dividing line between the two is ambiguous. Generally, though, East Texas has a wet climate and is characterized by cotton and ties to the Old South, while West Texas is dry and is characterized by cattle ranching and an affinity to the West.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Far from being merely wide, arid plains that are filled with cattle and cowboys, Texas comprises a series of gigantic steps, from the fertile and densely populated Coastal Plains in the southeast to the high plains and mountains in the west and northwest.

Stretching inland from the Gulf Coast, the Coastal Plains range from sea level to about 1,000 feet (305 metres). These flat, low prairies extend inland to form a fertile crescent that is well adapted to farming and cattle raising. Near the coast much land is marshy, almost swamp, except where drained by man-made devices.

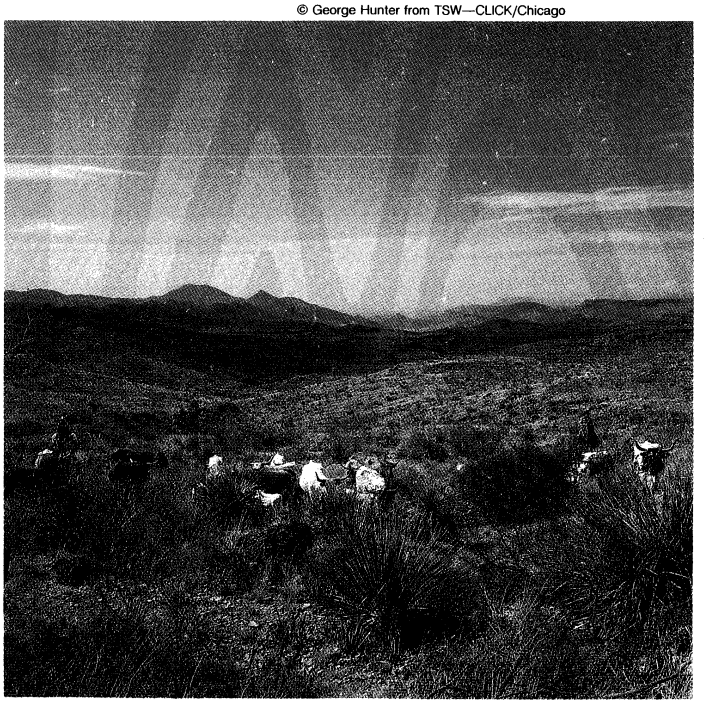
Agriculture on the coast

The western anchor of the Coastal Plains is the Rio Grande valley, where a heavy investment in citrus farming occasionally is damaged by disastrous freezes. It now features diversified farming, with vegetables and citrus. The low coastal lands between Port Lavaca and Port Arthur are ideal for rice cultivation. Inland from Houston the flatlands provide grazing for fine-breed cattle. Forests of pine and cypress grow extensively from Beaumont to the Red River and spill into Louisiana and Arkansas, making lumbering and paper mills important industries.

Access to water transportation, reservoirs of natural gas and oil, and availability of raw materials have made the coastal area the centre of industry in Texas. It is also the most densely populated part of the state. Houston, Texas' largest city, is a focal point, while Fort Worth, Dallas, Waco, Austin, and San Antonio form a line at the inner edges of the Coastal Plains. Corpus Christi, Galveston, and the Beaumont–Port Arthur–Orange complex augment Houston's port, one of the nation's largest.

The Coastal Plains, encompassing about two-fifths of the state's land area, break abruptly at the Balcones Escarpment, where in the distant geologic past the surface of the Earth cracked and slipped. Northwest of the fault, the land rears up into the Texas Hill Country and then into the tablelands of the Edwards Plateau to the south and the North Central Plains to the north. These last two regions are extensions of, respectively, the Great Plains and the

© George Hunter from TSW—CLICK/Chicago



Texas longhorn cattle grazing at the head of Fresno Canyon, Trans-Pecos region, Texas.

Central Lowlands. The entire region varies from about 750 to 2,500 feet above sea level. Farming and livestock raising constitute the basic economy. The Hill Country mixes orchard crops with ranching, small industries, summer camps, and tourism.

At the western edge of the North Central Plains lies the Cap Rock Escarpment, an outcropping of rock that stretches to the north and south for about 200 miles (320 kilometres). Protruding above the plains like a huge barricade, it is starkly visible in some places in cliffs that rise from 200 to almost 1,000 feet. Beyond that escarpment lies the third big step of Texas: the High Plains country and, to the south, the Trans-Pecos region.

From the High Plains country of West Texas emerged many of the legends of Texas weather and of the Texas cowboy. In this region lies the flat, dry area known as the Llano Estacado (Staked Plains). According to legend, as the Coronado expedition moved westward, it laid down stakes to serve as guides for the return trip, and even Indian tribes hesitated to venture across these lands. On these plains, sandstorms can obstruct vision in midday, filter tiny bullets of sand into the best-built homes, and scour the paint from exposed automobiles. Many wide, flat riverbeds remain dry most of the year, but they can become sluiceways for flash floods. Through the northern portal, beyond Amarillo, "blue northers" sweep out of the Rocky Mountains with a frenzy of freezing wind, ice, and snow. In the 19th century such famous ranches as the XIT, the Spur, the JA, and the Matador spread their cattle over these ranges.

The Llano Estacado

The North Plains subdivision, centred around Amarillo, depends on grain farming, ranching, oil, and small industries. The South Plains subdivision, with Lubbock as the principal city, has large underground water reservoirs that allow large-scale irrigated cotton farming.

The state's most rugged terrain lies to the west of the Pecos River. Trailing down from the Rockies, the Guadalupe Mountains lead into mountains of the Big Bend country, a name that is derived from a loop of the Rio Grande. The highest peak in Texas is Guadalupe Peak, which rises 8,749 feet (2,667 metres) above sea level. Big Bend National Park preserves the native ruggedness of the region.

Soils. There is immense variation in the types of Texas soil. The Piney Woods region of East Texas has a gray and tan topsoil that covers the red subsoil usually within a foot or two of the surface. The soil along the upper and middle Texas coast is black clay or loam, with lighter-coloured sandy soil on the coastal islands, bars, and spits. The soil of the southern Texas coast and inland to the Rio Grande is sandy, like that of East Texas, but is less eroded and leached.

The Blackland Prairie, a belt of fertile black clay to the west of the Piney Woods, extends southwesterly from the Red River to San Antonio. The soil of the Grand Prairie region, just to the west of the Blackland Prairie, is more rocky and resistant to erosion.

The Cross Timbers, a forest region with light-coloured, slightly acid, sandy loam soil, stretches across the prairies of northern Texas, enclosing part of the Grand Prairie. Red sandy and dark clay soils are found in the Llano Basin, in the centre of the state. The Edwards Plateau has thin, stony soil with a limestone bedrock.

Most of the soils of the western North Central Plains are red or tan-coloured and sandy, but some black clay is found in the region. The High Plains, just to the west, has dark brown to reddish clay loams, sandy loams, and sands. In the Trans-Pecos region are found reddish brown sandy soil in the mountains and grayish brown to reddish brown clay soil in the basins.

The rich fertility of the soils first attracted settlers to Texas. Much of the soil was lost through wasteful practices in the 19th and early 20th centuries, but since the 1930s efforts by federal and state governments have done much to promote soil conservation in the state.

Climate. Generalizations about the weather in Texas are almost meaningless. The Gulf Coast area around Houston has an average annual temperature of about 70° F (21° C) and rainfall of 45 inches (1,145 millimetres), whereas the Panhandle averages about 60° F (16° C) and less than 20 inches of rain. The driest region is the Trans-Pecos country, and the wettest is the southeast. Southern areas have freezing weather only rarely. In Brownsville, the southernmost city, no measurable snow has fallen in the 20th century, but the northwestern corner averages 23 inches annually.

Plant and animal life. A great variety of vegetation is found in Texas due to differences in the amount of rainfall and type of soil. Native longleaf, shortleaf, and loblolly pine provide most of the commercial timber in East Texas. A belt of post oak grows just west of the pine woods, as do blackjack oak, elm, pecan, and walnut. Marsh and salt grasses are found along the Texas coast, with bluestem and tall grasses growing a little farther inland.

The region south of San Antonio was originally brush country with mesquite, small live and post oak, prickly pear cactus, bluestem, buffalo grass, and bunchgrass. Irrigation in the 20th century has resulted in extensive vegetable and fruit production along the lower Rio Grande.

Bluestem, grama, Indian grass, switch grass, and buffalo grass grow in the prairies and plains regions of West Texas. Oak, pecan, elm, Osage orange, and mesquite are native trees found in the prairies and the Cross Timbers region. Cedar, mesquite, yucca, cactus, and some islands of cypress make up the vegetation of the Edwards Plateau.

Desert plants provide much of the vegetation of the Trans-Pecos region. Piñon pine, ponderosa pine, spruce, cedar, and oak grow in the higher mountains of the region.

Some 550 species of birds (nearly three-fourths of all species found in the United States) have been identified in Texas. Among the more exotic are the once nearly extinct whooping cranes that winter in the protected Aransas National Wildlife Refuge, near Corpus Christi, and the rare ivory-billed woodpeckers found in the Big Thicket National Preserve of East Texas.

Texas' bird species

Many of the domestic animals that are important in the economy of the state—cattle, horses, sheep, goats, and hogs—were introduced by the Spanish, but more than 100 mammals are native to Texas. Some, such as the bison, black bear, mountain lion, pronghorn, and red wolf, almost disappeared in the late 19th century and have been saved from extinction only through the efforts of conservationists. Some 100 species of snakes, including the poisonous copperhead, cottonmouth, rattlesnake, and Texas coral snake, are native to the state. The alligator is found in the lower reaches of all the major rivers and bayous.

The people. Texas has long been a huge reservoir for diverse streams of races and cultures. For the bands of prehistoric hunters, for the waves of Indian tribes, for the Spanish and Mexicans pushing northward, for the Anglo-Americans from the North and East, and for colonizers who came to Texas directly from Europe, there was more than enough room for settlement and for the opportunity to influence the institutions of the state. Some churches still conduct services in Swedish, Czech, or Spanish; throughout the south and west, Spanish remains the family language of many people.

Diverse ethnic stock

During the 19th century there were streams of migration into Texas. Between 1821 and 1836 an estimated 38,000 settlers, on promises of 4,000 acres (1,620 hectares) per family for small fees, trekked from the United States into the territory. In the 30 years before the Civil War, shiploads of Germans, Poles, Czechs, Swedes, Norwegians, and Irish arrived. Church-oriented, they took with them the Roman Catholic and Protestant faiths.

During the post-Civil War years numerous families moved from devastated Southern plantations to farms and ranches of the Southwest. From the north central states came communities of farming families with Swedish, Polish, and Irish backgrounds seeking relief from the tight economy. Others came from Europe, including Belgians, Danes, Italians, and Greeks, to become city dwellers, craftsmen, and keepers of small shops. Since the 1960s the Asian population has grown rapidly.

About 50,000 Indians live in Texas, but this figure fails to account for the many families who have some Indian ancestry. Most of the present-day Indians are city dwellers, but three tribes remain cohesive units. The Alabama-Coushatta Indians occupy one of the two reservations in

the state, in East Texas. The Tigua live on a reservation in El Paso, and the Kickapoo live near Eagle Pass.

Hispanic population

More than one-fifth of all Texans are of Hispanic descent, and their number continues to grow. Many of the communities along the U.S. side of the southwestern border are almost totally Hispanic, and larger cities such as Brownsville, Laredo, Corpus Christi, El Paso, and San Antonio carry the mark of Spain and Mexico in their architecture, names, and language. With the urbanization of the state and the decrease in the demand for agricultural workers, large Hispanic populations have converged on the major metropolitan centres that lie farther from the border.

The Civil War brought freedom for thousands of black slaves within the state. In the 20th century the black population clustered in the central parts of the larger cities, and more than 40 percent of blacks are now concentrated in the urban areas of Dallas and Houston.

The economy. Cotton, cattle, and oil—all based on land resources—dominated the successive stages in Texas' economic development until the mid-20th century, and they have continued to undergird the state's basic wealth. Retailing and wholesaling, banking and insurance, and construction have been among the activities reflecting the general affluence, urbanization, and diversification of the state's economy. Despite the growth of manufacturing and other industries, however, the Texas economy has remained heavily dependent on oil and gas. Fluctuations in oil prices—such as those occurring in the 1980s—have directly affected the economy, with the result that the classic Texas "boom and bust" economic cycle has continued.

Numerous national corporate headquarters have moved to Texas, and petroleum companies have explored for new sources of energy to continue their leadership in providing fuel for the nation. The Lyndon B. Johnson Space Center, an installation of the National Aeronautics and Space Administration (NASA), is located in Houston and is among many federal air installations in Texas. In addition, tourism has become a major business, and Dallas has attracted attention as a fashion centre—generally a low-wage industry, however, for most employees. Texas also has become preeminent in its oceanographic investigations into uses of the continental shelf and in the areas of medicine and surgery.

Agriculture. The fertile lands of East Texas attracted cotton farmers before the Civil War, and, following that struggle, cotton became the state's major crop. As mechanized farming developed, cotton production shifted to the High Plains country of West Texas, where irrigation and fertilizer fostered bountiful crops and maintained Texas' national leadership in cotton production. Occasional crop failures due to drought led to crop diversification. In total value of farm crops, Texas consistently has ranked in the top five among the states since the mid-20th century and has been a leading producer of grain sorghums, peanuts (groundnuts), and rice.

Nearly all of the mohair that is produced in the United States comes from the Angora goats of Texas. The state leads all others in the raising of beef cattle and sheep. The vast cattle empires of the 19th century have tended to shift to coastal areas during the 20th century, reversing the path of cotton.

Mining. New uses for oil were being developed when in 1901 the Spindletop gusher blew in near Beaumont. Texas leads all other states in oil and natural gas production. It also ranks first in oil-refining capacity. Oil deposits have been found under more than two-thirds of the state's area, though many finds seem too small for commercial development. Texas also leads all other states in the production of sulfur, crude gypsum, and magnesium.

Mineral wealth

Industry. Manufacturing began originally with the processing of local raw materials: cotton gins and cottonseed mills, meat-packing plants, flour mills, and fruit- and vegetable-canning plants. The production of electric and electronic equipment is now the largest manufacturing employer, followed closely by nonelectric machinery.

Oil refining is a major processor of raw materials, and oil-field equipment is manufactured. The Gulf Coast area is the centre for the petrochemical industrial complexes. A large percentage of the basic petrochemicals that are produced in the United States come from plants located from Beaumont to Corpus Christi. Development in aerospace, military, and health industries has led to some economic diversification in this area.

Reflecting a maturing economy, manufacturing has moved toward the fabrication of finished consumer products. The growth of the electronics industry has been outstanding, and other finished goods manufactured in quantity include air conditioners, furniture, boats, household appliances, machinery, leather goods, and clothing.

Transportation. The vastness of Texas and its contrasts in terrain originally posed great difficulties for transportation yet greatly stimulated its development. The desire to develop inland areas was one factor leading to the establishment of Austin as the capital. In 1852 the legislature granted public lands to railroads for each mile of track constructed, and in 1883 it authorized a county road tax for farm-to-market dirt roads. By 1900 railroads crisscrossed the state, and dirt roads straggled between most communities.

Today Texas leads the nation in road and rail mileage. It has a well-developed federal and state highway system, although concentrated in the more heavily populated east; this system is supplemented by an extensive network of roads maintained by counties and cities. As in other states, the actual mileage of mainline railroads has diminished, and passenger transportation has been discontinued over most lines. Operating freight revenues, however, have increased tremendously since the mid-20th century.

Texas was a pioneer in the development of the airplane. In or near San Antonio were located the first army flying schools, established at Fort Sam Houston in 1910; Kelly Field, which became a training camp for pilots in 1917; and Randolph Field, which by 1931 was serving as "the West Point of the Air." The need for air power in World War II brought air training to more than 40 military bases in Texas. Dallas–Fort Worth, Houston, and San Antonio are focal points for civilian air transportation. The Dallas–Fort Worth Regional Airport is the nation's largest in terms of land area and one of the busiest. The Houston Intercontinental Airport, one of two airports serving Houston, encompasses 7,300 acres and provides a computer-run train to carry passengers between terminals.

Water transportation

The discovery of oil and gas necessitated cheaper routes of water transportation to markets in the East and North. Federal aid permitted harbour improvements at Galveston, Sabine Pass (opening water routes to Port Arthur and Beaumont), Aransas Pass, and Corpus Christi. The opening of the 44-mile Houston Ship Channel has made Houston a major international port. In the 1930s an intercoastal canal was completed from New Orleans to Sabine Pass and from Galveston to Corpus Christi, and in 1946 the Gulf Intracoastal Canal was opened from Brownsville to Florida. Continuous dredging operations have opened lanes of ocean commerce to many smaller ports. Galveston, oldest among the major ports, is the headquarters for extensive commercial fishing enterprises.

Administration and social conditions. *Government.* The constitution of 1876 outlines the prevailing structure of Texas government. The governor, elected for a four-year term, may initiate legislation, call special legislative sessions, veto bills, and appoint boards and commissions. The governor's power is limited, however, because numerous officials and executive boards are elected rather than appointed. The bicameral legislature comprises the Senate of 31 members who serve for four years and the House of Representatives, with 150 members elected for two years. The top court for civil matters is the Supreme Court, with a chief justice and eight associate justices elected for six-year terms. The highest court for criminal matters is the Court of Criminal Appeals, with nine justices elected for six-year terms. There are 14 courts of civil appeal and more than 370 state district courts, with judges elected for four-year terms. Lower courts comprise county courts, justice of the peace courts, and municipal courts.

Texas comprises 254 counties; the largest of them, Brewster, with some 6,200 square miles, is roughly equal to the combined areas of Connecticut and Rhode Island. Within

constitutional limitations the legislature may create new counties. Each county is administered by a commissioners' court, which is an administrative rather than a trial body. Cities with a population of more than 5,000 may adopt their own home-rule charters.

The more critical problems of local government inevitably involve finances. The traditional source of local financing has been the property tax, but the movement of workers to suburbs and a hodgepodge of governmental agencies with taxing power has complicated the matter.

The Democratic Party has dominated elections since the Reconstruction period, pitting the many splinters of the party against one another in primaries that usually determine the eventual winners of state offices. Within the party the political philosophies of candidates have ranged from extreme liberalism to extreme conservatism. Certain contrary trends have emerged, however. The influx of new businesses and industries has brought many Republicans into Texas. The conviction that a two-party system would allow elections based on issues rather than personalities led some Democrats into the Republican camp, and strong support of civil rights by the national Democratic Party caused widespread defections by conservative Democrats. In addition, many influential Texas liberals have begun subtly to support Republicans, whom they consider more liberal than the old-time Democratic leaders.

Emergence of two-party political life

Politics also has become more organized. Both Hispanics and blacks have used the power of the ballot to elect city officials, to influence state decisions, and to move upward in the power structure.

Education. Efforts to meet, understand, and solve educational problems arising from the social, economic, and other changes since World War II have brought mixed results. Some schools have library self-study in buildings designed to accommodate innovation and to place the educational emphasis upon individual growth rather than grade classification. Conversely, there are also formal classrooms in which Hispanic children are forbidden to speak in their family language and where traditional instructional methods are used. In general, local school systems, despite minimum standards established by the state, vary greatly in accordance with local financial resources, prevailing adult educational levels, and demands for equal education for all segments of the population.

Public lands have been used to support education from the years of the Republic of Texas. The Texas Congress in 1839 set aside lands in each county to support schools. The state constitution of 1876 affirmed the endowment of 52,000,000 acres for public schools and another 2,000,000 acres for a state university and agricultural college.

The University of Texas system enrolls more than 100,000 students, nearly half of them on the main campus in Austin. The state has some 140 colleges and universities, including public senior institutions and junior colleges. The University of Texas and Texas A&M University have outstanding graduate and research programs. Rice University, a private institution in Houston, long has been recognized for its high academic standards.

Texas has a number of private or church-supported colleges and universities. Baylor University, in Waco, founded in 1845, is the only remaining university of the five established during the republic. Southern Methodist University, in Dallas, is a private institution affiliated with the United Methodist church.

Health and welfare. Programs for the mentally ill and mentally retarded have been slow to develop in Texas, but public concern has made itself felt in the legislature. There are several mental health hospitals within the state. For aged patients there are geriatric centres, and there is also a neuropsychiatric institute. Increasing attention has been given to outpatient clinic services. Several state centres care for the mentally retarded.

In medical education, research, and preventive medicine the state ranks among the nation's leaders. The University of Texas Health Science Center at Dallas and the Texas Medical Center, in Houston, are typical of these excellent programs.

Cultural life. The sense of the past has traditionally been strong in born-and-bred Texans. The emphasis long was based upon the heroics of living in a frontier land, upon individuals and their deeds, but an increasing appreciation of the diverse cultures that have enriched the life of Texas has helped to preserve and strengthen those customs. Throughout the state regional historical associations quietly search out and help to restore striking examples of 19th-century homes. San Antonio has re-created the early 18th-century Mexican-Spanish flavour in both restoration and in public shopping and walking areas in the heart of the city. Fredericksburg, with its German background, preserves many 19th-century customs and continues to cling to German as a family tongue. Even a metropolitan city like Houston has found space adjacent to its downtown area for restored historic homes. Laredo vividly dramatizes the marriage of Mexican and Anglo cultures with an annual Washington's Birthday parade and fete. Particularly evident are influences from Mexican culture, from the deep-rooted impact of the cattle country, and from the newer yet vibrant life fashioned by oil booms, wildcatting, refineries, and pipelines.

Appreciation of cultural diversity

Art, music, and literature occupy significant places in the lives of many communities in Texas. The Amon Carter Museum of Western Art in Fort Worth houses many paintings and bronzes of Western artists and maintains a microfilm collection of Western newspapers published before 1900. With the Fort Worth Art Museum, the William Edrington Scott Theatre, the Kimbell Art Museum, and the Fort Worth Children's Theater, the Amon Carter Museum provides a cultural centre for study and appreciation of the arts.

Houston's Civic Center nestles in 150 acres amid the city's tall downtown buildings. It serves as the home for the Houston Symphony Orchestra and the Houston Grand Opera. The world-famous Alley Theater is located nearby. In Dallas the Margo Jones Theatre and the Dallas Theater Center provide outlets for cultural and educational groups. The Dallas Symphony Orchestra is among the better known classical ensembles in the nation. Cultural interests, however, are not restricted to large metropolitan areas. Odessa, for example, supports the unique Presidential Museum, showing extensive memorabilia of the U.S. presidents, as well as an accurate replica of London's Globe Theatre, in which a summer program of Shakespearean and other Elizabethan plays is produced. Colleges and universities in the state are active in all areas of the arts.

Water has added new dimensions to popular recreation. In 1913 there were only eight major lakes or reservoirs in Texas, but today there are more than 180, many built to store water against periodic droughts. Several national parks and forests and more than 100 state parks dot the state, many of them providing fishing, swimming, camping, and picnicking facilities. Sports fishing has developed into a major recreation along the Gulf Coast.

Sports and amusements

Other entertainment centres include Six Flags over Texas, a westernized amusement park between Dallas and Fort Worth, and, near the Dallas suburb of Mesquite, a reservation into which motorists can drive and watch apparently free-roaming elephants, giraffes, and other African wildlife. Sea World of Texas is an aquatic theme park in San Antonio. In Houston the well-publicized Astrodome has become a centre of professional sports, rodeos, bullfights (with no killing), circuses, and other spectaculars.

Book publishing, though not a big business within the state, has gained a strong foothold. The University of Texas Press and the Texas A&M University Press have gained national acclaim through their scholarly and historical works, and the Southern Methodist University Press likewise has established discriminating standards. Several commercial publishing companies concentrate on books and monographs related to the history of the Southwest.

University libraries, art galleries, and special collections contain remarkable treasures. The Armstrong Browning Library at Baylor University houses more than 10,000 books and manuscripts by and about Robert Browning and Elizabeth Barrett Browning. At the University of Texas in Austin, the Lyndon Baines Johnson Presidential Library, operated as a branch of the Library of Congress, houses millions of documents on public affairs since the mid-1930s related to Johnson's public career. Also at the

university are a Latin-American collection, the Michener Collection of Art, and other special collections.

HISTORY

The forerunners of the West Texas Indians lived in camps that were made perhaps as much as 37,000 years ago. Possessing only crude spears and flint-pointed darts, these hunters survived primarily on wild game. In the more fertile areas of East Texas, some of the tribes established permanent villages and well-managed farms and evolved political and religious systems. Forming a loose federation in order to preserve peace and to provide for mutual protection, they came to be known as the Caddo confederacies. By 1528, when the first Europeans entered the interior of Texas, the area was sparsely settled, but the culture and habitation of the Indians exerted measurable influence on the later history of the region.

Settlement. By the 1730s the Spanish had sent more than 30 expeditions into Texas. San Antonio, which by 1718 housed a military post and a mission, had become the administrative centre. Missions, with military support, were established in Nacogdoches in East Texas, Goliad in the south, and near El Paso in the far west. The French also explored Texas. The explorations of Robert Cavelier, Lord de La Salle, and his colony at Matagorda Bay were the bases of French claims to East Texas.

Anglo-American colonization gained impetus when the United States purchased the Louisiana Territory from France in 1803 and claimed title to lands as far west as the Rio Grande. By 1819, however, the United States had accepted the Sabine River as the western boundary of the Louisiana Territory. Moses Austin secured permission from the Spanish government to colonize 300 families on a grant of 200,000 acres. When Mexico became an independent country in 1821, his son, Stephen F. Austin, received Mexican approval of the grant. He led his first band of settlers to the area along the lower Brazos and Colorado rivers. By 1832 Austin's several colonies had about 8,000 inhabitants. Other colonies brought the territory's Anglo-American population to about 20,000.

Revolution and the republic. Unrest throughout Mexico, including Texas, resulted in a coup by Antonio López de Santa Anna, who assumed the presidency in 1833. Texans, hopeful for relief from restrictive governmental measures, supported Santa Anna. Austin expected a friendly hearing about these grievances but instead was imprisoned in Mexico City for encouraging insurrection. He was freed in 1835 and returned home to find that skirmishes had already developed between the colonists and Mexican troops and that Santa Anna was preparing to send reinforcements. Texans formed a provisional government in 1835, and in 1836 issued a declaration of independence at Washington-on-the-Brazos. David G. Burnet was chosen ad interim president of the new Republic of Texas, Sam Houston was appointed its military commander, and Austin became commissioner to the United States with the mission of securing strategic aid and enlisting volunteers.

The famous siege of the Alamo in San Antonio lasted from February 23 to March 6, 1836. The strategic objective of the stand was to delay Mexican forces and thereby permit military organization of the Texas settlers. As the battle climaxed with a massive attack over the walls, the defenders (about 183) were all killed. Among the dead were the famous frontiersmen Jim Bowie and Davy Crockett. On April 21 Sam Houston led a surprise attack on the Mexican troops at the San Jacinto River, where he succeeded in capturing Santa Anna and in securing victory for the Texans.

The Texan revolution was not simply a fight between the Anglo-American settlers and Mexican troops; it was a revolution of the people who were living in Texas against what many of them regarded as tyrannical rule from a distant source. Many of the leaders in the revolution and many of the armed settlers who took part were Mexicans.

The Republic of Texas

The Republic of Texas was officially established with Sam Houston as president and Stephen Austin as secretary of state. Cities were named in their honour: Houston was the capital until 1839, when Austin was approved as the permanent capital.

The republic had a difficult 10-year life. Financing proved critical, and efforts to secure loans from foreign countries were unsuccessful. Protection against raids from Mexico and occasional attacks by Indians required a mobile armed force. During the republic a squad of armed men, the famous Texas Rangers, was maintained to ride long distances quickly to repel or punish raiding forces.

Annexation and statehood. As early as 1836, Texans had voted for annexation by the United States, but the proposition was rejected by the Jackson and Van Buren administrations. Great Britain favoured continued independence for Texas in order to block further westward expansion of the United States, but this attitude only helped to swing Americans toward annexation. Annexation was approved by the Texas and the U.S. congresses in 1845, and the transfer of authority from the republic to the state of Texas took place in 1846. One unique feature of the annexation agreements was a provision permitting Texas to retain title to its public lands.

The U.S. annexation of Texas and dispute over the area between the Rio Grande and the Nueces River brought about the Mexican War. Troops led by Winfield Scott and Zachary Taylor invaded Mexico, and Scott captured Mexico City on Sept. 14, 1847. In the Treaty of Guadalupe Hildago, signed on Feb. 2, 1848, Mexico gave up its claim to Texas and also ceded an area now in the states of New Mexico, Utah, Nevada, Arizona, California, and western Colorado. Texas claimed most of this additional area but later relinquished it in the Compromise of 1850.

Civil War and aftermath

Texas seceded from the Union on Jan. 28, 1861. The American Civil War brought disruption to the state. Governor Sam Houston strongly opposed secession, and, after refusing to take the oath of allegiance to the Confederacy, he was removed from office. During the war Texans had to defend themselves from Indian attacks, from Mexican encroachments, and from Federal gunboats and invading soldiers. Federal forces ultimately gained control of the lower Gulf Coast but were unable to move far inland.

The modern period. During the last three decades of the 19th century there were rapid developments in the population and economy of Texas. The state was readmitted to the Union under a new constitution in 1869. By 1875 the Comanche had been forced onto a reservation in present-day Oklahoma. Under waves of immigration, towns were established, farming spread throughout the central areas of the state, and the cattle industry began to thrive on the plains of West Texas. Railroad building and increased shipping fashioned new links with the rest of the world. Manufacturing, encouraged by the Civil War years, continued to grow. By 1900 the population had grown to more than 3,000,000.

The Lucas gusher that blew in at Spindletop (Beaumont) in 1901 opened a new economic era for the state. Oil companies were formed, oilmen began to search for and find new deposits in the state, and refining and marketing activities provided new jobs and incomes for Texas. Like the rest of the nation, Texas suffered throughout the Great Depression of the 1930s but later benefited from the tremendous industrial expansion that took place during World War II.

Economic and population growth continued in the postwar era. Oil refining, chemicals, and petrochemicals continued to dominate, but electronics, aerospace components, and other high-technology items became increasingly important in the last quarter of the 20th century. The population of Texas increased fourfold between 1900 and 1980, and by 1980 one-third of all Texans were either black or Hispanic.

Since the mid-20th century, Texans have played an increasingly important role in national politics. Sam Rayburn, of Bonham, served as speaker of the U.S. House of Representatives for 17 years, a tenure longer than that of any other person. Lyndon B. Johnson, who earlier had served as a Texas congressman, was majority leader of the U.S. Senate in the late 1950s, vice president of the United States from 1961 to 1963, and president from 1963 to 1969. In 1988 George Bush of Houston, who served as vice president of the United States from 1981 to 1989, was elected president. (DeW.C.R./R.A.Wo.)

THE MOUNTAIN REGION

Colorado

Colorado is classified as one of the Mountain states of the United States, although only about one-half of its approximately 104,091 square miles (269,596 square kilometres) lies in the Rocky Mountains. It borders Wyoming and Nebraska on the north, Nebraska and Kansas on the east, Oklahoma and New Mexico on the south, and Utah on the west. Colorado was admitted to the Union on Aug. 1, 1876, as the 38th state. The capital is Denver.

Colorado's history is written in the names of its cities, towns, mountain ranges, and passes. Indian and Spanish names alternate with those of frontier Americans, and many ghost towns are reminders of the thousands of prospectors and homesteaders who streamed into the territory in the mid-19th century to pursue dreams of gold and grain bonanzas. The vast cattle ranges and agricultural acreage fed by huge irrigation projects are characteristic of modern Colorado, as are the diversified industries and the educational and research facilities in its urban centres.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Colorado's natural landscape ranges from the flat, grass-covered High Plains of the Great Plains, through the rolling, hilly Colorado Piedmont paralleling the Rocky Mountain front, to the high and numerous mountain ranges and plateaus that make up the southern Rocky Mountains and the Colorado Plateau. Within these areas the state rises from about 3,500 feet (1,100 metres) in the east to more than 14,000 feet in the Rockies.

The High Plains

Lack of water is the dominant characteristic of Colorado's eastern plains region. The Arkansas and South Platte are the major rivers, but both rise in the mountains to the west. Many other rivers are dry during much of the year, and the land is flat. Underlain by layered rocks, sandstones, shales, and limestones covered by a short grass vegetation, the natural environment is inhabited by prairie dogs, jackrabbits, coyotes, rattlesnakes, antelope, and such birds as the meadowlark and lark bunting. The climate, flatness, and layered rocks have produced fertile soils that lack only moisture. Nearly all of the plains are covered by brown soils, which support a strong mat of buffalo and grama grass, a valued resource for cattle grazing.

About 50 miles (80 kilometres) wide and 275 miles long, the Colorado Piedmont is a picturesque, hilly to mountainous landscape sandwiched between the plains and the mountains. It encompasses all of the state's large urban complexes, the major transport arteries, most of the industry, most of the major colleges and universities, and four-fifths of the people. The layered rocks have been uptilted and dissected into prominent stream divides and deep valleys by the major rivers and numerous smaller streams that debouch onto the piedmont from the mountains. Terrain, ground cover, and climatic conditions provide suitable habitats for rabbits, waterfowl, pheasants, coyotes, deer, raccoon, and, on the arid foothills and unirrigated uplands, rattlesnakes. Many species of birds prevail, of which the meadowlark, crow, dove, and western magpie are most numerous. The climate and land of the Colorado Piedmont attract tourists, homeseekers, and farmers. The major cities and the wealthy farm areas lie where the streams have broadened the valleys. Among the attractive features of the landscape is the high, grotesque, and multicoloured agglomeration of sandstones northwest of Colorado Springs known as the Garden of the Gods. In the foothills southwest of Denver is one of the world's largest and most beautiful outdoor amphitheatres, Red Rocks Park. Since 1880, more than 400 reservoirs have been built in the piedmont to store water for irrigation. These sites are meccas for water sports, hunting, and house building.

The mountainous west

The western half of Colorado includes the huge mountain upthrust, comprising much of the southern Rocky Mountains and the Colorado Plateau, where mesas and mountain ranges alternate with broad, intervening valleys and deep, narrow canyons. With its copious amount of precipitation, this mountain land provides water for six states and Mexico. The drainage pattern from the Rockies is oriented by the mountains themselves, which form the Continental Divide, the main watershed boundary of the continent.

The mountainous portion of Colorado comprises a great number of individual mountain ranges. In the north and northwest the Front, Medicine Bow, Park, and Rabbit Ears ranges are major uplifts, and Rocky Mountain National Park (established 1915) is a major attraction. The western and southwestern extremity of the state comprises the tilted and acutely uplifted layered rock of the Colorado Plateau. The Grand Mesa and the White River Plateau, both above 10,000 feet, are major attractions. The region contains several national monuments and parks, most of them primarily scenic, while Mesa Verde National Park, designated by UNESCO as a world heritage site, preserves the remnants of cliff-dwelling Indian settlements.

The San Juan Mountains, a large, heavily ice-dissected volcanic plateau above 13,000 feet, rise in the southwest. The Sangre de Cristo Range is a linear range in the south central region of the state. At its western base are some of the largest sand dunes in the interior of the North American continent, an area of 60 square miles set aside in 1932 as the Great Sand Dunes National Monument.

The Sawatch, Colorado's highest range and the central core of the Colorado Rockies, consists of Mount Elbert—at 14,433 feet (4,399 metres) the highest point in the state—and many other elevations above 14,000 feet. The Colorado Rockies contain a significant share of the U.S. public domain in the form of 11 national forests, which total about 14,000,000 acres (5,670,000 hectares) of land. (Another national forest lies partly in the state.) There are 53 peaks more than 14,000 feet in elevation and 831 peaks between 11,000 and 14,000 feet.

Climate. Colorado may be divided into three climatic regions that largely reflect differences in elevation and proximity to the major mountain ranges: the eastern plains, the piedmont, and the mountains and high plateaus of the west. Summer temperatures on the plains average 75° F (24° C) for July and August, with extremes above 100° F (38° C). In the summer daily minimum and maximum temperatures may vary as much as 40° to 50° F (22° to 28° C), although the general variation is about 25° F (14° C). Winters are dry, cold, windy, and generally harsh. The plains form a playground for the wind, and, though snowfall is light, the winter blizzard becomes a dread element for both people and animals. Average January nighttime low temperatures range from 10° to 30° F (−12° to −1° C), with daily highs from 36° to 50° F (−2° to 10° C). Annual precipitation is erratic, ranging from 15 to 20 inches (380 to 510 millimetres). Approximately 70 percent of it falls during summer, and hail is frequent.

Climate of the plains

There is less precipitation annually in the Colorado Piedmont than on the plains, though about 70 percent of it falls in summer also, mostly in thunderstorms. July temperatures in Denver, the state capital, average 73° F (23° C); and January temperatures, 30° F (−1° C). Short hot and cold spells of above 90° F (32° C) and below 10° F (−12° C) are not uncommon. The chinook wind—a dry, descending winter airstream from the high mountains that is warmed by compression as it descends—often raises temperatures 30° to 40° F (17° to 22° C) in less than an hour, melts the snow cover, and can produce violent winds that have been recorded in excess of 100 miles per hour.

The rugged topography of the mountains and plateaus of western Colorado produces a complex pattern of local climates. Elevation, amount of exposure to the direct rays of the sun, and the orientation of mountain ranges and valleys to the general air circulation are major factors determining the climate of a particular location. Typically, wide variations in climate occur within short distances. July temperatures average about 60° F (16° C) at many mountain locations, while the lower plateaus and valley

bottoms average some 20° F (11° C) higher. Winter temperatures often are more extreme with elevation. Leadville, at 10,000 feet, has an average January temperature of 17° F (−8° C), but temperatures can drop to −50° F (−46° C) at higher elevations. Arid conditions prevail over much of the Colorado Plateau, while at elevations above 5,500 feet (1,700 metres) rainfall and snow are sufficient to support thick forests. Precipitation generally increases rapidly with elevation, with amounts ranging from about 20 to 50 inches. Snow may fall during any month in the mountains, and total annual accumulations are heavy, regularly reaching 300 inches at some stations.

Ecological zones

Plant and animal life. There are four broad ecological zones from the plains to the high mountain peaks. The plains are dominated by short-grass prairie, or steppe. In the foothills zone, from 5,500 to 7,000 feet, oak, mountain mahogany, juniper, and piñon pine are the dominant vegetation. Higher zones, from 7,000 to about 11,500 feet, feature a coniferous forest in which the ponderosa pine, Douglas fir, and blue and Engelmann spruce are dominant, interspersed with aspen and other deciduous, broad-leaved species. The alpine tundra zone, above 11,500 feet, has sparse vegetation, mainly mosses, lichens, and sedges.

Most animal species have no permanent habitat in the Colorado Rockies. They move to high elevations where food and cover are plentiful during summer and return to the warmer lower elevations during winter. Deer, elk, and mountain sheep are the most common game animals. Among the furbearers, the coyote, wildcat, badger, marten, muskrat, and beaver are prevalent.

Settlement patterns. An overall consideration of Colorado's population is most meaningful in a regional context.

The demography of Colorado's eastern plains is much affected by the region's rigorous physical geography: its dryness, bareness, wind, and capricious precipitation. The seven plains counties constitute nearly one-sixth of Colorado's land area but have a dwindling population, the density of which rarely exceeds five persons per square mile (two per square kilometre). The towns of the plains, all located on highways and railroads, serve vast rural hinterlands where livestock raising is important and where wheat and sorghum are major products. Limon, Burlington, Cheyenne Wells, and Yuma are the largest towns.

Ready availability of water, a climate conducive to outdoor work and recreation, and proximity to the mountain front are mainly responsible for the large population growth of the Colorado Piedmont. The 22 counties occupy a third of the land area, and about 80 percent of the state's people live in the metropolitan areas of Denver–Boulder, Colorado Springs, Pueblo, Fort Collins, and Greeley.

Terrain, isolation, severe winters, and separation from the piedmont counties by the Continental Divide are major limiting factors in population density and distribution in Colorado's mountain and plateau counties. The 34 counties occupy half of the state's land area, but some have fewer than two people per square mile. Unlike the plains, however, the population is increasing. The rural population is settled mainly in restricted mountain valleys, where ranching and irrigation farming support the family unit. Ski resorts enhance the local economies of such areas as Aspen and Vail, while energy production is important to the economy of Grand Junction.

The people. The first Colorado territorial census, in 1860, showed a population of 34,277, 86 percent of it rural. This pattern continued until 1910, when half of the nearly 800,000 inhabitants were urban. After 1950 the urban percentage rose sharply, reaching more than 80 percent of the population in the late 20th century. The number of blacks and American Indians is small, but a significant percentage of Coloradans are of predominantly Mexican descent. American Indians are concentrated in two areas of the state: metropolitan Denver, with migrants from tribes throughout the United States, and the southwestern corner of Colorado, where two Ute reservations are all that remain of that tribe's once-vast presence in the area. As in most of the nation, minority groups are hampered by inadequacies in education, housing, and economic opportunity. The conditions of seasonal migratory labour have been of increasing concern at all levels of government.

The economy. Location, soil, minerals, water, space for expansion, and physical beauty are positive resources in Colorado's growth. Among the Rocky Mountain states Colorado accounts for nearly two-fifths of the population but about one-half of all manufacturing employment.

Agriculture. From the outset, agriculture has been basic to Colorado's economy. Colorado was the first state to abrogate the riparian doctrine of water use, based on English common law, which gave prior water rights to owners of adjoining lands. It evolved instead a totally new concept for use of water resources based on the rights of the larger public, which has been adopted and adapted by most of the 17 Western states. The state ranks high among the U.S. states in the amount of land under irrigation. Corn (maize), wheat, and hay are the major crops.

Cattle production

Colorado is a major cattle producer and also raises large numbers of hogs and sheep. Weld, Morgan, Larimer, and Boulder counties are the national centre for the production of cattle fattened in feedlots rather than on the open range. The piedmont landscape offers the spectacular sight of acres of fat cattle feeding on corn and alfalfa (lucerne) near Greeley. There is much corporate farming, and generally it is highly mechanized.

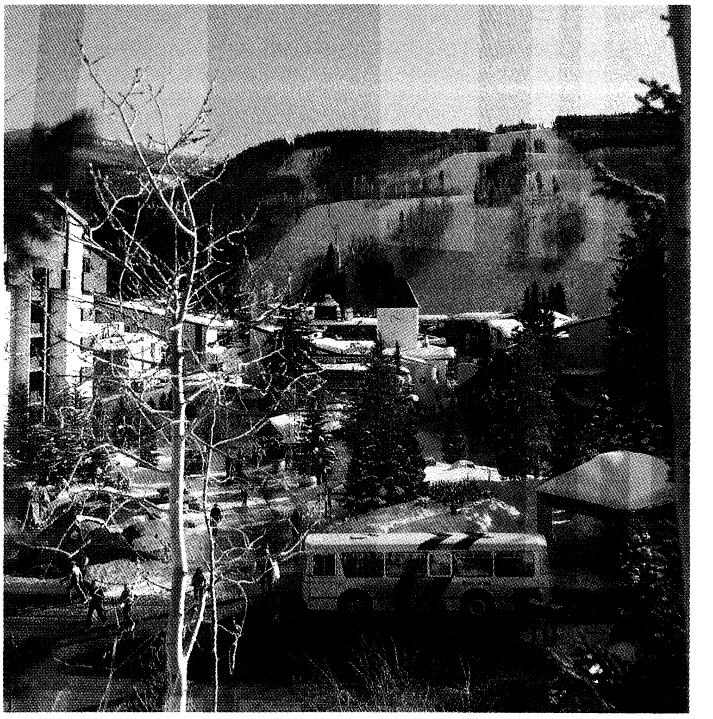
Mining. Although not the leader that it was in the mining bonanzas of a century ago, Colorado's mineral industry continues to make substantial contributions to the economy. Among the principal minerals are coal, petroleum, molybdenum, and sand and gravel. Northwestern Colorado has some of the largest and most valuable coal deposits in the United States, but the industry is relatively dormant because of decreasing demand. Petroleum and natural gas reserves are mostly in the form of oil shales.

The state's power resources are important. Power generation is based on coal, hydropower, petroleum, and natural gas. Consumption is immense, and demands are difficult to meet. About two-fifths of the total capacity and production is privately owned.

Industry. Major industries include printing and publishing, machinery production, food and food products, metal production, lumber and wood products, and military ordnance and accessories.

Tourism. Although manufacturing, agriculture, and summer tourism are the mainstays of Colorado's econ-

Grant Heilman—Grant Heilman Photography



Lodge and (background) chair lifts and slopes at Vail, a ski resort in Eagle county, Colo.

Winter sports

omy, winter sports have grown at an almost alarming rate. Transport, housing, and lift facilities are continually expanding to meet the annual ski invasion.

Transportation. Colorado has a well-developed transportation system and ranks first among the mountain states in road mileage. Main highways tend to be east–west, circumvent high mountain masses, and follow valleys and canyons to their heads in the 32 mountain passes over the Continental Divide. The highest of the passes, at 12,183 feet (3,713 metres), is on the seasonal Trail Ridge Road in Rocky Mountain National Park. A number of other passes exceed 10,000 feet in elevation. One of the nation's major east–west interstate highways runs through the state, utilizing twin vehicular tunnels under the Continental Divide west of Denver.

Denver's Stapleton International Airport is a major centre in the nation's air traffic pattern. It is served by several major airlines, and carriers also link Denver with other Colorado cities and with neighbouring states. Several large airfreight depots adjoin Stapleton Airport. Railroad lines in Colorado are mainly bulk-freight carriers using multilevel railcars and flatcars for containerized freight, although a main east–west Amtrak passenger route passes through Denver and the Rockies.

Administration and social conditions. *Government.* In 1875 a convention drew up the constitution for the prospective state, which was admitted to the Union the following year. Because Colorado's admission occurred 100 years after the signing of the Declaration of Independence, it became known as the Centennial State.

The executive branch has five offices: governor, lieutenant governor, secretary of state, attorney general, and treasurer. All elected officials serve four-year terms. Numerous commissions, boards, and examiners are appointed to discharge the executive functions of state government. At the county level the constitution provides for several kinds of officers.

The Colorado General Assembly, which meets annually, comprises a Senate of 35 members elected to four-year terms and a House of Representatives of 65 members elected to two-year terms. The Legislative Council, created by statute in 1953, is a 13-member fact-finding agency of the General Assembly, and the Joint Budget Committee, established in 1959, is the permanent agency for fiscal and budgetary review. Under the Administrative Organization Act of 1968, the Commission on Uniform State Laws was created; and the Commission on Interstate Cooperation dates from 1937.

Local governments in Colorado have a jurisdictional balance with the state government. Counties can opt for either a home rule charter or for a constitutional government with offices and powers outlined by the state legislature. Three of the counties (Denver, Pitkin, and Weld) have home rule charters. Constitutional local governments generally consist of the following officers (variations depending on population): commissioners, clerk, sheriff, coroner, treasurer, surveyor, assessor, and attorney. Most terms are for four years.

The Colorado judiciary comprises the seven-member state Supreme Court, district courts encompassing one or more counties, and the county courts. In addition, the constitution provides for juvenile and probate courts, as well as municipal courts, which provide the grass-roots core of the judicial system.

Political balance

Since World War II each state political campaign has involved the issue of extension of federal activities, with Democrats generally committed to extension, and Republicans opposed. Since the 1920s each of the two parties has won control of the legislature in about one-half of the elections, indicating a fairly even balance within the state.

Education. The people of the Colorado Territory created the University of Colorado in 1861. The school did not open until 1877, but publicly supported primary education began in 1862, with secondary schools opening in the following decade. Population increases since the end of World War II have produced problems in funding, in instructional space, and in busing to achieve improved instructional programs. Curriculum and instructional revision has been undertaken in all areas, with foreign languages and mathematics particularly stressed.

In addition to the University of Colorado, at Boulder, the University of Denver (1864), the Colorado School of Mines, at Golden (1869), and Colorado State University, at Fort Collins (1870), were founded before statehood. Colorado has more than 40 colleges, junior colleges, and universities, of which about half are publicly supported. Of special note is the U.S. Air Force Academy, authorized by Congress in 1954. In 1958 it moved into its campus near Colorado Springs.

Observatories

Colorado has three important observatories. The High Altitude Observatory in Climax and the Chamberlain Observatory in Denver are operated by the University of Denver. The National Center for Atmospheric Research is a cooperative project of more than 40 U.S. universities, sponsored by the National Science Foundation. It has ties also to the Department of Astrophysics and Atmospheric Physics of the University of Colorado.

Health and welfare. Colorado's income per capita is higher than the national average, as is its expenditure on public assistance. There is considerable income disparity between the rural counties and the urban counties of the Front Range corridor. For example, Arapahoe's per capita income, the highest among the area's urban counties, is more than three times Conejos', the lowest among the area's rural counties. Provision of health, education, and other services is affected by this disparity, although the state government attempts to equalize services.

Cultural life. Red Rocks Park, a large natural amphitheatre in the foothills west of Denver, is Colorado's best known theatre; it hosts frequent musical events. The Central City Opera House, dating from 1878, has a summer season of opera and drama. Summer fare is available as well at Elitch Gardens Theatre, opened in Denver in 1891, and at festivals in Aspen and Boulder. The Denver Center for the Performing Arts was completed in the early 1980s. The Fine Arts Center in Colorado Springs is a regional art centre. The Denver Art Museum houses collections of Renaissance and Peruvian paintings as well as Oriental and pre-Columbian works.

Libraries have an important function in Colorado's cultural milieu. There has been a continuing trend to organize regional libraries to provide adequate service to every community. The Bibliographical Center for Research, Rocky Mountain region, is one of three such centres in the United States. The Colorado State Library is responsible for furnishing all Colorado institutions with research, reference, and general reading services.

The State Historical Society maintains nine historic properties. These include the Colorado State Museum in Denver, several houses and forts dating from the early days of the state, and the Ute Indian Museum in Montrose and the El Pueblo Museum in Pueblo. On the western and southwestern plateaus the Black Canyon of the Gunnison (established 1933), Colorado (1911), and Dinosaur (1915) national monuments are preserved as scenic attractions. Because of their cultural and historic value, Mesa Verde National Park (1906), Hovenweep (1923), and Yucca House (1919) national monuments in the southwest—all relics of former Indian civilizations—are preserved for public use and for archaeological study and exploration.

HISTORY

The earliest inhabitants. The influence of Indian culture on Colorado has been strong. Indian place-names have enriched the English vocabulary; Indian folktales, music, and dances have been assimilated into American culture; and Indian food and artwork have made valuable and unique contributions to the Colorado economy. The cliff dwellings in Mesa Verde National Park are among the physical remains of early Indian communities.

Indian and Spanish heritages

The Plains Indians, mainly Arapaho and Cheyenne, guided the explorers, traders, and trappers across the plains. The Indians knew the streams, the natural routes, the sources of fresh water and firewood, the areas of natural protection, and the feeding grounds of the buffalo. The Great Basin Indians, mainly the Ute, made similar contributions to knowledge of the Rocky Mountains.

The Indians, however, were displaced by Spanish explor-

ers from Mexico in search of cities of gold and silver. Fearing attacks by the United States, they strengthened the Spanish frontier in the 1840s with huge land grants reaching as far north as the Arkansas River. On these grants were established the first permanent white settlements in Colorado and, in 1851, the first recorded irrigation. The Spanish language is imprinted on Colorado geography. The state was named from the Spanish *colorado* ("red," or "ruddy"). Twenty large streams in Colorado are called *ríos,* and numerous cities, villages, and mountain ranges and peaks have Spanish names.

The U.S. territory. American exploration of Colorado began immediately after the purchase of the Louisiana territory by the United States in 1803. Dispatched to map, explore, and record scientific data about the new land were Zebulon Pike in 1806, Stephen Long in 1820, and John C. Frémont in 1842. As knowledge of the area spread, fur traders and trappers followed. Permanently stamped on the land are the names of such frontier scouts as Kit Carson and Jim Bridger. Fort Bent and Fort Saint Vrain served as collection points for furs, places for food and supplies, and shelter and protection from Indians.

The Colorado gold rush

In 1859 gold was discovered. A sudden great influx of people took place to the cry of "Pikes Peak or bust," and the bustling gold-dust towns of Central City, Black Hawk, Gold Hill, and Cripple Creek made mining history. The first gold was panned from the streambeds, after which came the search for the mother lode in the mountains.

In these frontier mining districts, civil and criminal codes were drawn up, and penalties for crimes were established. Of the thousands of seekers for gold, only a few found their bonanza. By the 1890s, the boom was over, and the mountains were largely vacated except for a few permanent mining towns.

Contemporaneous with the mining rushes was Colorado's period of territorial government. In 1861 congressional legislation provided for administrative officials to be appointed by the president. Seven governors were appointed in 15 years, and none served a full four-year term. In 1875 a constitution was drawn up and ratified by the territorial assembly, and in 1876 Colorado became a member of the Union.

Economic and social growth. Shortages of food during the gold rush led enterprising pioneers to initiate a new and significant component to the regional economy. Water was taken from the streams and put onto the land in what has been called the single most significant event in Colorado history. An entirely new social code and economy and a Western water law evolved. The industries and inhabitants of cities and towns came to depend upon irrigation agriculture. Sugar factories, which extracted the juice from the sugar beet, sprang up across the landscape.

By 1881 the buffalo herds on Colorado's plains had been replaced by cattle and sheep. From its mountain valleys, plains, and feedlots, Colorado became a major producer of meat. Automobiles, railroads, and a tunnel through the mountainous backbone united the mountains and high plateaus of western Colorado with the flat eastern portion of the state, and the flow of resources set the pace for industrial development. Also in the 1880s, steel was first produced in Pueblo, based on local deposits of iron ore and coal, and Pueblo became a major steel producer. Drought and the Great Depression of the 1930s triggered rural emigration but also spurred the construction of a large-scale transmountain water diversion project.

World War II and after. Fear of enemy attack on both U.S. coasts during World War II stimulated the development of government facilities in Colorado because of its interior, yet accessible, location. The Denver Federal Center, Rocky Mountain Arsenal, Camp (later Fort) Carson, Camp Hale, and other installations brought thousands of newcomers to the state. After the war many of these newcomers stayed on to develop ski resorts, the electronics industry, and the U.S. Air Force Academy. By the 1980s, some three-fifths of the state's population had been born outside of Colorado. Tourism replaced mining and agriculture as the mainstay of the economy, while official state policy promoted orderly growth of the economy and the infrastructure to support it. (M.J.Lo./J.L.Di.)

Idaho

Idaho, admitted as the 43rd state of the Union on July 3, 1890, is one of the Mountain states, but it is often also classified as part of the Pacific Northwest region, a region unified by the Continental Divide as an eastern boundary and by the Columbia River drainage basin, which covers virtually the entire area. The name is an invented one, formerly thought to be an Indian name (Ee-dah-hoe) meaning "gem of the mountains."

With 83,564 square miles (216,432 square kilometres), including 1,153 square miles of inland water, Idaho has twice the combined area of the six New England states. Its boundaries are both historical and geographic in derivation. The boundary with the Canadian province of British Columbia on the north follows the 49th parallel of latitude, while the southern border with Utah and Nevada follows the 42nd parallel; both lines were established by treaty—the northern between the United States and Britain in 1846 and the southern between the United States and Spain in 1819. The state's northeastern border with Montana—in the Idaho panhandle—follows the Continental Divide, while the eastern border with Wyoming incorporates a small slice of Yellowstone National Park. On the west, Idaho's border with Oregon and Washington is a 480-mile (770-kilometre) straight stretch except between Weiser and Lewiston, where Hells Canyon of the Snake River serves as a natural boundary.

Geographic and historical borders

Idaho is shaped much like a logger's boot, thereby accidentally reflecting the state's rugged forest and mountain terrain in which logging and mining play major roles. The residents of Idaho enjoy some of the largest unspoiled natural areas in the United States, including about 2,500,000 acres (1,012,000 hectares) of wilderness and primitive land in which roads and vehicles are seldom to be found. Since its development in 1936 Sun Valley has become an internationally known area for winter sports. Idaho also has large supplies of groundwater. Hot springs are found in many parts of the state and are used to heat some homes and buildings in Boise, the capital, whose name (French *boisé,* "wooded") reflects its settlement as an oasis for explorers who once crossed the desolate Snake River Plains. A frontier character is still evident in the individualism of voting that makes the crossing of party lines, especially to support liberal issues and candidates, a frequent occurrence in an otherwise fairly conservative climate.

PHYSICAL AND HUMAN GEOGRAPHY

The land. Diversity of the natural environment is characteristic of Idaho's landscape, creating a sectionalism that is reflected in its community life, politics, economy, and cultural development, as well as in the varieties of its soils and animal and plant life. Altitude is often a more important factor in controlling Idaho's climate than is latitude. The northern areas of the state are lower in elevation on the average than are much of the central and southern areas. Prevailing westerly winds from the Pacific Ocean blanket most of the state, especially the northern and southwestern regions. A drier, colder, continental climate is more noticeable in the southeastern counties, but Idaho has a milder climate than most of the states located in the same latitudes east of the Continental Divide.

Mountain ranges and rivers

Relief and drainage. Parts of four major physiographic provinces are included within Idaho: the Northern Rocky Mountains, the Middle Rocky Mountains, the Columbia Basin, and the Basin and Range. The Northern Rockies extend from the Canadian border to south central Idaho and occupy nearly half the state's area. Peaks of between 10,000 and 12,000 feet (3,000 and 3,700 metres) are common in central Idaho, and in the Lost River Range, Mount Borah at 12,662 feet (3,859 metres) is the state's highest peak. Other notable ranges in this part of the state include the Sawtooth, Pioneer, Continental Divide, Beaverhead, and Bitterroot. Peak elevations generally decrease northward to about 6,000 feet at the Canadian border.

The Middle Rocky Mountains occupy a narrow strip along the Idaho–Wyoming border. The area comprises several ranges that trend north–south and northwest–southeast and rise to between 7,000 and 10,000 feet.

Grass- and sagebrush-covered plateaus and valleys and a few small lakes are found between the ranges.

The Basin and Range Province extends into southeastern Idaho as a roughly triangular-shaped desert with its base along the Idaho–Utah border. A series of low north–south-trending block-faulted ranges separate small gravel-floored basins.

The remaining part of the state is included within the Columbia Basin, which in Idaho is occupied entirely by the Snake River plateau. The Snake River follows the plateau in a broad crescent across southern and western Idaho; next to the Northern Rocky Mountains, it is the major natural feature of the state. It rises in the southeastern part of the state, with tributaries in Yellowstone National Park, and flows from east to west through "sagebrush Idaho." With huge reclamation projects, the river valley contains most of Idaho's irrigated farmland. The course of the Snake includes Hells Canyon—at 7,900 feet (2,400 metres) North America's deepest gorge—and 212-foot-high Shoshone Falls. Its valley is a geologically complex sequence of lakes, lava beds, mesas, buttes, canyons, and desertscape, symbolized by the barren craters and cones of the Craters of the Moon National Monument.

The Snake River

Idaho has some 2,000 lakes, and water is the state's greatest single resource. A major portion of its industry, agriculture, and population lies along the Snake, which furnishes water in abundance for one of the nation's largest irrigated areas and developed hydroelectric power sources.

Climate. Idaho's mountainous topography produces an extremely diverse climate pattern. In general, as altitude increases precipitation increases and mean temperatures drop. On a different scale, the high mountains in eastern Idaho tend to hold back the cold winter air that predominates in Montana and Wyoming, giving the state higher mean temperatures than are found at the same latitude and altitude in those states and in more midcontinental locations. Mean January temperatures range from 17° F (−8° C) at Deadwood Dam in the central mountains to 31° F (−1° C) at Orofino in the central panhandle. July temperatures range from 60° F (16° C) at Deadwood Dam to 77° F (25° C) at Grand View in the southwest.

Idaho is situated in the rain shadow of mountains to the west in Washington and Oregon, and only higher elevations receive adequate amounts of rainfall. Most of the Snake River plateau receives less than 10 inches (250 millimetres) of precipitation annually, making it the driest part of the state. At the other extreme, large sections of the Northern Rockies record annual totals of more than 50 inches of precipitation, much of it in the form of snow.

Animal and plant life. Idaho is home to numerous bird, mammal, fish, amphibian, and reptile species. They occupy the state's six ecoregions, which vary from the sagebrush plains of the Snake River plateau to the alpine grasslands found in the higher mountain elevations. Vast evergreen forests cover about two-fifths of the state, largely the mountainous terrain. Western white pine, white fir and other true firs, and Douglas fir predominate in northern forests, while Douglas fir, lodgepole pine, and ponderosa pine dominate the southern forests. Aspen, maple, willow, birch, and mountain ash often carpet the lower slopes of higher mountains, especially in the north.

Idaho is one of the few states in which grizzly bears and timber wolves roam free, although in very small numbers. Grays Lake National Wildlife Refuge, in the southeastern corner of the state, is the site of a long-term attempt to introduce the whooping crane, one of North America's endangered birds, and to use sandhill cranes as surrogate parents to further increase the birds' population size and range. Several other endangered plant and animal species occur in Idaho, including Macfarlands four o'clock (*Mirabilis macfarlanei*) and the woodland caribou.

Endangered species

Idaho is known for its wilderness areas and for its cold-water fish populations. Fishing for trout, including the Kamloops (Kootenai) and steelhead varieties, is found on many of the thousands of miles of rivers and streams in the state. Some of the most remote mountainous country in the nation—the Frank Church–River of No Return, Selway–Bitterroot, and Gospel Hump wilderness areas—constitutes the heart of Idaho and the largest contiguous wilderness area in the coterminous 48 states. These wilderness areas and adjacent forested lands provide hunting for elk, Rocky Mountain goats, bighorn sheep, and mule deer. Idaho is also one of the few states in which there are large numbers of sage grouse and sharp-tailed grouse.

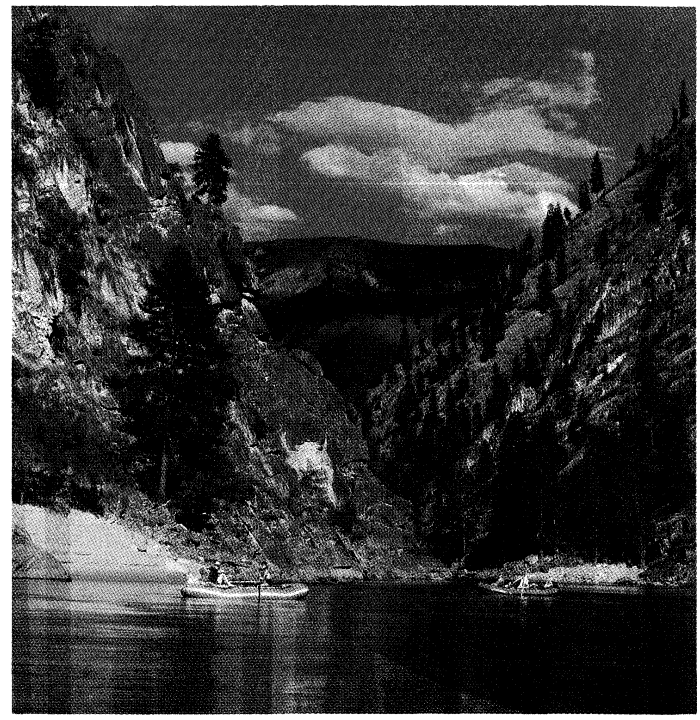
Settlement patterns. Many factors—religion, agriculture, transportation, topography, industry, cultural ties, and sectional pride—have contributed to Idaho's diverse regional characteristics. For many years writers and politicians consistently referred to the division of Idaho as northern Idaho—the 10 northern counties—and southern Idaho—the rest of the state. Studies of voting behaviour, however, indicate that four sections with distinct voting patterns have emerged: the 10 northern counties and three separate areas in the south, roughly the southwestern, the south central, and the southeastern sections. A more realistic regionalism has developed around trading and marketing centres, sometimes crossing the state boundaries. It consists of the following areas: Lewiston and Spokane, Wash., in the north; Boise, Twin Falls, Pocatello, and Idaho Falls in the south; and the Logan–Ogden–Salt Lake City axis in northern Utah.

Sectional characteristics

With the exceptions of mining and lumbering settlements, most of the settlements in southern Idaho tend to follow the course of the Snake River. A narrow agricultural strip runs northeastward from Pocatello. Agriculture continues its dominance to the west as far as the Boise Valley, where the state's largest concentration of population is located. The Palouse and Camas prairies are primarily agricultural, while the Lewiston area is industrial and service-oriented. Mining, lumbering, and agriculture are important throughout the north, while rural villages centre around a community life that includes churches, schools, commercial trading, banking, and service businesses to support the region's population.

The people. The rural counties of Idaho continue to lose people to the cities, while farms and ranches continue to get larger. Most of the immigration comes from the Western, North Central, and Southern states, whereas the bulk of emigration goes to the West. The population is more than 95 percent white, most of whom trace their ancestry to the United Kingdom, Germany, Ireland, France, Italy, or Poland. American Indians constitute the second largest group, and there are also significant numbers of Asians, blacks, and Hispanics.

© Charles Gurche



Boating on the Salmon River in the Frank Church–River of No Return Wilderness, Idaho.

Nearly half of Idaho's population are church members, and about half of them are Mormons. The next largest denominational groups are Roman Catholic, Methodist, and Presbyterian. The proximity to Mormon headquarters in Salt Lake City has resulted in strong religious ties to Utah, and the populations of some of the cities in the southeastern part of the state are more than 90 percent Mormon.

The economy. Economically, Idaho occupies a position between the highly developed and the underdeveloped states. Industrial expansion has replaced dependence on agriculture, lumbering, and mining in the 20th century, and Idaho has also emerged as one of the top states in tourist income. Government furnishes the second largest portion of Idaho's income. Labour, except in agriculture and small business, is heavily organized.

Agriculture and forestry. Huge herds of beef cattle and sheep graze not only in the prairie regions but also among the plateaus of the mountain regions. Idaho has some of the richest agricultural land in the United States, especially the irrigated region of the Snake River plain. Of the farm crops, potatoes have become almost synonymous with Idaho, though wheat, lentils, barley, oats, sugar beets, peas, beans, and alfalfa seed are important sources of farm income. Some two-fifths of the state's total area is in forests, and a huge quantity of lumber is cut from commercial timberlands each year. The primary commercial trees are Douglas fir, ponderosa pine, and western white pine.

Major crops

Mining. Although the discovery of gold and the subsequent gold rush created Idaho's mining industry, gold is no longer important to the state's economy. Idaho, however, ranks among the three leading states in silver, lead, antimony, and molybdenum production. Phosphate mining and processing is important in the southeast.

Hydroelectric power, much of it provided by power stations on the Snake River, is the main source of energy for both business and private users in Idaho. Natural gas and oil have been used increasingly, while waste wood products have declined in importance. The Idaho National Engineering Laboratory near Arco, operated primarily as a research and testing site for nuclear reactors by the federal government, also is used for energy production.

Industry. Value added by manufacturing exceeds the contributions of agriculture to the economy. Much of it is related to the processing of foods and forest and mining products, however, indicating how dependent the economy remains on primary products.

Transportation. The wilderness and the mountains have made transportation difficult. Idaho has only one major highway connecting the southern and northern parts of the state. Almost all interstate highways that pass through the state run from east to west. Three transcontinental railroads cross the panhandle, and one railroad serves the southern portion. Geographic conditions influence air travel as well, with many small airfields providing service to remote areas. These airfields are used largely by private and contract fliers. Idaho has a water route to the Pacific Ocean from Lewiston by way of the Snake and Columbia rivers. Due to slack water that permits oceangoing barges to dock at Lewiston, the city is an important industrial and shipping centre.

Administration and social conditions. *Government.* Idaho operates under its original constitution of 1889, and, typical of states admitted to the Union after the Civil War, it has a constitution that establishes the usual separation of executive, legislative, and judicial powers but limits the governor's strength. The constitution is detailed and includes many provisions that rightly belong in the statutes; it has been amended more than 90 times.

The only change in the government between 1890 and 1914 was the creation of numerous service and regulatory commissions and boards largely independent of the governor. Administrative reorganization after World War I consolidated these agencies in an effort to make them democratically responsive. The Great Depression of the 1930s brought on dozens of new commissions and boards, however, and the growth has continued. In 1974 the state government was again reorganized. The executive branch consists of six elected officers, independent of the governor, and 19 departments, under which more than 100 boards, commissions, councils, and committees are placed. The governor and other executive officials are elected to unlimited four-year terms. The legislature, which meets annually, comprises 42 senators and 84 representatives, and both senators and representatives serve two-year terms. Justice is administered by the Supreme Court, a court of appeals, and seven district courts and by county magistrate's courts. The district courts may originate cases and hear appeals.

Few voters in the nation are as independent as those of Idaho, and party cohesiveness is difficult to maintain; officials elected at one time often show a diversity of party and ideological stances. This independence usually is issue-oriented in state and national elections and personality-oriented in local elections. The two major political parties generally have dominated Idaho's political life. The voters have chosen Republican candidates slightly more than half the time, but the crossover vote, usually liberal and issue-oriented, can swing the outcome of any statewide election. The Republicans have long controlled the state legislature, but the governorship often is won by the Democratic Party. The preprimary party convention has been replaced by open primaries.

Voter independence

Idaho has more than 1,000 units of local government, including counties, municipalities, school districts, and special-purpose districts, the latter having limited taxing power. Most activities of local government are carried on by counties and cities. County commissioners, with a combination of legislative and executive functions, are very powerful. The state legislature has for many years refused to pass home rule legislation.

State debt is limited constitutionally to $2,000,000. The difficulty of achieving an equitable base for a sound system of public finance is increased by federal and state ownership of about 70 percent of Idaho's land area. The state's major revenue comes from personal and corporate income taxes and a sales tax, most of which is returned to public school districts. The state controls virtually no businesses or utilities except liquor sales, and among conditions made favourable to business development is the state's stance as a service rather than as a regulatory agency.

Education. Indian mission schools were supplemented by classes for white students when settlement began during the 1860s, and by the time of statehood the land-grant university had been chartered. The state Board of Education, dating from 1912, supervises appropriated funds, teacher certification, and related functions. The junior college system began on a district basis in 1939 and became a state function in 1965. The publicly supported University of Idaho (created in 1889 at Moscow), Idaho State University (1901, in Pocatello), and Boise State University (1932), as well as the private College of Idaho (1891, in Caldwell) and Northwest Nazarene College (1913, in Nampa), all offer advanced degrees. The University of Idaho is both a college of agriculture and the state's major educational institution. The university offers bachelor's and advanced degrees in areas that are related to the state's economy—engineering, mining, forestry, and wildlife and range science—and in other areas of business, education, and arts and letters.

Colleges and universities

Health and welfare. The electorate of Idaho is generally conservative on economic matters, but allocations for social and educational programs are liberal and are endorsed by both political parties. Nearly one-fifth of the state tax revenues goes into public health and public assistance programs. In addition, notable achievements have been based on a sense of social ethics, including a superior civil rights law. Living standards are relatively high because labour contracts follow national patterns, and living costs are below those of many states.

Idaho has several health districts that provide public health care throughout the state. Although Idaho has a low ratio of physicians to population and many rural hospitals find it difficult to remain open, the range of health services is comparable to that found in most U.S. states. All of the major cities have high-quality hospitals and excellent private medical services.

Cultural life. The opera houses in the mining camps, with various types of musical shows and serious drama,

were Idaho's first "culture." The missionaries and the churches set the patterns of cultural development for a long period. The University of Idaho has taken a leading role in developing programs in music, art, architecture, creative writing, and theatre. Students who return to small towns are, in many instances, the only college-educated people in the community, with the exception of the local attorney and physician. Other institutions of higher learning have also developed strong fine arts programs.

Culturally young, Idaho has contributed artists with wide reputations, including Vardis Fisher, a novelist whose writing decried dogma and tyranny; and Carol Ryrie Brink, who wrote books for adults and children. Ernest Hemingway wrote many of his books while living in Idaho, which he enjoyed for its wilderness aspects.

All of the colleges and universities have symphony orchestras, choral groups, and theatre programs, and a number of cities—including Boise, Pocatello, and Moscow—have orchestras. The University of Idaho and the cities of Coeur d'Alene and Lewiston have summer theatres. The Idaho Commission on the Arts has sponsored and promoted the development of art exhibits, lectures, literature, films, theatres, and music throughout the state.

Idaho has created a statewide system of parks. In addition, there are several national parks, and part of Yellowstone National Park is in Idaho. Craters of the Moon National Monument is one of the rarest geologic creations in the United States. The U.S. Forest Service maintains many campgrounds throughout the state.

HISTORY

Early history and settlement. Before the 1840s, when the buffalo herds disappeared and the wagon trains of settlers who were bound for California began to arrive, Indians had lived in the Idaho region for at least 10,000 years. In the north were the Kutenai, the linguistically identical Salish (Kalispel), the Coeur d'Alene, and the Nez Percé. Northern Paiute lived in the west central region, while the western Shoshoni and the northern Shoshoni occupied most of the southern lands. Most of these groups lived in small villages, consisting largely of family groups that moved according to the fishing, hunting, and gathering seasons, and the ties between them were weak. The tribes still live in approximately the same areas, some on the several reservations that are located within the state.

Indian groups

When the Lewis and Clark Expedition reached Idaho in 1805, about 8,000 Indians lived in the region. A trading post was erected on Pend Oreille Lake in the north in 1809, and fur traders were followed by missionaries. Gold seekers by the thousands poured through the area on their way to California in 1848, but many returned eastward after gold was discovered in northern Idaho in 1860. The settlers who followed wanted land and political stability, which had hitherto been uncertain; and slowly agriculture acquired economic dominance.

Territorial period. Idaho originally was in Oregon country, which was claimed first by Spain and then by Russia, Great Britain, and the United States; after the latter two had settled on the 49th parallel as the northern U.S. border, the Oregon Territory was created in 1848. It included the present state of Idaho, as well as what is now Oregon, Washington, and part of Montana. From 1853 to 1859 Idaho was divided between the Oregon and Washington territories. It then was part of Washington until it was organized separately as the Idaho Territory in 1863.

Frontier days

From a population of fewer than 17,000 in 1863, the territory expanded to nearly 90,000 at the time of statehood in 1890. Many new arrivals were Confederate refugees who, in the years following the Civil War, often dominated the legislature and opposed the Republican governors who were appointed by the federal government. Political strife and vigilante committees were elements of frontier life during the territorial decades. Among the events and trends that coloured the state's political and social life were the religious conflicts between the polygamous Mormons (Church of Jesus Christ of Latter-day Saints) and other sects; a strong sectionalism that divided various regions of the territory; a pioneer democracy that emphasized the rights and achievements of the individual; the completion of railroads, which fostered economic and population growth; the beginning of lead and silver mining in the mountains; and the creation of the University of Idaho in 1889 by the last territorial legislature that was convened prior to statehood.

Statehood. Labour protests that often erupted into violence were features of the 1890s era in Idaho. Through his unsuccessful prosecution in 1907 of William D. Haywood, an organizer of the Industrial Workers of the World (IWW), Senator William Borah became Idaho's major national figure until his death in 1940. During the 20th century Idaho has been engaged in developing its agriculture, forestry, and industry, while maintaining the more satisfying aspects of modern life at the doorstep of a natural wilderness. (B.A.M.)

Montana

Although its name is derived from the Spanish *montaña* ("mountain," or "mountainous region"), Montana has an average elevation of only 3,400 feet (1,040 metres), the lowest among the Mountain states. The mountains sweep down from the Canadian province of British Columbia, trending northwest–southeast into western Montana, into Idaho on Montana's western and southwestern border, and southward into Wyoming. The eastern portion of the state, however, is a gently rolling landscape, with millions of grazing cattle and sheep, and with only scattered evidence of human habitation. It forms a part of the northern Great Plains, shared with the Canadian provinces of Alberta and Saskatchewan to the north, with the U.S. states of North and South Dakota to the east, and with northeastern Wyoming to the south.

Only three states—Alaska, Texas, and California—have an area larger than Montana's 147,046 square miles (380,848 square kilometres), and only two states—Alaska and Wyoming—have a lower population density. The residents are relatively far from the markets for their products, as well as from the nation's manufacturing and supply centres. The state is strongly oriented toward the outdoors, toward summer and winter sports, toward hunting and fishing, and toward the long-distance trip for socializing and entertainment or as a cure for prairie- or mountain-born restlessness.

Sparse settlement

In spite of its northern location, Montana is very much a Western state. The main street of the capital, Helena, is Last Chance Gulch, the city's original name and a reminder of the prospectors who invaded the hills in the 1860s to pan for gold. By 1889, when Montana became the 41st state of the Union, the cattle drive was an institution, and the state had begun to emerge as one of the leading copper-mining centres of the nation.

PHYSICAL AND HUMAN GEOGRAPHY

The land. The western two-fifths of Montana falls within the Rocky Mountains, and the eastern three-fifths lies upon the Great Plains. Rocky Mountain Montana is a land of high mountains, deep valleys, green forests, and treeless crest lines, whereas Great Plains Montana is a vast, horizontal sweep of yellow rangeland, golden grainfields, and brown fallow strips. This contrast between mountain and plain is the most powerful geographic feature of the state.

Relief. In Rocky Mountain Montana the mountain ranges are aligned generally from north-northwest to south-southeast. They are made up of ancient, hard rocks that were compressed, folded, faulted, and otherwise contorted by the mountain-building forces that created the Rockies.

Rocky Mountain Montana

During the last Ice Age, glaciers carved the mountain crest lines and high valleys from rounded, convex terrain into sharp, rugged, concave topography and, when they melted, left the loose earth material that they had gouged out of the mountains as glacial deposits in the valley bottoms. The glaciers in Montana today are very small compared to the great tongues of ice of the past. The bottoms of the valleys between the mountain ranges consist mainly of alluvial floodplains and terraces; of benchlands and foothills carved on young, soft rocks; and of plains, terraces, and foothills made up of glacial deposits.

There is a contrast within Rocky Mountain Montana between mountains with narrow valleys and those with broad valleys. In the narrow-valley regions, which are the most rugged and spectacular of the state, the valley floors are humid and forested. There are two narrow-valley regions. One is northwestern Montana, which includes Glacier National Park with most of Montana's glaciers. The other lies in south central Montana at the northern end of Yellowstone National Park; this area contains the highest point in Montana, Granite Peak, which has an elevation of 12,799 feet (3,901 metres) above sea level. These two narrow-valley regions are separated by a broad-valley area in west central and southwestern Montana. There the valley bottoms are wide, dry, and grassy, permitting sweeping panoramic views of the mountain ranges.

Great Plains Montana

Most of Great Plains Montana is rather rough land. The country south of the Yellowstone River is mainly scattered hills, which make up the Breaks of the Missouri around Fort Peck Lake, with genuine plains found in the "Golden Triangle" north of Great Falls and plateaus elsewhere. Some of the hills, breaks, and valley bluffs form rugged badlands. The valleys of the major rivers flowing from the Rocky Mountains across eastern Montana are deeply incised. Scattered upon the plains and plateau surfaces are eight small mountain masses called Rocky Mountain outliers, which are like islands of the Rockies set out upon the plains.

The rocks underlying Great Plains Montana, except for the mountain outliers, are young, soft, and more or less horizontal. Roughly north of the Missouri River the plains rocks are covered by glacial deposits left by the continental ice cap, which occupied the area at the same time that alpine glaciers were sculpting the mountains to the west. The bottoms of the incised valleys are made up of alluvial floodplains, terraces, and soft-rock benchlands.

Drainage. Montana is the only state in the Union from which waters flow to Hudson Bay, the Gulf of Mexico, and the Pacific Ocean. The northwestern section of the state lies west of the Continental Divide and is drained to the Columbia River—and thus ultimately to the Pacific—by the Kootenai River, the Clark Fork, and its major tributary, the Flathead River. The Flathead flows into and then out of Flathead Lake, the largest natural lake in the state. The Kootenai flows out of Montana at the lowest elevation in the state, 1,820 feet (555 metres) above sea level. East of the Continental Divide Montana is drained by the Missouri River and its principal tributary, the Yellowstone. The Missouri is a tributary of the Mississippi River, which flows into the Gulf of Mexico and thus is part of the Atlantic Ocean drainage system. Finally, a small portion of Montana on the eastern slope of Glacier National Park drains ultimately to Hudson Bay.

Soils. On the mountain ranges and outliers the soils are mostly derived from the underlying, hard bedrock and are too steep and stony to be cultivated. In Great Plains Montana the soils south of the limit of glaciation are derived from the underlying, soft, Great Plains rocks. Most of the soils in the north, formed in glacial deposits, are undulating to rolling and generally are better for cultivation than the rougher lands with residual soils to the south. In the valley bottoms most of the soils are formed from unconsolidated alluvial and glacial deposits and are productive for crops when irrigated.

Climate. The climate of most of Great Plains Montana is semiarid, with warm summers and cold winters. Average annual precipitation is scant, about 13 inches (330 millimetres). Total snowfall is light. The chinook, a warm winter wind that blows on the plains near the foot of the Rockies, periodically interrupts the bitter cold—with January temperatures averaging 18° F (−8° C)—for which Montana is notorious.

In Rocky Mountain Montana there are several different climates arranged one above the other in altitudinal zones. The climate of the lowest zone—the dry valley bottoms—is similar to that in eastern Montana. The climates of the other zones become progressively cooler, wetter, and more snowy with higher altitude.

Plant and animal life. For the most part, Rocky Mountain Montana is forested and Great Plains Montana is grassland. In the Rocky Mountains and on the mountain outliers the vegetation, like the climate, is arranged in altitudinal zones. In the dry valley bottoms below the lower timberline are grassland, brushland, and open stands of trees called parklands. In the timbered belt on the mountainsides the vegetation is coniferous forest dominated by Douglas fir. The treeless crest lines above the upper timberline are made up of low alpine tundra vegetation, barren rock, and glaciers. In most of Great Plains Montana, where the land is not cultivated, the vegetation is short-grass grassland. Along the base of the Rockies are foothill prairie grasslands. Many of the hilly areas on the plains are covered by prairie parklands.

Wildlife

Montana has an abundance of rare and imposing species, notably grizzly bears, Rocky Mountain goats, bighorn sheep, and moose. These animals live mainly in the mountains along with more common species, such as American elk, mule deer, black bears, mountain lions, bobcats, and forest grouse. The grassland animals of Great Plains Montana and of the grassy western valleys include pronghorn, mule deer, coyotes, badgers, and plains grouse. Distributed nearly statewide along and near streams and lakes are white-tailed deer, beavers, muskrats, mink, bald eagles, ring-necked pheasant, ducks, geese, and swans.

Settlement patterns. The population of Montana is small, sparse, and unevenly distributed. In the two narrow-valley mountain areas the population is aligned in thin strips along the valley bottoms and towns are small. The broad-valley region has a heavy population by Montana standards, and some of the state's major cities and large towns, including Missoula, Butte, Helena, Bozeman, Kalispell, Anaconda, and Livingston, are located in the irrigated districts on the dry valley floors.

In Great Plains Montana there are continuous bands of relatively dense population along the irrigated bottoms of the major incised valleys. All of the cities and major towns, except Lewistown, are lined up like beads on a string along these strips. Along the Yellowstone River are Billings, Miles City, Glendive, and Sidney; and along the Missouri River and its tributary, the Milk River, are Great Falls, Havre, and Glasgow. Population is moderately dense and fairly evenly spaced in the dryland grain areas, but it is sparse and uneven in the livestock-ranching areas.

The people. Most of the inhabitants of Montana today trace their ancestry to the countries of western, northern, and, to a lesser degree, eastern Europe—mainly Great Britain, Ireland, Germany, France, The Netherlands, the Scandinavian countries, and Poland.

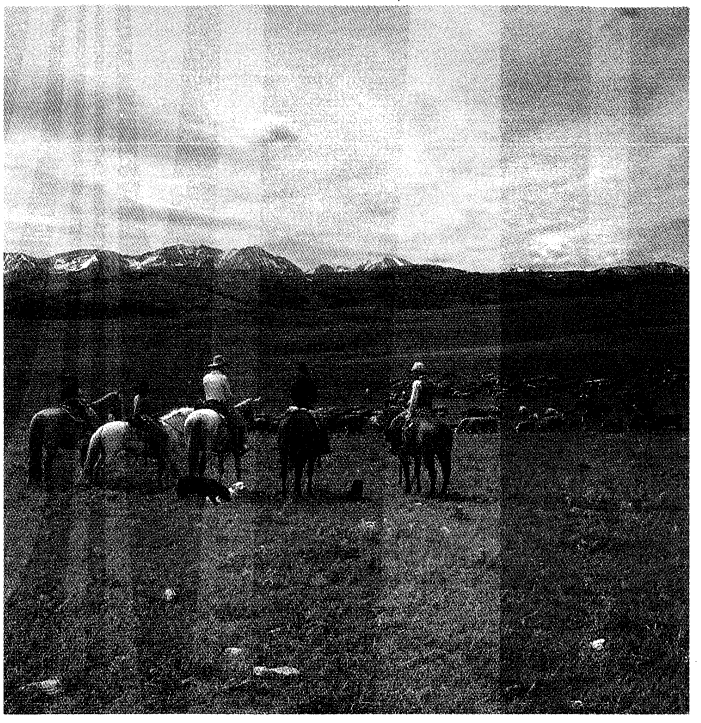
Indian population

The one significant exception is the American Indians. Montana has seven Indian reservations and an Indian minority of about 5 percent. Nearly two-thirds of the Indians live on the reservations and most of the rest in the cities near the reservations, notably Missoula, Great Falls, and Billings. Blacks make up only a tiny fraction of the Montana population. Few blacks are permanent citizens of the state but—like the small Asian and Hispanic populations—are temporarily in Montana for employment or educational opportunities.

About half the inhabitants of Montana are affiliated with organized religious groups. Mormonism has increased in Montana. Many of the Indians were nominally converted to Roman Catholicism by missionaries.

The economy. Montana's economy is dominated by the primary sector—agriculture, forestry, mining, and energy production. The outdoor recreation industry, however, has also become important, and some high-technology industries have come to the state.

Agriculture. Beef cattle, sheep, grain, sugar beets, potatoes, and fruit are produced on irrigated farms in the broad, dry valleys of Rocky Mountain Montana and in the incised valleys in Great Plains Montana. Wheat and barley are grown on large dryland grain farms throughout the two major grain-growing regions—the Golden Triangle and northeastern Montana—and in patches elsewhere. Most of the rest of the state is rangeland and is used in the livestock-ranching industry for the production of beef cattle and sheep. There is some meat packing, flour milling, and sugar refining in Montana, but most of the farm and ranch products are processed outside the state.



Cowboys grazing their cattle on the summer range west of Gallatin Gateway, southwestern Montana. The Spanish Peaks, part of the Madison Range, appear in the background.
© James Fain

Mining and water resources. Coal is one of the major mineral resources of Montana. Along with petroleum and natural gas, coal is extracted from the young, soft rocks of Great Plains Montana. Coal-fired, thermoelectric plants are located at Colstrip and Billings. Silver, gold, copper, platinum, talc, phosphate, vermiculite, sapphire, and other minerals are mined from the old, hard rocks of Rocky Mountain Montana and of the mountain outliers. There is a lead smelter located at Helena and an aluminum plant at Columbia Falls.

Oil and gas production

Petroleum was discovered in commercial quantities at Elk Basin in 1915. The Elk Basin, Kevin-Sunburst, and Cut Bank fields led in production of petroleum and natural gas for several years. The great Williston Basin was developed in 1951, but the Bell Creek field in Powder River county has been the most productive. Large petroleum refineries are located at Billings and Laurel.

Montana's tremendous water resources provide for hydroelectric power production and for uses in other sectors of the economy. Runoff from the forested sides of the Rocky Mountains and the outliers recharges groundwater, fills lakes and reservoirs, and generates the flow of the great rivers. There are several large dams, power stations, and reservoirs, which are clustered mainly in the mountains of northwestern Montana.

Forestry. Lumbering and the manufacture of forest products are vital to western Montana. Of the more than 14,000,000 acres (5,700,000 hectares) of commercial forestland, more than two-thirds is owned by the federal and state governments. The forestry industry includes the manufacture of plywood and of pulp and paper products. About one-third of Montana's forest products are sold to states in the Midwest.

Services. A large majority of all jobs in Montana are in the service sector; Billings, Great Falls, Missoula, and Butte are the state's major regional service centres. Tourism has become a significant component of Montana's economy.

Transportation. Railroad passenger service and total route length in Montana have decreased since the 1960s. Total highway mileage in Montana is relatively low, but there is a well-developed network of interstate and primary highways between population centres. Several major airlines and air-taxi lines serve the state, and many small planes are privately owned.

Administration and social conditions. *Government.* Montana's original constitution, adopted upon statehood in 1889, was replaced by a new one in 1972. The new document provided for a voter initiative process that has been used vigorously in attempts to enact new laws and to amend the constitution itself.

Allegiance is so evenly divided between the Republican and Democratic parties that election patterns are not predictable. Ticket splitting is common. The governor and the lieutenant governor are elected on one ticket, and both serve four-year terms. The state has an open primary, voters need not declare party affiliation, and registration is permanent unless a voter does not vote within four years.

The executive branch comprises 19 cabinet-style departments, with virtually all appointments directly controlled by the governor. The two-house legislature meets in odd-numbered years for 90-day sessions and is composed of 50 senators and 100 representatives.

Judges are elected without party designation. The highest court is the seven-member Supreme Court. The state is divided into 19 judicial districts, in which 36 district judges serve. On the lowest level are justices of the peace and police judges.

The county is the highest level of local government. Its powers and duties are defined and limited by state statutes. Three elected commissioners are the chief administrators, though a full-time manager may be employed instead. Municipal governments, like those of counties, derive all their authority from the state. They can, however, enact local ordinances, whereas counties cannot. Municipalities have police forces, and each county has an elected sheriff, who appoints deputies and has jurisdiction outside towns and cities. Because cattle rustling is a continuing problem, some sheriffs and deputies act as brand inspectors to prevent the sale of stolen livestock.

Education. Although rural schools continue to consolidate and introduce bus transportation, many children attend small and not always adequate one- or two-teacher country schools. School districts are corporate bodies headed by a county superintendent of schools and governed by elected school boards.

Higher education

The state's system of higher education, chartered in 1893, includes universities at Missoula and Bozeman; regional four-year colleges at Billings, Havre, and Dillon; and a college of mineral science and technology at Butte. There are also a number of church-affiliated private colleges, junior colleges, and public postsecondary vocational-technical schools.

Health and welfare. Montanans pay various penalties for the wide-open spaces they enjoy. An increasing number of communities have no dentist, physician, or hospital. Montana's welfare program is state-supervised and administered by county departments of public welfare. Because of sparse population and few private social service agencies, costs are relatively high. Welfare departments are caught between pressures from taxpayers to curb programs and from organized low-income groups to provide wider services. There are special programs on Indian reservations.

Cultural life. Most artistic activity centres in the cities with colleges and universities, several of which sponsor visits by lecturers and professional artists of various kinds in addition to the presentation of the work of faculty members and students. Several cities have symphony orchestras that include some professional musicians.

The Montana Institute of the Arts, founded in 1948, is a grass-roots organization that ties together the scattered, often isolated practitioners of various arts and crafts through publications, an annual festival, and traveling exhibits. The Montana Arts Council, a state agency affiliated with the National Endowment for the Arts, funds dozens of local cultural organizations, primarily for music, drama, dance, literature, and the visual arts. There are summer theatres in a number of communities; and the Montana Repertory Theater, based in Missoula, tours both inside and outside the state.

Several Indian tribes hold traditional dance ceremonies at which outsiders are welcome. There is an annual reenactment of the Battle of the Little Bighorn on the Crow Indian Reservation. Rodeos abound, as do square-dance

groups, and Montana is a thriving centre for old-time fiddling. In Red Lodge an annual nine-day Festival of Nations, originated to ease tensions among European ethnic groups of coal miners, has become a tradition.

Historical and art museums

The Montana Historical Society maintains a museum, art gallery, and specialized library in Helena. The C.M. Russell Museum, in Great Falls, specializes in the works of the cowboy artist Charles Marion Russell. Billings is the location of the Yellowstone County Fine Arts Center and the Yellowstone County Museum. The excellent, small Museum of the Plains Indians is in Browning. Many communities nurture art galleries and small museums of local historical interest.

Montana's spectacular scenery is one of its premier recreational attractions. A large proportion of the state's land is given over to state and national parks, forests, recreational areas, wildlife refuges, and wildernesses, which provide venues for a wide range of outdoor recreational activity.

HISTORY

Settlement. At the beginning of the 19th century the Crow Indians occupied the south central portion of present-day Montana, the Cheyenne the southeastern corner, the Assiniboin and Atsina the northeastern corner, the Blackfoot the central and north central area, and the Kutenai the northwestern corner. The Pend d'Oreille had a territory around Flathead Lake, the Kalispel were in the mountains west of there, and the Flathead occupied the Clark Fork and Bitterroot valleys. The southwestern corner was disputed territory. The Flathead were later forcibly moved to their present reservation in the Flathead valley. Most of the other tribes are now living on reservations within their respective territories.

The first white explorers known to have set foot in Montana were the members of the Lewis and Clark Expedition (1804–06). Fur trappers and traders followed, setting up forts to trade with the Indians. The only early trading post to survive as a present-day town was Fort Benton, which was established in 1846 and became an important port on the Missouri River. Roman Catholic missionaries followed the fur traders and in 1841 established Saint Mary's Mission near present-day Stevensville, believed to be the first permanent settlement in Montana. Trailblazers carved the northern Overland Route to Montana from the east, the Bozeman Trail from the southeast, and Mullan Road westward from Fort Benton, the head of navigation for steamboats on the Missouri.

Mining era

Gold prospectors flocked in after rich placer deposits were discovered in the early 1860s. Montana Territory was established in 1864 with Bannack, in Grasshopper Creek, its first capital, and Virginia City, in Alder Gulch, its second. As pressure from white settlers increased, the Indians fought to protect their hunting grounds. The Dakota (Sioux) and Cheyenne won their last major victory in June 1876 at the Battle of the Little Bighorn. A band of Nez Percé under Chief Joseph won a battle in the Big Hole Basin the following year and fled toward Canada, only to be met and defeated by U.S. troops a few miles south of the international boundary.

Hard-rock mining began in the 1880s, and Montana became a state on Nov. 8, 1889, with Helena as the capital. Butte began as a gold camp, but shaft mining commenced when vast deposits of copper were discovered there. Butte subsequently became known as the "Richest Hill on Earth," and the world's largest smelter was built at nearby Anaconda. The so-called War of the Copper Kings was won by Marcus Daly, whose Anaconda Company became one of the largest mining conglomerates in the world. The company smashed the mining unions, influenced the state legislature, acquired almost all of Montana's daily newspapers, and virtually controlled the state for three-quarters of a century.

Cattle and sheep grazing in Great Plains Montana started in the 1860s, when herds were driven overland from Texas. The vast grasslands seemed ideal for cattle, but a severe winter in 1886–87 virtually wiped out the herds. Beginning around 1900, homesteaders began pouring into the plains country to bust the sod and grow grain on dry land. After a few years of bumper crops and high prices, a series of dry years brought financial disaster and mass exodus.

Oil and natural gas production began in Great Plains Montana in 1915, expanded greatly in the 1950s, and peaked in the 1960s. Coal mining, which began in the days of coal stoves and steam locomotives, increased dramatically in the 1970s.

The closing of the copper mines at Butte, the smelter in Anaconda, and the copper refinery at Great Falls in the early 1980s marked the end of the copper century and a turning point in Montana's history. The state no longer relies so heavily on the primary economic sector; more emphasis is placed on tourism and on new and innovative businesses that provide jobs without causing deterioration of the state's magnificent mountains, crystal waters, and wide-open spaces. (D.M.J./J.M.C.)

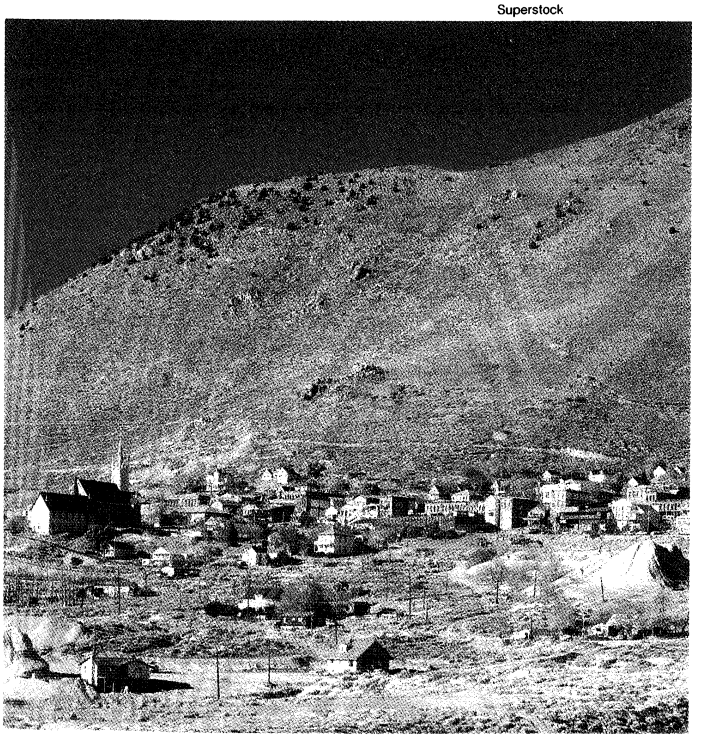
Nevada

Located in a mountainous region that includes vast semiarid grasslands and sandy alkali deserts, Nevada became the 36th state of the United States on Oct. 31, 1864. Its name was derived from the Spanish *nevada* ("snow clad"), a reference to the high mountain scenery of the Sierra Nevadas that are located on the state's southwestern border with California. Other neighbouring states sharing many of the geographic features that make Nevada the nation's most arid state are Oregon and Idaho to the north, Utah to the east, and Arizona to the southeast. Nevada is the seventh largest of the 50 states, but it is also one of the most sparsely settled. Carson City is the state capital.

Nevada appears far removed from the days when Virginia City was a fabled frontier town, thriving on the rich silver mines of the Comstock Lode. However, many frontier qualities persist, though subtly transformed by a sophisticated urban environment. The prospector digging against odds to find a bonanza has been replaced by the fortune seekers in the gambling casinos of Las Vegas and Reno, and the erstwhile "saloon diversions" have evolved into lavish nightclub entertainment.

While the great majority of Nevadans live in the two main cities—more than one-half of them in the Las Vegas metropolitan area and almost one-fourth in that of Reno—the undeveloped lands of the state provide a largely unknown resource. Combined with the major sci-

Superstock



Virginia City National Historic District on the eastern slope of the Sierra Nevada range, western Nevada.

entific activity related to the federal government's atomic research facilities, the modern cities and desert reaches make Nevada a unique phenomenon among U.S. states.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Most of Nevada lies within the Great Basin section of the Basin and Range Province, where the topography is characterized by rugged mountains, flat valleys with occasional buttes and mesas, and sandy desert regions. More than 30 north–south mountain ranges cross the state; the highest elevations are Boundary Peak, at 13,143 feet (4,009 metres), and Wheeler Peak, at 13,063 feet. The southern area of the state is within the Mojave Desert, and the lowest elevation, 470 feet (143 metres), is in the Colorado River Canyon.

Drainage. The state's rivers depend on the melting of winter snows and on spring rainfall. Almost all of the rivers drain into lakes that have no outlets or into shallow sinks that in summer evaporate into alkaline mud flats. The Humboldt, the largest of Nevada's rivers, provides the state's only major east–west drainage system. The Truckee, Carson, and Walker rivers, which rise in the Sierra Nevada, serve extensive irrigation and reclamation projects in their areas. The Muddy and Virgin rivers in southern Nevada are related to the Colorado River system.

Several lakes provide scenic and recreational attractions. Lake Tahoe on the California–Nevada border is particularly notable for its clarity, depth, and scenic beauty. Pyramid, Walker, and Winnemucca lakes are remnants of an ancient sea. In relation to its area, however, Nevada has little surface water. The increasing demands of urbanization, industry, and agriculture are exhausting both groundwater and surface resources, and scarcity of water is an increasing concern. The impounded waters of Lake Mead, extending for 117 miles (188 kilometres) behind Hoover Dam, provide reserves for the southeastern area.

Scarcity of water

Climate. The high Sierras along the state's western boundary often cause clouds of Pacific origin to drop their moisture before reaching Nevada, thus producing a semi-arid climate. The driest regions are in the southeast and near Carson Sink, where annual rainfall seldom exceeds four inches (100 millimetres). The northeast has as little as eight inches of precipitation, whereas that of the northwestern mountains often reaches 24 inches. Temperatures vary as widely. In the north July temperatures average 70° F (21° C), and in the south 86° F (30° C). In January the averages range from 24° F (−4° C) in the north to 40° F (4° C) in the south. The northern and eastern areas have long, cold winters and short, relatively hot summers, whereas in southern Nevada the summers are long and hot and the winters brief and mild. Regional differences are pointed up by variations in the growing season: Las Vegas has 239 days, Reno 155, and Elko only 103.

Plant and animal life. Despite aridity and rugged terrain, Nevada shows considerable variety in vegetation. In the lower desert areas, mesquite, creosote, greasewood, yucca, and more than 30 varieties of cacti abound, while sagebrush and Joshua trees flourish at the higher elevations. Throughout the state, particularly during the period after the spring rains, more than 2,000 varieties of wildflowers have been identified. Mountain forests contain pine, fir, and spruce as well as juniper and mountain mahogany. The piñon pine is characteristic in the high mountain regions, and the rare bristlecone pine—one of the oldest living species of trees—is native to the Toiyabe Range.

The animal population of Nevada includes those species that are best adapted to temperature extremes and to lack of moisture. Among the larger animals are bighorn sheep, several varieties of deer, and the pronghorn. Rabbits and other rodents are found in abundance. The desert harbours such reptiles as geckos, horned toads, tortoises, and sidewinder rattlesnakes. Predators such as the coyote and bobcat are common. The permanent bird population of the state is somewhat limited, but there are seasonal visitations by a great variety of migratory birds. Game birds that can be found in the state include sage grouse, pheasant, and quail; and Nevada's rivers and lakes contain large quantities of bass, trout, crappie, and catfish.

The people. The majority of Nevadans are of European ancestry, almost 90 percent of whom were born in the United States. An estimated 5,000 to 10,000 persons trace their ancestry to Basques recruited as sheepherders from their Pyrenean homeland. Spanish-speaking Americans, mainly those of Mexican and Cuban origin, are concentrated in the southeast. Descendants of the Paiute, Shoshoni, and Washoe Indians live on several reservations. Blacks, mostly in the Las Vegas and Reno areas, make up a very small percentage of the state's population.

The predominant religious groups are Mormons and Roman Catholics. There are a variety of Protestant denominations and a Jewish minority.

From the 1950s through the '70s Nevada's population grew by about 70 percent, and by the mid-1980s it was expanding three times faster than that of the nation as a whole. In spite of a birth rate slightly above, and a death rate slightly below, the national average, this growth was largely the result of immigration from other states. The impact of this immigration has been felt most strongly in Las Vegas and surrounding Clark county and in Reno and surrounding Washoe county. Most Nevadans are urban and are engaged in the booming economies of those two metropolitan areas.

Immigration and urbanization

The economy. Although the traditional bases of Nevada's economic life, mining and agriculture, remain important, they are far overshadowed by manufacturing, government, and tourist-related services.

Mining. One of the richest mineral regions of the nation extends eastward from California across Nevada and into Arizona. Copper production, which had been the largest component of mineral production, dropped dramatically in the 1970s and '80s, when the state's leading copper producers shut down operations. Copper is now produced only as a by-product of gold-mining operations. Gold has replaced copper as the most commercially valuable of the state's minerals, and the annual output is among the highest in the nation. Nevada is also the leading producer of barite and mercury. The McDermitt Mine in Humboldt county is the largest single source of mercury in the United States. Although silver production dropped in the late 1970s, new mines began operation in the 1980s. Other important minerals include gypsum, sand and gravel, crushed stone, tungsten, and magnesium. Petroleum was discovered in Nye county in 1954, and commercial production began in the 1970s.

Agriculture. Nevada's agriculture depends on irrigation. Even in the river valleys, farmers and ranchers pump additional groundwater for their crops and livestock. About 750,000 acres (300,000 hectares) are classified as cropland, compared with about 7,600,000 acres of pasture and rangeland. In the 20th century farms and ranches have increased in acreage while declining in number.

Croplands are devoted mainly to forage and feed crops, with alfalfa the major commercial crop. Livestock ranching, however, is the primary source of agricultural income. The large cattle and sheep ranches are chiefly in Elko, Humboldt, and Lander counties. Most of the cattle are shipped to California or the Midwest for fattening and marketing. Dairy and poultry farms have become important in western and southeastern Nevada, where horse ranches also have been developed.

Ranching

Nearly one-fifth of Nevada's total acreage is devoted to forests and woodlands. More than 5,000,000 acres have been designated as national forests, and private holdings support only a small-scale lumber industry. Aside from lumber production, the forests are of importance for the conservation of water and wildlife and in providing recreational opportunities.

Industry. Manufacturing has expanded and diversified, and most of the larger enterprises are located in Clark or Washoe counties. The leading product groups are stone, clay, and glass products; printing and publishing; food and food by-products; and chemicals. The largest industrial complex is located in Henderson, where major factories process titanium ore and produce industrial chemicals.

Most of the electricity is generated by coal and natural gas, but a small amount is produced by hydroelectric plants. Coal and natural gas are used in power plants in southern Nevada. Hoover and Davis dams are major

power sources, supplemented by imports of hydroelectric power from California and Oregon.

About 85 percent of Nevada's land is owned by the federal government. Following establishment of the Nevada Test Site by the federal government in the 1950s, a complex of research and development enterprises, mainly in the aerospace, civil defense research, biomedical environmental protection, and electronics fields, developed in the Las Vegas area. These industries have come to rival similar industries in California and in the Boston and Washington, D.C., areas. The test site itself is a major centre for underground nuclear detonation and for 15 years, until 1972, for nuclear rocket development. Thousands of military personnel are stationed also at Nellis Air Force Base and Fallon Naval Air Station.

Tourism. Tourism and its related activities bring millions of visitors; contribute more income than mining, agriculture, and manufacturing combined; and employ about one-third of the work force. Although millions of people visit Lake Mead and other recreational and scenic areas, the tourist industry centres on several attractions that largely are unique to Nevada among the U.S. states.

Legalized gambling and related lures

The 24-hour-a-day gaming casinos bordering The Strip and Glitter Gulch in Las Vegas are the most publicized aspect of the legal gambling industry. Important adjuncts to the casinos are the luxury hotels, gourmet restaurants, golf courses, and nightclubs that have made Las Vegas—and, to a lesser extent, Reno and Lake Tahoe—a major centre of live entertainment in the nation. Small towns also emphasize the hospitality industry and tourism. Unique to the rural counties of central Nevada is legal prostitution.

Transportation. Its vast size makes Nevada heavily dependent upon air transportation. The state is served by several national airlines. There are numerous airports and airfields, and both Las Vegas and Reno have been designated as international ports of entry.

Three major railroads cross the state, while short lines serve as feeders where truck competition has not caused their discontinuance. Nevada's public roads include primary and secondary highways as well as municipal and rural roads. Two of the federal highways are part of the interstate system.

The three major transportation and trade centres of the state are Reno, the principal distributive centre for northwestern Nevada and northeastern California; Elko and Ely, in northeastern Nevada; and Las Vegas, the commercial centre for southern Nevada and nearby areas of Utah and Arizona. Warehousing and trucking industries flourish because of Nevada's strategic geographic location and the "free port" tax exemption for goods continuing in transit.

Administration and social conditions. *Government.* Nevada is governed under its original constitution, adopted in 1864 but since amended in many respects. The chief officials, including the governor, lieutenant governor, attorney general, secretary of state, controller, and treasurer, are elected to four-year terms. In addition to the usual departments and agencies supervising areas of public concern, the state Equal Rights Commission oversees areas of discrimination of various kinds, while the Gaming Control Board oversees operations of the gambling industry.

Nevada's bicameral legislature comprises the Senate of 21 members elected for four-year terms and the Assembly of 42 members elected for two-year terms. It convenes in January of odd-numbered years.

The highest judicial body is the Supreme Court, composed of a chief justice and four associate justices. There are also district courts, subdivided into departments on a population basis. Cities and townships have courts staffed by municipal judges and justices of the peace. All judicial offices are subject to nonpartisan elections.

Local government comprises 16 counties, 16 cities, and 56 townships. As Nevada traditionally has been rural-oriented, the county remains the primary unit of local administration. Each county has a public administrator, board of commissioners, district attorney, sheriff, and other officials. Cities and towns are incorporated under charters granted by the legislature, most of them with a mayor–council form of government.

Finance. Nevada's fiscal policies have been markedly conservative. The constitution rigidly limits both taxation and indebtedness. The bonding capacity cannot exceed 2 percent of the total assessed valuation of real property in the state, and there is a maximum tax rate on real estate. Even more unusual is the absence of state taxation upon inheritances and all types of income. A gaming tax and the sales tax are the principal sources of state income. State taxation provides about two-thirds of general revenue, with most of the balance coming from federal grants and subventions.

Education. The public school system is controlled by an elected Board of Education, which delegates administrative responsibilities to an appointed superintendent of public instruction. Local school districts, coextensive with the counties, receive supplementary funding from the state. School attendance is compulsory for those between the ages of seven and 17.

Institutions of higher education

The University of Nevada originally was established at Elko in 1874 under the provisions of the Morrill Land-Grant College Act; 12 years later it was moved to Reno. In 1951 an extension branch was established in Las Vegas, which since has become the autonomous University of Nevada, Las Vegas. There are two-year community colleges in Elko, Carson City, Reno, and North Las Vegas. To supplement campus instruction the Desert Research Institute and the Agricultural Experiment Stations provide statewide research services.

Health and welfare. Nevada's welfare programs and its custodial institutions are administered by the Department of Human Resources. Old-age and welfare allotments are given on the basis of need, and public assistance is available for the blind and other handicapped residents. Support payments are provided for dependent children, and orphanages are located near Carson City and Boulder City. The state mental hospital is in Sparks, and there is a mental health facility in Las Vegas. Penal and rehabilitation institutions include the state prison in Carson City, auxiliary prisons in Jean and Indian Springs, a girls' training centre at Caliente, and an industrial school for boys at Elko. Health care, housing, and public safety are responsibilities of local government or private enterprise.

Cultural life. Nevadans traditionally have mingled rural conservatism and the individualism of the Old West. Until the mid-20th century its population was small and dispersed, and cultural values were those of an agrarian society. With the establishment of resort industries and increases in population, however, Las Vegas and Reno developed marked metropolitan characteristics. Not only has the economy diversified but cultural activities also have burgeoned. Recognizing this trend, in 1967 the state legislature established the Nevada State Council on the Arts to coordinate and stimulate cultural activities.

Both major cities have well-established programs in the performing arts. The universities sponsor lectures, concerts, and theatrical productions, while the tourist industry regularly features some of the most famous entertainers in show business. Both Reno and Las Vegas support symphony orchestras and have commercial and public art galleries. Traditional Indian arts and crafts have been revived on reservations and in urban colonies.

Nevada's frontier heritage is commemorated by annual pageants and festivals. During Helldorado Week, held in Las Vegas each May, the townspeople wear Western garb and stage a series of rodeos and parades. There is a Basque Festival in Elko, and the Reno Rodeo is an outstanding Fourth of July celebration. The state observes its anniversary, Admission Day, on October 31, highlighted by a parade and costume ball in Carson City.

Old West festivals

The Nevada State Museum, in Carson City, emphasizes the mining industry and mineral collections. Anthropological artifacts are featured at the Lost City Museum in Overton, at the Museum of Natural History in Las Vegas, and at the Southern Nevada Museum in Henderson. The Mackay School of Mines Museum, on the Reno campus of the state university, is oriented toward metallurgical, mineralogical, and geologic specimens. The Nevada Historical Society, also in Reno, has pioneer mementos, the most complete holding of Nevada newspapers, and a sizable historical reference library. The library of the University

of Nevada, Reno, has the largest collection of books in the state, while the Nevada State Library in Carson City is notable for its excellent collection of legal works.

The state and federal governments maintain parks, forests, historical monuments, and recreational areas. The Valley of Fire State Park, near Overton, is known for its brilliantly coloured rock formations and Indian petroglyphs. Mormon Station Historic State Monument, in Genoa, is the site of the first permanent settlement in Nevada; the Death Valley National Monument is on the border between Nevada and California. Red Rock Canyon, with unusual high-desert terrain and a spectacular multicoloured escarpment, is near Las Vegas, and Cathedral Gorge State Park, near Pioche, displays red and gold rocks that resemble church spires. Humboldt National Forest, in east central Nevada, includes the Jarbidge Wilderness and the Ruby Mountains. The Lake Mead National Recreation Area, which contains Hoover Dam, has fishing, boating, and swimming facilities; and Great Basin National Park, featuring the Lehman Caves and the Wheeler Peak area, is located near the Nevada–Utah border.

HISTORY

Archaeological evidence indicates that prehistoric Indian settlements existed in Nevada more than 20,000 years ago. Cave dwellers left picture writings on rocks in southern Nevada, and Basket Makers and Pueblo Indians also flourished there. Explorers of the early 1800s found Mohave, Paiute, Shoshoni, and Washoe Indians at various locations within Nevada.

Explorers and settlers. Missionaries and fur traders were in the vanguard of the exploration of the Nevada area. The missionary travels of Francisco Garcés from New Mexico to California in 1775–76 were imitated by other Spanish Franciscans. In 1825 Hudson's Bay Company trappers explored the northern and central region, and two years later Jedediah Smith led a party of Americans into the Las Vegas Valley and across the Great Basin. By 1830 the Old Spanish Trail was bringing traders to the area from Santa Fe and Los Angeles, and in 1843 and 1845 John C. Frémont's explorations publicized the Great Basin and the Sierra Nevada region. During the 1840s pioneers followed the Humboldt Valley–Donner Pass route to the Pacific Coast, and the gold rush of 1849 greatly expanded migration through Nevada to California.

Nevada, which came under U.S. sovereignty through the Mexican cession of 1848, was a part of California until it was incorporated into the newly organized Utah Territory in 1850. In 1849 a settlement was made at Genoa (then Mormon Station) in Carson Valley, but the population remained sparse until the discovery of the famous Comstock Lode in 1859. From that time on Nevada ceased to be merely a highway for gold seekers on the way to California. Virginia City became the most famous of all the Western mining camps, and the rapid influx of prospectors and settlers resulted in the organization of Nevada Territory in 1861.

Discovery of mineral wealth

The American Civil War gave strategic importance to the new territory. President Lincoln realized that Nevada's mineral wealth could help the Union, and he also needed a Northern state to support the proposed antislavery amendments to the Constitution. Although Nevada Territory had only about one-fifth of the 127,381 persons required for statehood, Nevadans were encouraged to seek admission to the Union. Congress accepted the proposed state constitution and voted for statehood in 1864.

Mining and cattle-ranching decades. In its early decades Nevada's economy was dependent on mining and ranching. The rich Comstock mines reached a maximum annual output of $36,000,000 in silver in 1878. During the 1870s, however, the federal government limited the role of silver in the monetary system, causing a decline in silver prices, the closing of many Nevada mines, and the decay of once thriving communities into ghost towns.

As mining declined, cattle ranching became a major industry. Beef prices, however, were unpredictable, high railroad rates were burdensome, and severe winters often killed thousands of cattle. In the late 1880s many cattle ranchers were forced into bankruptcy. Depressed in mining and ranching, the state's population dropped from 62,000 in 1880 to 47,000 in 1890.

Prosperity returned to Nevada only after the beginning of the 20th century, when rich silver ores were discovered near Tonopah and major copper deposits around Ely and when a major gold strike occurred at Goldfield. Thousands of miners answered the lure of these bonanzas, and the railroads built extensive branch lines to bring in equipment to the mining areas and haul out the ore. Accessible railroads and reduced rates also encouraged cattle ranchers to renew large-scale production. Irrigation of fertile river valleys produced sizable hay crops. Thus assured of winter feed, ranchers further expanded their herds in the upland regions. World War I demands for Nevada's beef and metals kept the boom going, but the failing markets of the 1920s brought the return of economic depression.

During its first three decades as a state, Nevada was oriented to Republican control. Reflecting the lax standards in national politics, the state was often manipulated by corrupt politicians. Mine owners and ranchers frequently subsidized government officials, and there were accusations that rich men in the state had bought seats in the U.S. Senate. Monetary issues became of paramount importance in the 1890s, and the Free Silver Party swept four consecutive state elections. By 1900, however, the traditional two-party system was again in control, and since then Nevada has voted consistently with overall national trends.

Creation of a modern economy. Nevada began its transition to a modern economy during the Great Depression of the 1930s. After the legalization of gambling in 1931 and the reduction to six weeks of the residence requirement for divorce, Nevada became a marriage, divorce, and resort centre. The principal resort areas are Las Vegas, Reno, Laughlin, and Lake Tahoe. Las Vegas attracts many tourists from southern California and foreign countries and also hosts business and professional conventions. Reno draws many pleasure seekers from the San Francisco Bay area and from the Pacific Northwest. Laughlin emerged as a tourist centre in the 1980s, and Lake Tahoe continues to serve as a fashionable playground.

Advent of the resort and scientific communities

Construction of Hoover Dam on the Colorado River substantially aided the economy of southern Nevada, and its cheap hydroelectric power opened the way for manufacturing. The importation of hydroelectric power from Bonneville Dam on the Columbia River and piped-in natural gas also has brought industrial development in the northwestern region.

In the 1950s the establishment of the Nevada Test Site by the federal government expanded employment opportunities and stimulated the development of technical industries within the state. Overshadowing the new industrialization, and fundamentally responsible for the current prosperity, is the diversification and expansion of the tourist trade to include not only the gaming and entertainment facilities of the Reno and Las Vegas areas but also the scenic and recreational opportunities statewide. (R.J.Z.)

Utah

A Mountain state, Utah became the 45th member of the Union on Jan. 4, 1896. Utah represents a unique episode in the settlement of the United States, a story of a religious group that trekked and was driven across three-quarters of the continent in search of a "promised land." The state capital, Salt Lake City, is the world headquarters of the Church of Jesus Christ of Latter-day Saints, commonly known as the Mormon church, and the spiritual home of adherents throughout the world. With Mormons making up nearly 70 percent of the state's population, the beliefs and traditions of the Mormon church continue to exert profound influences on many facets of the state's life and institutions.

Before the arrival of the first Mormon pioneers, Utah was inhabited by several Indian tribes including the Ute, for whom the state is named. From the beginning of Mormon settlement in 1847, the pioneers set about wresting a green land from the deserts, gradually supplementing their crops with the products of industry and the earth. The economy

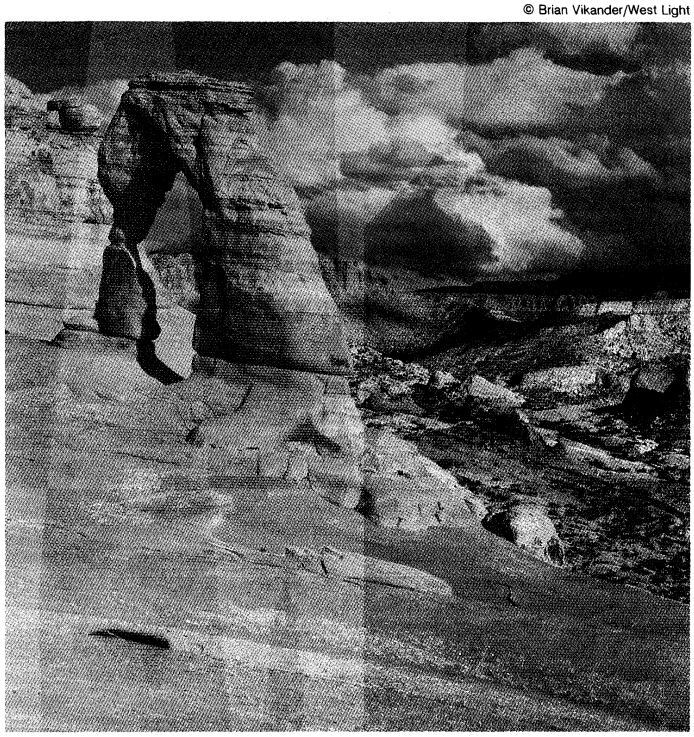
of present-day Utah is based on manufacturing, tourism, and services, in addition to agriculture and mining.

Mountains, high plateaus, and deserts form most of Utah's landscape. The state's 84,899 square miles (219,889 square kilometres) lie in the heart of the West, with Idaho to the north, Wyoming to the northeast, Colorado to the east, Arizona to the south, and Nevada to the west. At Four Corners, in the southeast, Utah meets Colorado, New Mexico, and Arizona at right angles, the only such meeting of states in the nation.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The Colorado Plateau comprises slightly more than half of Utah. Relatively high in elevation, this region is cut by brilliantly coloured canyons. Utah's growing tourist industry relies upon the attraction of the region's fiery, intricately sculptured natural bridges, arches, and other masterpieces of erosion.

© Brian Vikander/West Light



Delicate Arch in Arches National Park, eastern Utah.

Desert and mountain landscapes

The western third of the state is part of the Great Basin of the Basin and Range Province, a broad, flat, desertlike area with occasional mountain peaks. Great Salt Lake lies in the northeastern part of the region: to the southwest of the lake is the Great Salt Lake Desert, covering some 4,000 square miles, which include the Bonneville Salt Flats, famous for land speed racing. During the Pleistocene epoch, from 1,600,000 to 10,000 years ago, the region's huge Lake Bonneville covered an area as large as Lake Michigan. Great Salt Lake, saline Sevier Lake, and Utah Lake are the remnants of Lake Bonneville.

The Middle Rockies in the northeast comprise the Uinta Mountains, the only major mountain range in the United States running in an east–west direction, and the Wasatch Range. Along the latter runs a series of valleys and plateaus known as the Wasatch Front. The Wasatch Range exhibits many glacially formed features such as cirques and moraines. Canyons have been formed by various streams.

Altitudes range from 13,528 feet (4,123 metres) at Kings Peak in the Uintas to about 2,000 feet in the southwestern corner of the state. The Oquirrh and Deep Creek ranges of the Great Basin are important for their deposits of copper, gold, lead, and zinc.

Drainage. Utah contributes to three major drainage areas—the Colorado and Columbia rivers and the Great Basin. The Colorado and its tributary, the Green, drain eastern Utah. The Upper Colorado River Storage Project includes several dams and many lakes in that area. Rivers in the western part of the state include the Bear, Weber, Provo, Jordan, and Sevier, all of which flow into the Great Basin. All of the river systems are important for their irrigation and power potential.

Irrigation and water conservation

Irrigation was among the first Mormon pioneer efforts in 1847, and since then irrigation and water conservation have become increasingly important. The irrigation complex in Utah comprises a number of dams, reservoirs, canals and ditches, pipelines, and flowing wells, exclusive of the large Glen Canyon and Flaming Gorge dams. State boards and departments regulate water use, while the division of health maintains quality water standards under the Water Pollution Control Act of 1953.

Soils. The desert soil that covers most of the state lacks many organic materials but contains lime. Lack of adequate drainage in the Great Basin has damaged surrounding soils with saline materials and alkali salts. The richest soils are in the centre of the state, from the Idaho border almost to Arizona, where most of the farming is done. Mountain soils provide a habitat for conifers and other trees.

Climate. Utah's geographic location in relation to the mountain systems of the West, which divert much of the area's precipitation, makes it basically an arid state. Southwestern Utah, which has a warm, almost dry, subtropical climate, however, is referred to as Utah's "Dixie." The southern part of the Colorado Plateau has cool, dry winters and wet summers, with frequent thunderstorms. Northern Utah is affected by air masses from the North Pacific and continental polar air; it receives most of its precipitation in the cool season.

The state has four distinct seasons. The average temperature in July is about 70° F (21° C). In winter the average temperature is slightly below freezing except in Dixie. Daily temperatures vary widely: when Salt Lake City has July highs of 90° F (32° C) or above, the nights are 55° to 65° F (13° to 18° C). Relatively low humidity prevails; average precipitation is 11 inches (280 millimetres) a year, varying from less than eight inches annually over the Great Salt Lake Desert to 50 inches in the Wasatch Mountains. The average annual snowfall is 4½ feet, ranging from none in the southwestern valleys to more than 10 feet at ski resorts. The average growing season is 131 days.

Plant and animal life. Utah's 4,000 plant species represent six climatic zones, from the arid Lower Sonoran in the southwestern Virgin Valley to the Arctic on mountain peaks. In the south are found creosote bush, mesquite, cactus, yucca, and Joshua tree; the alkaline deserts are the habitat of shad scale, saltbush, and greasewood. Juniper and sagebrush grow in the foothills and mountain valleys, as do piñon pine, cedar, and native grasses for grazing. In the mountains grow pines, firs, aspen, and blue spruce. Timber covers more than 15,000,000 acres (6,000,000 hectares), but only about one-fourth of the forestland is commercially valuable.

Characteristic flora

The mule deer is the most common of Utah's large animals since bison, timber wolves, and grizzly bears have largely disappeared. Coyotes, bobcats, and lynx are hunted. Game birds include grouse, quail, and pheasants; golden eagles, hawks, owls, and magpies are numerous. Great Salt Lake bird refuges are the home of sea gulls, blue herons, and white pelicans. Several species of game fish are native, while others have been introduced. Reptiles and amphibians, both poisonous and nonpoisonous, are native.

Settlement patterns. About two-thirds of Utah's land is federally owned, and 7 percent is owned by the state. About 4 percent is reserved for Indian use.

The Wasatch Front, extending north–south from Ogden to Provo and including Salt Lake City, is the main area of urban and industrial development. Salt Lake City is the political, cultural, and religious capital of Utah. Historically a trade centre, it continues to be a hub for industry, commerce, and interstate transportation.

The Front has not only the largest part of the population but also the best farmland in the state. Although tens of thousands of acres of cropland have been urbanized since 1958 and an urban trend continues, a rural society is still observable. Rural settlements typically have a "Mormon village" flavour, with a readily recognizable Mormon

chapel or tabernacle within the town, wide streets, and a cultivated area surrounding the town itself.

The people. The population is about 95 percent white, mainly of northern European ancestry. The remainder are Hispanics, Indians, Asians, blacks, and other minorities. Except for Indians, nearly 80 percent of the minority population lives in the three Wasatch Front counties of Salt Lake, Davis, and Weber.

The population of San Juan county is about one-half Indian, containing almost 30 percent of Utah's Indians. These are mostly Navajo, who reside primarily in the Four Corners region of the Navajo Indian Reservation. The Ute live on the Uintah and Ouray Indian Reservation. Annually sponsored events include the bear dance in the spring, the sun dance in July, and the Uinta Basin Industrial Fair in August or September. A number of Southern Paiutes, among the most economically depressed of the tribes, live on several small reservations in southern Utah.

Traditional Indian cultural activities

People of Hispanic origin constitute the state's largest minority group. Increasing attention is being paid to the problems of educating and acculturating this group, many of whom are low-income workers in agriculture, mining, manufacturing, and services.

Although Mormons represent 90 percent of all religious adherents in Utah, Roman Catholics can be found throughout the state. Baptists, Lutherans, and other Protestant denominations are also represented.

The economy. From 1847 to 1868 the Mormons built a self-sufficient economy based on agriculture, handicrafts, and small industry. From 1869 to 1896 this cooperative economy was supplemented by a non-Mormon enclave devoted to mining and trading. After statehood the exportable resources of the state were exploited to an increasing extent by outside corporations and enterprisers, and the agriculture of the state turned toward range cattle, wool, and such commercial crops as sugar beets. The economic depressions of 1921 and of the 1930s were severe, but federal programs and the welfare program of the Mormon church helped the state to recover. During World War II several defense plants and air bases were built, and Utah had a uranium boom. In the late 1950s several large plants were erected to build rocket engines for missiles.

The state's economy is highly diversified. The agricultural and mining sectors have been supplemented by light and heavy manufacturing, finance, transportation, and tourism. Salt Lake City is a regional centre of finance and trade, and many large enterprises have offices there.

Mining. A fair percentage of the nation's new copper is produced annually in Utah, but a decline in world copper prices in the 1980s brought Utah's copper production facilities to a standstill by 1985. Reorganization of the industry has allowed some production to continue. Utah is the world's foremost producer of beryllium, and it is a major producer of gold, silver, lead, uranium, and molybdenum. Salt (sodium chloride) was once the only mineral extracted in quantity from the Great Salt Lake, but sophisticated chemical industries now operate on the shores of the lake, using its brines to also produce magnesium, potassium sulfate, and sodium sulfate for industrial use throughout the world.

Mineral reserves and sources of income

Utah is a major producer of coal west of the Mississippi, and it is the only state producing Gilsonite, a source of road oil, paving binder, and asphalt tile. In addition to steam plants, Utah has many hydroelectric plants.

Agriculture. Following the national trend, farm employment and the number of farms in Utah have declined since 1960, but productivity has increased. Almost three-fourths of Utah's farm income comes from livestock products, the remainder from field crops, fruit, and canning crops.

Industry. Business and military services and state and federal government employment continue to increase at a faster rate than other sectors of the economy. The proportion of personal income derived from manufacturing is below average, however. Printing and publishing, food processing, petroleum refining, and the production of transportation equipment, computer hardware and software, nonelectrical machinery, rocket engines, and fabricated-metal products are the major manufacturing sectors.

Transportation. Utah's transportation industry, with easy access to all national markets, is the basis for the state's development as a major distribution centre for the West. Although railway mileage has decreased, road traffic has expanded; several interstate highways supplement the state system. In addition to the international airport serving Salt Lake City, there are excellent feeder line facilities in Ogden, Logan, Provo, Cedar City, and St. George.

Administration and social conditions. *Government.* Utah's constitution, dating from statehood, guarantees basic personal freedoms consistent with the federal Bill of Rights, prohibits sectarian control of public schools, forbids "polygamous or plural marriages," and grants equal civil, political, and religious rights, including suffrage, to all citizens. Voting requirements follow national patterns, though for elections affecting tax levies a voter must have paid a property tax the previous year.

Constitutional framework of the government

The governor is aided by a lieutenant governor (who also performs the duties of a secretary of state), auditor, treasurer, and attorney general, while much of the administration of routine state affairs is done through more than 50 state agencies. Each of these officials serves a four-year term. The governor has the right to veto any bill, but that decision may be overruled through repassage of the bill by a two-thirds majority of each house of the legislature. Any bill passed by the legislature and not acted upon by the governor within 10 days while the legislature is in session automatically becomes law. The governor, lieutenant governor, and attorney general together form the State Board of Examiners, which reviews all official state transactions.

Legislative power is vested in the Senate and House of Representatives, as well as in the voters, who have the power to initiate legislation and to hold a referendum on all laws not passed by a two-thirds majority of both houses. The legislature consists of 29 senators serving four-year terms and 75 representatives serving two-year terms.

The legislature meets annually in 45-day sessions. Special sessions may be called by the governor. Four councils provide investigation and research of specific legislative and state problems; advice on budgetary matters and appropriation requests; and legislative administration.

The highest judicial authority is the state Supreme Court, composed of five justices elected to 10-year terms, one every two years. Judges of the seven district courts are elected for six-year terms. The state also has circuit courts and justices of the peace. A juvenile court system has its own districts and judges.

All of Utah's counties are political subdivisions of the state and carry out administrative, judicial, law enforcement, financial, health, educational, and welfare functions assigned by the state and federal governments. All but one of the counties are governed by the traditional three-member commission form of government. The other, Cache County, has an elected executive with part-time council members who perform judicial and policy-making functions. Counties perform municipal-type services in unincorporated areas as citizens demand, and they perform other services demanded or requested by citizens and permitted or not prohibited by state statutes.

Forms of municipal government vary according to population. Salt Lake City, the only city with a population of more than 90,000, elects a mayor and city council. Cities between 15,000 and 90,000 elect a mayor and two commissioners, while smaller cities elect a mayor and five council members. Incorporated towns are governed by a president and four trustees. Any city commission or town council has the power to appoint a city manager.

Although Utah is referred to frequently as a Republican state, actually no party can claim dominance. Elected officials from both parties work well together and show a reasonable degree of harmony. This has been true since the early 1890s, when the normally homogeneous Mormon populace was divided into political parties by church leaders to comply with federal requirements for statehood.

Utah's broadly based tax structure appears to distribute the costs of government among all segments of the economy. The corporate income tax rate is lower than that of most Western states. A liberal free-port tax law granting tax exemptions on goods warehoused and processed in Utah is an incentive to commerce.

Education. More than half of Utah's governmental expenditure is for education. Utah has the highest proportion of its population in public schools, the highest proportion of high school graduates, and the highest median level of school years completed of any state in the nation.

The school districts levy taxes that pay for almost half of educational expenses, the remainder being paid by the state. General public school regulations are administered by the state Board of Education; elected local boards exercise more specific control. There is a growing number of private elementary and secondary schools.

Higher education

The largest of Utah's state universities is the University of Utah, in Salt Lake City. It was founded in 1850 as the University of Deseret and has a reputation for outstanding graduate and professional schools of medicine, law, and pharmacology. Utah State University, in Logan, founded in 1888 as a land-grant school, has achieved national status in the fields of agriculture, forestry, education, engineering science, upper-atmosphere research, and the fine arts. Weber State College (1889), in Ogden, and Southern Utah State College (1897), in Cedar City, are schools with rapidly expanding programs and facilities. Dixie College (1911), in St. George, and College of Eastern Utah (1938), in Price, are state junior colleges. Community colleges offering technical and other courses are located in Salt Lake City and Provo.

Brigham Young University, in Provo, is operated by the Mormon church. It is the largest church-related university in the nation, with 12 colleges and professional schools. Westminster College of Salt Lake City (1875) is operated by three Protestant denominations.

Health and welfare. The state, county, and local governments have developed programs to improve the economic and social status of minority groups.

Health, welfare, and housing services are administered by the Department of Social Services. County health services are supervised and coordinated by the state Board of Health, which also works with school boards for child health care. Outstanding hospital systems are administered by independent health organizations and by the Roman Catholic and Episcopal churches.

The state's welfare program includes comprehensive old-age assistance, unemployment insurance, worker's compensation, and other social benefits. Efforts have been made to improve and upgrade outdated labour and hazardous-occupation laws. A division of low-income housing within the Department of Community Affairs facilitates better planning and coordination in that area. The Mormon church also has an extensive welfare program.

Cultural life. Because the population of Utah is overwhelmingly Mormon, the church has a strong influence on the state's cultural life and traditions. The church is divided into "stakes," consisting of six to 10 local congregations, or "wards," of about 500 members each. Each stake has a "tabernacle," or stake centre, with one or more chapels and recreational and cultural facilities. Each ward, or occasionally two or three wards together, owns a chapel with a centre for collective worship, classrooms, a basketball court, and a dance hall. Mormon culture emphasizes closely knit family life, widespread interest in family genealogy, prohibitions against consumption of alcoholic beverages and use of tobacco, a relatively small amount of nightlife, and participation in sports and personal-development programs. Other denominations also are active in cultural areas. Particularly notable is the annual St. Marks Arts Festival, which includes music, dancing, poetry reading, drama, and other creative arts.

Impact of Mormonism on the cultural milieu

The Utah State Historical Society has a collection of manuscripts, publications, and photographs and publishes the *Utah Historical Quarterly,* monographs, and full-volume diaries. The National Society Daughters of Utah Pioneers, with about 1,000 branches throughout the nation, maintains a museum and monuments, preserves old landmarks, marks historical sites, keeps a library of historical matter, and collects data and relics to document the lives of the Utah pioneers. *The Western Historical Quarterly,* the official publication of the Western History Association, is published by Utah State University.

Park City and Pioneer Village in Salt Lake City are Old West towns containing original buildings and furnishings. Every county holds a fair in the autumn, highlighted by displays and competitions, concessions, and often a rodeo.

On July 24 almost all communities hold Pioneer Day, commemorating the entrance of the Mormon pioneers into the Salt Lake valley. It includes parades, fireworks, rodeos, orations, and reminders of Utah's early settlers.

The Division of Fine Arts, founded in 1899, is the oldest state arts agency in the nation. Its purpose is to promote all branches of the fine arts. It sponsors the Utah Governor's Conference on the Arts, and it allocates funds advanced to the state by federal agencies.

The most famous buildings in Utah are the many-spired Mormon Temple and the turtleback Mormon Tabernacle, both in Salt Lake City. The latter was built in the 1860s. It holds up to 8,000 people and has rare acoustical qualities that enrich the sounds of its world-famous organ with some 10,700 pipes. There are also notable Mormon temples in St. George, Manti, Provo, South Jordan, Ogden, and Logan.

Music, dance, and theatre

Among the performing arts, music is emphasized. The Mormon Tabernacle Choir, 325 members with trained but nonprofessional voices, presents concerts and national weekly radio and television broadcasts. The Utah Symphony Orchestra, Utah Opera Company, and Salt Lake Oratorio Society are other major ensembles. The major universities have symphonies and choral groups performing in winter, as well as summer festivals and concerts.

Dance companies in Utah include Ballet West, which features classical ballet, and the Repertory Dance Theatre, which features modern dance. The University of Utah Children's Dance Theatre and the Brigham Young University folk-dance troupes are well known.

Utah gained an early start in drama with the opening of the Salt Lake Theater in 1862. A replica has been constructed on the University of Utah campus, and performances are held there regularly. The Mormon church emphasizes folk drama in its youth organization; more than 2,000 wards produce at least one play a year, many of them written locally. These culminate biennially in a large drama festival in Salt Lake City. The annual Utah Shakespearean Festival is held in Cedar City.

Salt Lake City has professional basketball, hockey, and baseball teams, and auto racing on the Bonneville Salt Flats has gained international importance. The Mormon church sponsors competitive team sports involving thousands of players, with basketball, softball, and golf tournaments that are among the largest in the nation.

Utah's nine national forests and other undeveloped areas offer great tracts of land for hunting, fishing, camping, hiking, skiing, and snowmobiling. Other natural attractions include the national parks (Arches, Bryce Canyon, Capitol Reef, Zion, and Canyonlands), the national monuments (Cedar Breaks, Dinosaur, Natural Bridges, Timpanogos Cave, Rainbow Bridge, and Hovenweep), the national recreation areas (Flaming Gorge and Glen Canyon), and the Golden Spike National Historic Site. There are 45 state parks, including Pioneer Trail State Park at Salt Lake City.

HISTORY

Prehistory and white exploration. As early as 10,000 BC, small groups of hunters and gatherers lived in caves by the great inland sea, prehistoric Lake Bonneville. This desert culture was replaced about AD 400 by the more advanced Pueblo, or Anasazi, culture, which came into Utah from what is now the Southwest and Mexico. These Indians constructed superb communal cliff dwellings and raised corn (maize), squash, and beans. They left Utah about 1250, perhaps because of an extended drought.

Early inhabitants and cultures

When white explorers and settlers came to Utah in the 18th and 19th centuries, they encountered Shoshoni Indians—the Southern Paiute, Gosiute, and Ute—some of whom raised corn and pumpkins by irrigation. The Ute in eastern Utah lived in a region of higher rainfall; having acquired horses from Plains Indians, their nomadic life centred around the buffalo.

Two Franciscan fathers, Francisco Atanasio Domínguez and Silvestre Vélez de Escalante, explored Utah in 1776, and afterward Utah was visited by occasional Spanish

trading parties. Fur trappers and immigrants to California and Oregon were in the region in the 1820s and '30s. The first four of some 16 annual rendezvous between trappers and buyers were held in Utah from 1825 to 1828, indicating the early importance of the area to the fur trade. The "mountain men" who explored and established trading posts included James Bridger, who first visited the Great Salt Lake in 1824, and Jedediah Smith, who first traversed the state from north to south and west to east in 1826–27. Explorers sent by the government included John C. Frémont, who led expeditions to northern Utah in 1843 and the western Great Salt Lake area in 1845.

Mormon settlement and territorial growth. The period of settlement and territorial status is notable for the ending of the quest (1845–47) for a Mormon homeland, wrestling with an arid environment, the contest for sovereignty between Utah and the United States, and the conflict with indigenous Indians over the use of the land.

When wagonloads of Mormon pioneers under the leadership of Brigham Young first entered the valley of the Great Salt Lake in 1847, they were determined to transform the arid valley land into a green and wholesome "kingdom of God." From Salt Lake City (until 1868 called Great Salt Lake City) settlers were directed to colonize in all directions until they had developed a prosperous and stable economy and political structure in a territory that was originally 210,000 square miles in area, stretching from the Rockies to the Sierra Nevada and from the Columbia River in Oregon to the Gila in Arizona. Immigrant converts continued to stream into Utah from Europe and the eastern United States; they were organized into colonizing parties based on allocations of skills and leadership abilities and sent out to build the territory. By 1860 more than 150 self-sustaining communities with a total of 40,000 residents had been established, irrigating crops with water from mountain streams carried through canals to the alluvial valley lands. Utah's place in the national scene was symbolized by the driving of the golden spike at Promontory in 1869, uniting the eastward- and westward-reaching lines of the nation's first transcontinental railroad.

Conflict with the Indians was held to a minimum until the colonizers became more and more ubiquitous and local Indians began to raid the settlements. The Ute were eventually placed on a reservation in the Uinta Basin, the Southern Paiute and Shoshoni on smaller reservations, and later the lands south of the San Juan River were incorporated into the Navajo Reservation.

Mormons' conflicts with federal government

The propensity of the Mormons to establish their own political and social system and the incompetency of federal territorial officials led to an era of conflict with the federal government. In 1857 President James Buchanan, believing the Mormons to be in a state of open rebellion, ordered some 2,500 soldiers to Utah to replace Young, who had served as governor during the early years. This episode is referred to as the Utah War, although no armed clashes occurred. With the outbreak of the American Civil War in 1861, a new camp was established east of Salt Lake City under the command of Colonel Patrick Connor. Connor openly supported his troops in prospecting for minerals and sought to "solve the Mormon problem" by initiating a miners' rush to Utah. A substantial enclave of non-Mormon miners, freighters, bankers, and businessmen arrived, and there ensued three decades of conflict between Mormons and non-Mormons.

Before and after statehood. The Mormon settlers applied for statehood in 1849 under the name Deseret, a word from the sacred *Book of Mormon* meaning "honeybee" and signifying industry. This bid was rejected, as were the efforts of five subsequent constitutional conventions between 1856 and 1887. Before the U.S. Congress and the national administration would assent to statehood for Utah, Mormon leaders were required to discontinue the church's involvement in politics through its People's Party, withdraw from an economic policy in which Mormons dealt primarily with each other, and discontinue the practice of polygamy.

After its acceptance into the Union in 1896, Utah moved rapidly into the mainstream of the nation. The political structure changed from theocracy to a conventional democracy: non-Mormons were elected to important positions, including the office of governor. The Mormon church has been officially neutral in politics since the early 1900s, and the influence of economic blocs has become far more important. That Utah's traditional spirit of community-mindedness survives was illustrated by the thousands who joined, on short notice, to counteract the devastation of heavy spring flooding in Salt Lake City in 1983. (L.J.A.)

Wyoming

Wyoming is the ninth largest U.S. state, with an area of 97,809 square miles (253,326 square kilometres). It was admitted to the Union as the 44th state on July 10, 1890. The state's name is derived from a Delaware Indian word meaning "land of vast plains," an apt description of Wyoming's spacious natural environment, which is home to nearly as many pronghorn as people. The state's residents are spread across the land in small ranching and farming towns, in mining settlements, and in communities offering unparalleled outdoor recreational opportunities. Each year millions of people visit Yellowstone and Grand Teton national parks. Cheyenne, the state capital, is located in Wyoming's southeastern corner.

Wyoming shares boundaries with six other Great Plains and Mountain states: South Dakota and Nebraska on the east, Colorado on the south, Utah on the southwest, Idaho on the west, and Montana on the northwest and north. Tens of thousands of pioneers crossed Wyoming along the Oregon, Overland, Mormon, Bozeman, and Bridger trails during the 19th century. The route of the short-lived Pony Express crossed the state along the Oregon Trail in 1860–61, as did the tracks of the Union Pacific Railroad when they first connected North America's east and west coasts in the late 1860s.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Wyoming's topography is dominated by several large basins and the mountain ranges of the Rockies that border them. The broad basins are synclines, while the mountains dominating Wyoming's horizon were formed during a period of mountain-building activity known as the Laramide orogeny, which affected the region from 70,000,000 to 40,000,000 years ago. The land surface of Wyoming has a mean elevation of 6,700 feet (2,040 metres) above sea level, the highest in the United States after Colorado. Three-quarters of the state lies more than a mile (1,609 metres) high, and 40 percent exceeds 7,000 feet (2,100 metres) in elevation. Wyoming's lowest point of 3,125 feet (953 metres) lies in the channel of the Bell Fourche River as it flows from the state into South Dakota, and its highest point is Gannett Peak (13,804 feet [4,207 metres]) of the Wind River Range in west central Wyoming.

Basins and mountains

Wyoming has six physiographic regions: the Black Hills; Great Plains; Southern, Middle, and Northern Rocky Mountains; and Wyoming Basin. The Black Hills extend into the state from South Dakota and are of generally low relief. Wyoming's Great Plains region occupies the easternmost one-third of the state, gradually increasing in elevation from the state's eastern border to the many mountain ranges that mark the region's western margin.

The Southern Rocky Mountains extend from northeastern Colorado along the Laramie, Medicine Bow, and Sierra Madre ranges, making their farthest extension into Wyoming along the Laramie Range, where the mountain system terminates just south of the North Platte River near the city of Casper. The Northern Rocky Mountain region extends south from Canada across Montana and Idaho and enters Wyoming at the northwestern corner of Yellowstone Park. The much larger Middle Rocky Mountain region occupies most of the northwestern quarter of the state, extending south along the Idaho–Wyoming border into Utah. Included in this region are the scenic Big Horn and Wind River mountain ranges, the geysers and fumaroles of Yellowstone Park, the igneous Absaroka Plateau on the park's eastern margins, and Gannett Peak.

Wyoming's Rocky Mountains

The Wyoming Basin is composed of interspersed smaller

mountains and intermontane basins and is located between the Southern and Middle Rocky Mountains. This region includes Flaming Gorge, created by the Green River, and the Great Divide Basin that encloses an area of interior drainage with no outlet.

Drainage. The Continental Divide crosses Wyoming from the south central portion of the state, trending northwest and leaving the state through Yellowstone National Park. Partly because of the presence of the divide, Wyoming contributes to the headwaters of four major North American drainage systems—the Colorado, Columbia, and Missouri rivers and the Great Salt Lake. The most significant of these to the state is the Missouri system, which drains approximately three-fourths of Wyoming's land area. It is estimated that 75 percent of the state's contributions to these drainage systems originates as snowmelt in Wyoming's mountain ranges.

Soils. Wyoming's several hundred soil types may be grouped into three broad categories determined largely on the basis of the state's variable elevation and climatic zones. Varieties of mountain soils are found throughout the many ranges in Wyoming, with their greatest concentration in the northwest. These soils are frequently acidic and of limited value to commercial cropping, although they may support alpine meadows used for summer pasture and scattered forests used for timber products. The southwestern to north central portions of the state contain numerous varieties of desert soils that are frequently alkaline and used mostly for winter range, although others are suitable for agricultural crops when irrigated. Plains soils, found in the eastern third of Wyoming, are of reasonable fertility and support moderate levels of dryland farming, including the production of wheat.

Climate. Wyoming's climate is influenced by its interior location on the North American landmass (a condition termed continentality) and by its high mean elevation. The state includes areas of arid desert, semiarid steppe (short-grass prairie), and alpine climates. The arid desert regions are all found in the western half of the state and have average annual rates of precipitation of between four and eight inches (100 and 200 millimetres). More than 70 percent of the state is considered semiarid steppe and averages nine to 16 inches (230 to 410 millimetres) annually. Wyoming's mountains may receive much larger quantities of precipitation. In some mountainous areas total snowfall can exceed 200 inches (5,100 millimetres) annually and remain on the ground more than 150 days per year.

Average monthly temperatures vary greatly across Wyoming. January mean temperatures range from a low of 10° F (−12° C) in the mountains to 28° F (−2° C) in the southeast. Mean July temperatures range from 50° F (10° C) in the mountains to 75° F (24° C) in the Big Horn Basin in north central Wyoming.

Plant and animal life. Approximately 80 percent of Wyoming is covered with grasses and semidesert to desert shrubs. The state's forests are found largely in the mountains and along streams where sufficient soil moisture is available. Though there are some limited areas of hardwood trees, most of Wyoming's forests are composed of conifers, principally ponderosa pine in the northeast, lodgepole pine in the south central area, and Douglas fir, Engelmann spruce, and lodgepole pine in the northwest.

Pronghorn

Wyoming supports abundant animal life, including the largest number of pronghorn found anywhere. Pronghorn are located in every part of the state, with their greatest concentration in areas of sagebrush and grasses. The state also supports large numbers of whitetail and mule deer and lesser populations of wapiti (American elk) and moose. Black bears live in most of Wyoming's forested mountain areas, with grizzly bears in the high mountain and wilderness areas in and surrounding Yellowstone Park. Herds of American bison (buffalo) are found in Yellowstone and Grand Teton national parks. Grouse, quail, partridge, and pheasant are found in some of the state's uplands, and wild turkeys are common in many of the state's open woodlands.

Settlement patterns. Wyoming's earliest pattern of sedentary occupancy by Europeans was determined by the locations of military posts such as Fort Laramie (1834–90) and Fort Bridger (1843–90). The building of the Union Pacific Railroad in the late 1860s led to the founding of several early settlements, including Cheyenne, Laramie, Rawlins, Rock Springs, and Evanston.

Wyoming's current pattern of settlement is based upon its agricultural, mining, and recreational activities. There is no major metropolitan area, and the state's two largest urban areas, Casper and Cheyenne, are small cities by the standards of most states. The remainder of the state's towns and cities are typically small in population, having long expanses of Wyoming's wide open spaces between them. They are service centres for surrounding ranches and farms, mining operations, and recreational lands.

The people. More than 95 percent of Wyoming's residents are Caucasian. Although most of the state's Caucasian residents trace their roots directly to Europe, Mexican-Americans now account for about 5 percent of Wyoming's population. Blacks constitute less than 1 percent of the total population, most living in the Cheyenne area. Although Chinese immigrants were instrumental in the construction of the Union Pacific Railroad, Wyoming's Asian population today is small and numbers fewer than 2,000. Most Asians live in the state's southern counties in the cities of Cheyenne, Laramie, and Rock Springs. Nearly 2 percent of Wyoming's population is composed of American Indians, mostly of the Arapaho and Shoshoni tribes. More than half of the Indian population lives on the 3,500-square-mile (9,065-square-kilometre) Wind River Reservation in the west central portion of Wyoming.

The economy. Wyoming's economy is heavily tied to mining and agriculture (primarily the marketing of beef cattle and sheep). The state also has an important and growing tourist industry, serving the hundreds of thousands of visitors to the state's parks and historic sites. Manufacturing is of only minor importance.

Mining. Wyoming is one of the top coal-producing states. Coal ranked as Wyoming's most valuable mineral resource prior to 1920 but is now third behind the state's oil and natural gas reserves. Of the oil and gas deposits found across the state, the largest known deposits are those in the northeast. There are also significant oil shale reserves in the southwest.

Uranium and other ores

Wyoming has substantial uranium deposits, estimated to account for one-third of the total reserves in the United States. The state's largest uranium deposits are found in the Red Desert, Shirley Basin, South Powder River basin, Gas Hills, and Pumpkin Buttes areas. Wyoming also contains quantities of trona (unrefined soda ash), bentonite clay (used as drilling mud), gypsum, limestone, and iron ore. Because of a rise in gold prices in the mid-1980s, gold exploration has increased in Wyoming, especially in the southern tip of the Wind River Range.

Agriculture. As Wyoming's nickname, the Cowboy State, implies, ranching has historically been important to the state both economically and culturally. The state's rangelands are well suited to livestock production, and approximately 70 percent of the state's land area is devoted to livestock grazing. Wyoming is responsible for the production of much of the sheep, lambs, and beef cattle in the United States. The principal breed of sheep is Rambouillet; the principal breeds of cattle are Angus and Hereford.

The major crop-producing areas in Wyoming are in the southeast and in the Big Horn and Wind River basins. Wyoming's most valuable grain crop is wheat; other important crops include oats, barley, hay, and corn (maize). Wyoming is a major producer of sugar beets, dry beans (including great northern and pinto beans), and potatoes.

Tourism. Tourism and recreation are major growth industries in Wyoming. They make a substantial contribution to the state's economy and account for approximately 10 percent of the total employment. The state government has increased its advertising of Wyoming's spectacular scenery and recreational opportunities. Among the principal sites for tourists are the state's parks and historic sites, Yellowstone and Grand Teton national parks and such attractions as Big Horn Canyon National Recreation Area, Fort Laramie National Historic Site, and Devil's Tower and Fossil Buttes national monuments.

Major east–west routes

Transportation. The original path of the transcontinental railroad still serves as one of Wyoming's major transportation corridors. The tracks of the Union Pacific Railroad, which continue to carry substantial quantities of freight across the state, now share the corridor with Interstate 80, one of the country's most important east–west highways. A second east–west transportation corridor is Interstate 90. Wyoming's primary north–south transportation corridor is Interstate 25. The state is served by a network of paved highways and roads that include the scenic Yellowstone Highway, which connects Wyoming's largest city, Casper, with Yellowstone National Park.

There is no passenger rail service in Wyoming, but commuter air carriers serve the state's major cities and recreational destinations, such as Cheyenne, Laramie, Casper, Cody, and Jackson Hole. Most commuter air service operations originate in Denver, Colo., or Salt Lake City, Utah. Additional flights are scheduled during the winter to serve skiing destinations such as Jackson Hole.

Administration and social conditions. *Government.* Wyoming's constitution, adopted in 1889, specifies three branches of state government: a legislature, an executive, and a judiciary. There are five elected executives—the governor, auditor, treasurer, superintendent of public instruction, and secretary of state; there is no lieutenant governor. All executive officers serve four-year terms. Each of the five elected state administrators supervises an area of state government with a substantial degree of autonomy.

Wyoming's constitution specifies a bicameral legislature, including a House of Representatives and a Senate. The Senate has 30 members who are elected for four-year terms, while the House has 64 representatives who are elected every two years. Representation in both chambers is based upon county or district populations. Wyoming's legislature is composed of part-time citizen lawmakers who meet for limited legislative sessions each year.

Wyoming's constitution also establishes a three-tier court system that includes local courts, nine district courts, and a Supreme Court. District court justices stand for reelection every six years. The state's Supreme Court has five justices who stand for reelection every eight years in nonpartisan elections. Local courts include country courts and justices of the peace and municipal courts.

Local government

At the local government level there are 23 counties and numerous municipalities, school districts, and special districts. The form of municipal government is by local option, the strong and weak mayor and the manager forms all being used. All counties use the commission form of government. The average county government provides services to a sparse population spread over 4,226 square miles, an area more than twice as large as the state of Delaware. Sweetwater county, in the southern portion of the state, alone accounts for 10,429 square miles. These large areas require a strong commitment to the effective provision of services on the part of local government officials and a measure of self-reliance on the part of Wyoming's population.

Wyoming is politically conservative and has traditionally favoured the Republican Party in presidential contests. Wyoming has also sent a greater number of Republican senators and representatives to the U.S. Congress. Although Democratic victories are not uncommon in state executive-branch positions, there has not been a Democratic majority in the state Senate since 1936 or in the House since 1964.

Education. Compared with national averages, Wyoming's schools are small and have favourable teacher–pupil ratios. In the state's most isolated areas, kindergarten to eighth-grade students are still taught in one-teacher schools.

The University of Wyoming, located in Laramie, was founded in 1886 and is the state's only public four-year institution of higher learning. The university offers undergraduate study in more than 100 fields and has a variety of graduate degree programs. Wyoming also maintains two-year community colleges in Casper, Cheyenne, Sheridan, Powell, Rock Springs, Torrington, and Riverton.

Health and welfare. Wyoming has unparalleled outdoor recreational opportunities, a low rate of crime, and little pollution. The state's population is well educated, with a proportion of both high school and college graduates above the national average. Personal income per capita in Wyoming is variable because of the state's dependence on mineral extraction but generally is among the highest in the Rocky Mountain region. Quality health-care facilities are located in the state's larger towns and cities, but there remains a demand for resident health-care professionals in many of the state's rural areas.

Cultural life. Western traditions and culture remain very much a part of Wyoming life. Annual festivals that celebrate the state's Western heritage include county fairs, the Wyoming State Fair, held in Douglas each August, and Jubilee Days, held in Laramie in July. Many of these events are held in conjunction with rodeos. The world's largest rodeo is held each summer in Cheyenne during Frontier Days. Frontier Days has been held annually since 1897 and draws visitors from all parts of the world to watch events such as bronc riding, bull riding, calf roping, and barrel racing.

Music and theatre

Music companies and music festivals include the Cheyenne Symphony Orchestra, Casper Civic Symphony, Grand Teton Music Festival in Teton Village, and the Wyoming Summer Music Festival in Laramie, which highlights chamber music. Several cities and towns in Wyoming have active theatre companies. There are also a number of museums in Wyoming, many preserving the state's colourful historical past. Every county in Wyoming has at least one library, with the state's largest being the William Robertson Coe Library, at the University of Wyoming in Laramie.

Wyoming's numerous state and national parks provide nearly unparalleled opportunities for camping, hiking, and observing wildlife. Nearly the entire expanse of the world's oldest national park, Yellowstone, is found within the borders of Wyoming, making it readily accessible to the state's residents. A large number of Wyomingites also regularly take advantage of the state's excellent hunting and fishing opportunities.

HISTORY

Prehistory to white exploration. The first occupants of Wyoming were Paleo-Indian hunters and gatherers who arrived from Siberia through Alaska more than 20,000 years ago. The total number of these peoples was never large because they were highly dependent upon local game populations. By the time the first well-documented visits

© James Fain

Calf roping at the Cody Stampede, a rodeo held annually in Cody, Wyo.

by white men to Wyoming occurred, the state's population likely did not exceed 10,000. The Shoshoni were the largest tribal group in Wyoming around 1800, but there were also smaller numbers of Arapaho, Crow, Cheyenne, and Oglala and Brulé Dakota (Sioux).

First European visitors

The first known white men to enter Wyoming were the French-Canadian brothers Francois and Louis-Joseph, sons of Pierre Gaultier de Varennes, Lord de La Vérendrye. The brothers visited the northeastern corner of the state in 1743 while unsuccessfully searching for a route to the Pacific Ocean. Although the Lewis and Clark Expedition (1804–06) missed Wyoming by 60 miles, a member of the group, John Colter, broke away from the main party and trapped in northern Wyoming for some time; the official journal of the expedition includes Colter's route and descriptions of the Jackson Hole and Yellowstone Park areas.

Fur trade and the Union Pacific Railroad. The early explorers were followed later by small numbers of fur traders. Although there were likely never more than 500 of these mountain men in Wyoming at any given time, the state's economy between 1825 and 1840 was heavily dependent on the activities of such famous trappers and traders as Jim Bridger, William Sublette, Jedediah Smith, and Thomas Fitzpatrick.

The number of people entering the Wyoming area increased with the westward movement of the American population. As many as 400,000 emigrants crossed Wyoming between 1841 and 1868 on the trails leading to what is now Oregon, Washington, Montana, Utah, and California. In 1850 alone it is estimated that as many as 55,000 crossed the future state. Pony Express riders, including Buffalo Bill Cody, carried the mail across Wyoming between April 1860 and October 1861.

In November 1867 the first train of the Union Pacific Railroad reached Cheyenne and opened Wyoming as never before. Cheyenne grew from a handful of people to more than 6,000 in the first year, though the town consisted largely of tents and shacks with a limited number of commercial buildings. This rapid population growth continued in southern Wyoming as the Union Pacific tracks continued across the state, finally entering Utah in 1868. The building of the railroad focused attention on the West, and the Wyoming Territory was created on July 25, 1868.

The state. The state's constitution was approved by a vote of the territorial population on Nov. 5, 1889, although Wyoming was not admitted to the Union until 1890. Wyoming's constitution was the first in the world to grant full voting rights to women. Wyoming was also the first state to elect a woman governor when Nellie Tayloe Ross won the position in 1924. Because of these developments Wyoming has been called the Equality State.

Development of the cattle industry

In the years preceding statehood Wyoming developed its thriving cattle industry. The state's immense rangelands fostered the initiation of the cowboy era that was chronicled in Owen Wister's *The Virginian* (1902), based on his experience in turn-of-the-century Wyoming. Although frequently exaggerated by Hollywood, this era was marked by violence on the range between cattlemen, homesteaders, and sheepherders that continued well after 1900.

Although Wyoming retains its Western heritage and personality, employment in the state is now more characterized by mining than by the cowboy life. The state's reliance on the energy industries of coal, oil, natural gas, and uranium has made Wyoming subject to "boom-and-bust" cycles that depend on world prices for its products. During the energy boom of the 1970s, for example, the state's population grew at nearly four times the national rate and had one of the highest incomes per capita in the country. The world oil supply glut of the 1980s, on the other hand, caused a substantial downturn in the state's economy that led to significant population out-migration. The state is making an effort to diversify its economy in such areas as tourism, but there is little doubt that Wyoming's long-term economic future is tied to mining. (G.R.We.)

THE PACIFIC COAST

Alaska

When it became the 49th state of the United States on Jan. 3, 1959, Alaska increased the nation's size by nearly 20 percent. The new area included vast stretches of unexplored land and untapped resources. When Secretary of State William H. Seward negotiated its purchase from Russia in 1867, however, Alaska was known as Seward's Folly. Its settlement and exploitation have been hindered by its distance from the rest of the nation and by geographic and climatic impediments to travel and communications; Alaska continues to be the country's last frontier. More than half of the state's inhabitants live in the Greater Anchorage area. The capital is Juneau, 573 miles (922 kilometres) to the southeast in the panhandle region.

Alaska lies at the extreme northwest of the North American continent and is the largest peninsula in the Western Hemisphere. Its 591,004 square miles (1,530,700 square kilometres) include some 15,000 square miles of fjords and inlets, and its three faces to the sea have about 34,000 miles (54,400 kilometres) of indented tidal coastline and 6,600 total miles of coast fronting the open sea. It borders the Arctic Ocean on the north and northwest, the Bering Strait and the Bering Sea on the west, and the Pacific Ocean and Gulf of Alaska on the south. The land boundaries on the east cut across some 1,150 miles of high mountains to separate the state from the Canadian Yukon Territory and British Columbia province. Rimming the state on the south is one of the Earth's most active earthquake belts. In the Alaska Range north of Anchorage, Mount McKinley, at 20,320 feet (6,194 metres), is the highest peak in North America.

The question of development versus preservation has been heightened by commercial and ecological uses of land: the Alaska Highway gas-pipeline project, native Alaskans' land claims, noncommercial whaling by native peoples, and related matters. The conflicts between conservationists and petroleum companies over the Trans-Alaska Pipeline, which runs from the oil-rich North Slope on the Arctic Ocean to Valdez in the south, was a continuation of the century-long effort to find a balance between conservation and development in this enormous land.

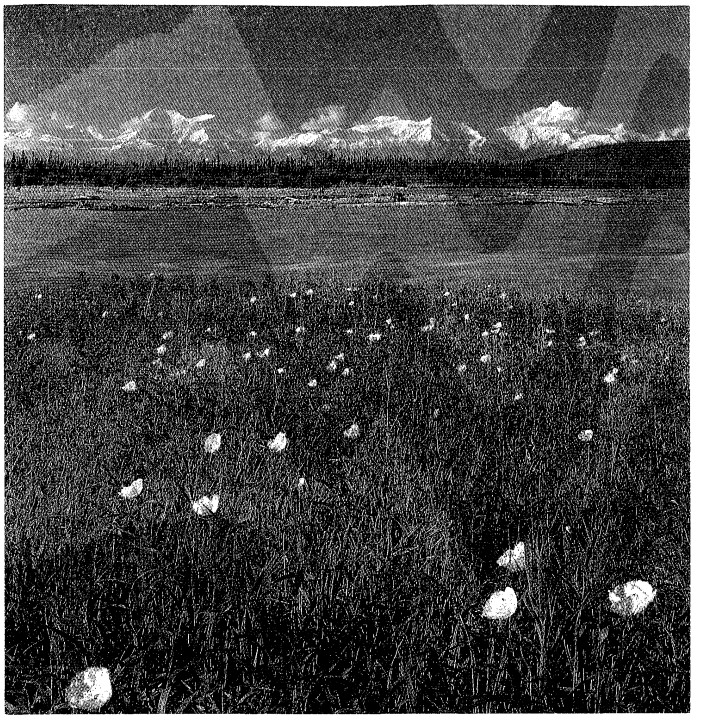
PHYSICAL AND HUMAN GEOGRAPHY

The land. The immense area of Alaska has a great variety of physical characteristics. Nearly one-third of the state lies within the Arctic Circle and has perennially frozen ground (permafrost) and treeless tundra. The southern coast and the panhandle at sea level are fully temperate regions. In these latter and in the adjoining Canadian areas, however, lies the world's largest expanse of glacial ice outside Greenland and Antarctica. Off the extreme western end of the Seward Peninsula, Little Diomede Island, part of Alaska, lies in the Bering Strait only 2.5 miles (four kilometres) from Soviet-owned Big Diomede Island; both countries have shown a tacit tolerance of unintentional airspace violations, which are common in bad weather.

Relief. Alaska is composed of nine distinct physiographic and environmental regions. Much of the mainland panhandle region, a narrow strip of land 25 to 50 miles wide lying east and south of the St. Elias Mountains, is composed of the Boundary Ranges. There are several large icefields, and the peaks include Mount St. Elias (18,009 feet), from whose summit the Alaska–Yukon border swings due north. The western extension of this mountain chain is the Chugach Range, a giant arc at the northernmost edge of the Gulf of Alaska. Many remote valleys and high ridges are still unexplored, and the relief and glaciation inhibit exploitation. The coast is characterized by frequent and intense oceanic storm systems that have produced dense rain forests on the coastal mountain flanks. In the valleys rivers produce devastating annual floods often associated with excessive snowmelt and glacial meltwaters.

The panhandle region

The region of the south coastal archipelago and the Gulf



Alaska Range seen from across the Tanana River.
E. Cooper—H. Armstrong Roberts

of Alaska islands includes the Alexander Archipelago in the panhandle region, with 11,000 islands, plus Kodiak Island and its satellites south of Cook Inlet. These islands, extensions of the southern region, are lower, less rugged, and less glaciated. All receive heavy rain and are affected by waters warmed by the Kuroshio Current.

The Aleutian region includes the narrow Alaska Peninsula, which forms the south shoreline of Bristol Bay, and the 1,100-mile-long Aleutian chain that separates the North Pacific from the Bering Sea. The chain includes 14 large islands, 55 significant but smaller ones, and thousands of islets. The largest are Unimak, Unalaska, and Umnak. On the occasionally clear summer days, active volcanoes and such glacier-covered peaks as symmetrical Shishaldin Volcano (9,372 feet [2,857 metres]) on Unimak can be seen. Such magnificent views represent the Aleutians at their scenic best. Usually, however, the weather is wet and stormy, the winds horizontal and cutting, and the fog all-pervading.

The broad Alaska Range region connects the Aleutian Range across the southern third of mainland Alaska to the Wrangell Mountains, which abut against the vast complex of the St. Elias Mountains. The Wrangell Mountains have large active volcanoes and high valley glaciers. The flanks of this subarctic range are largely tundra-covered.

The low-lying interior basin region between the Alaska Range in the north and the Chugach–Wrangell–St. Elias mountains to the south and east enjoys a relatively temperate climate. The lower valleys contain good farmlands, and it is there that most of the people of Alaska live.

The central plains and tablelands of interior Alaska constitute a vast region west and north of the Alaska Range; they reach as far north as the Brooks Range. The area is rolling and dissected by numerous streams tributary to the Yukon and Kuskokwim rivers. The plains extend from the Canadian border to Norton Sound, the Seward Peninsula, the Yukon delta, and south to the northern rim of Bristol Bay on the Bering Sea. The region is characterized by river flats and truncated upland tablelands. With abundant game, it is an important nesting ground for waterfowl, including great numbers of migrating birds.

Brooks Range

A major mountain chain running west to east in the area north of the central plains and extending from the sea nearly to the Yukon border, the Brooks Range gradually slopes northward to a narrow linear coastal plain bordering the Arctic Ocean and westward to lower hills north of Kotzebue Sound. There are a few high Arctic glaciers, and the area is semiarid. The lower flanks and valleys are tundra-covered, with permafrost features.

The coastal lowland north of the Brooks Range, sometimes called the North Slope, is the home of great herds of caribou. The environment is truly polar, with the sea waters along the coast frozen eight months of the year and the ground permanently frozen except for a thin zone of summer melting. It is treeless, and, in summer, grasses and Arctic alpine flowers abound. The National Petroleum Reserve-Alaska is located in the western sector, while the Prudhoe Bay oil fields and part of the Arctic National Wildlife Refuge occupy the eastern sector.

The islands of the Bering Sea represent a small but unique Arctic maritime environment, typified by St. Lawrence, Nunivak, and St. Matthew islands and the Pribilof group. These tundra-covered islands are surrounded by sea ice in winter and serve as protected refuges for the world's largest herds of fur-bearing seals and sea otters, as well as sea lions and walrus. A large herd of domesticated reindeer is tended by Eskimos on Nunivak Island.

Climates. Five general climatic zones may be delineated in Alaska, excluding the great mountain ranges.

Southern coastal and southeastern Alaska, the Gulf of Alaska islands, and the Aleutians have average temperature ranges in the summer of 40° to 60° F (4° to 16° C) and in the winter of 40° to 20° F (4° to −7° C). Rainfall varies locally from 60 to 160 inches (1,525 to 4,065 millimetres), and the panhandle and southern islands are covered with Sitka spruce, hemlock, and other evergreens. The Cordova–Valdez region and parts of the west central panhandle have the state's highest precipitation, 220 inches or more. At Valdez 200 inches of snow is not uncommon. Precipitation is less in the Aleutians, but even there about 250 rainy days occur annually.

The interior basin ranges from 45° to 75° F (7° to 24° C) in summer and 20° to −10° F (−7° to −23° C) in winter. The region is drier than the coast and only slightly colder in winter, with Anchorage receiving about 15 inches (380 millimetres) of precipitation annually. The pleasant conditions and proximity to the sea have helped to make the area the centre of the state's population.

The islands and coast of the Bering Sea have summer temperatures of 40° to 60° F (4° to 16° C) and winter temperatures of 20° to −10° F (−7° to −23° C). Tempering influences of the Pacific dissipate north of the Pribilof Islands, and Arctic sea ice often reaches this area.

The central plains and uplands range from 45° to 75° F (7° to 24° C) in the summer and −10° to −30° F (−23° to −34° C) in the winter. Average rainfall is 10 to 20 inches, though less than 10 inches is common.

The ameliorating effects of the Arctic Ocean keep temperatures of the North Slope at 35° to 55° F (2° to 13° C) in the summer and −5° to −20° F (−21° to −29° C) in the winter—less severe than those of the interior plains. About five inches of precipitation nonetheless remain on the ground as snow for some eight months of the year. The 24-hour sunlight of summer can produce strong buildups of radiant energy, sending temperatures to 90° F (32° C). The deep chill of winter, however, maintains the permafrost character of the High Arctic zone. Ice clogs the northern coast nine months of the year, while ice fog frequently extends southward to Fairbanks.

Settlement patterns. A large percentage of Alaskans live in the southern interior basins around Anchorage; most of the remainder live in the interior plains around Fairbanks or in the panhandle region, where Juneau is the major city and the administrative centre of the state. Tiny pockets of people are scattered in small villages, the most sparsely occupied being the Arctic plains, the Bering shores, and the Aleutians. Many frontier conditions persist: a male-to-female ratio of 5 to 1 in 1910 has been reduced to near equality, but in many places bars are as numerous as churches.

Population distribution

The people. English, Russian, Spanish, and French place-names reflect early European exploration, but equally prominent are dozens of names from the pre-Western era. The name Alaska itself is derived from the Aleut *alaska* and the Eskimo *ālakshak,* both meaning "mainland."

Native population and cultures

Long before Bering's voyages the Tlingit Indians lived in the southern and southeastern coastal area; the Aleuts on the Aleutian Islands and the Alaska Peninsula; and the Eskimos on the Bering shore and the Arctic Ocean coast. The interior natives were the Tinneh Indians, whose language was Athabascan, that of the Plains Indians of the interior continent to the south. The Indian groups are presumably descendants of the earliest immigrants across the Bering Land Bridge from Asia, perhaps more than 15,000 years ago, and they reflect the migratory wave that reached as far as the southern extremity of South America as early as 10,000 years ago. Eskimos and Aleuts appear to be much later immigrants, having arrived, probably in boats made of animal skins, perhaps 8,000 to 3,000 years ago. All groups have been involved in the debates and adjudications over public land grants.

The first wave of immigration from the "South 48"—which occurred in the decade before World War I as an aftermath of the gold rush—was a response to Alaska's initial concentration on its mineral, fish, and timber resources. The discovery of oil fields and the emergence of Alaska as an international air crossroads added impetus to the influx of the 1940s and '50s and construction of the Trans-Alaska Pipeline to that of the 1960s and '70s. By 1980 only about 20 percent of the white population of the state was born in Alaska.

Of the current population about one-seventh are Eskimos, Aleuts, and Indians. The remaining citizenry include military personnel and their families and a melting pot of mixed American, Russian, Filipino, Japanese, Chinese, and other nationalities.

The economy. The Alaskan economy is conditioned strongly by the state's frontier stage of development, but its formerly inadequate tax base for state and municipal growth ended with the development of the North Slope oil fields. High costs of labour and transportation and complicated environmental and land-use constraints still tend to discourage outside investment. Nonetheless, development of the state's natural resources has assisted markedly in the transition from a federal military to a commercial self-supporting economic base.

Government. Alaska's economy has been dominated by government since territorial days. From 1940 to 1960 the federal government invested nearly $2,000,000,000 in the development of military bases in Alaska. Nothing else in Alaska's history has produced such long-term results, bringing thousands of residents into the territory and creating jobs and a vast array of transportation and communications facilities extending to remote corners of the state. Combined with state and local government, the defense installations continue to add much to Alaska's economy.

Agriculture. Only a small sector of Alaska's economy is agricultural, but a viable in-state market is still under development. More than 3,000,000 acres (1,200,000 hectares) of tillable land are available for farming, but much clearing has yet to be done. Most acreage is near Anchorage and on the Kenai Peninsula, though there is some near Fairbanks, and stock ranching is practiced on Kodiak and Unimak islands. As a result, all farm products are sold locally and most foods must be imported, tremendously increasing the cost of living. Closure of the Homestead Act, ending settlement of the native land claims issue, has further curtailed development of new land. In spite of a short growing season, the long hours of summer sunlight are adapted to the production of oats, barley, potatoes, hay, and cool-climate vegetables. Livestock and greenhouse crops are also successful.

Salmon industry

Fishing, forestry, and furs. Alaska's most constant source of revenue is derived from fishing. Fish are found mostly in waters off the southern coasts, salmon being of especial importance. The centre of the world's salmon-packing industry is at Ketchikan, on Kodiak Island, and at Bristol Bay ports in the southern Bering Sea. Fleets also bring in quantities of herring, cod, pollack, and halibut, as well as Dungeness, king, and Tanner crabs. International fishing of Alaska's waters is regulated by the 200-mile-wide exclusive economic zone and the U.S.-Canadian Pacific Salmon Treaty (1985), which assigns ownership of fish to the country in which they spawn.

Most of Alaska's timber resources are in the Tongass and Chugach national forests, in the panhandle and on the southern coast, respectively. Timber is produced mainly for export to Asia, with the pulp of Ketchikan and Sitka exported to Japan.

Pribilof sealskins represent more than half of the state's annual fur production. Other furs, largely from controlled farms, are processed as well. The production of reindeer hides from a herd on Nunivak Island is managed by the Alaska Native Association.

Power. Alaska's immense waterpower reserve is virtually untapped. The largest project is at Lake Eklutna, near Anchorage. A hydroelectric development near Juneau delivers power to the panhandle area, and another project on the Kenai Peninsula is planned to deliver power to the central and southern regions. Most other communities depend on diesel and coal plants to produce much of the required municipal power.

Mining. Petroleum was first extracted and refined between 1917 and 1933, but the development of the Kenai oil field in 1961 made the petroleum and natural gas industry Alaska's most important mineral production. Oil seeps were known as early as the 1880s in the North Slope region, which today has become a field of major economic importance to both the state and the nation. Alaska ranks second only to Texas in oil production.

Since 1880 hard-rock ore minerals have been mined in Alaska, about 95 percent of which yield gold, copper, zinc, and silver. Prospecting continues, with modern scientific technology and aerial exploration. The areas of maximum mineral potential lie in the panhandle, the Chugach and Alaska ranges, and the Seward Peninsula at locations unlimited by environmental regulation.

Alaska's gold production declined drastically after World War II, but the mining of gold especially and of zinc, silver, and lead began to rebound in the 1980s. Copper mining as a major industry ended with the closing of the Kennecott Mine in 1938, although there are new prospects elsewhere. Coal has remained an important industry. Another important activity is the extraction of sand, gravel, and clay to serve the construction industry.

Tourism. Alaska has had an upsurge of tourism. Travelers arrive mainly by air or sea and can now cover large areas by airplane and road. The influx is partly the result of the 500-passenger, 100-car ferries that operate as the Alaska Marine Highway. One ferry system connects Kodiak with mainland Seward and the Alaska Railroad, another links Cordova and Valdez, and a third serves panhandle communities from Ketchikan to Skagway, with service also from Prince Rupert, British Columbia, and Seattle, Wash.

Transportation networks

Transportation. High costs of transportation continue to sap Alaska's economic development, largely because the major transportation links, both internal and external, are by air, which provides the fastest way to cross Alaska's great distances and formidable terrain. Two dozen airlines serve Alaska, with daily service for passengers and cargo from the South 48 and Canada, Europe, Hawaii, Korea, and Japan. Some 800 airfields, seaplane bases, and emergency strips are in use, and few villages are without service at least by bush pilots. Most of the state's roads are surfaced. The Alaska Highway and its Haines and Skagway cutoffs connect Alaska's internal road network to the outside and provide relatively easy access for tourists. A 416-mile (669-kilometre) haul road from Fairbanks to Prudhoe Bay connects with the existing highway system to provide an overland route from the ice-free southern ports to the Arctic Ocean. The public, however, is restricted to the southern half of this highway and may use it only in the summer.

The government-owned Alaska Railroad runs for about 500 miles (800 kilometres), linking Seward, Anchorage, and Fairbanks. Ocean shipping connects Seattle, Vancouver, and the trans-Canada railhead of Prince Rupert to towns in the panhandle and westward to Cordova, Valdez, Seward, and Kodiak. Ocean vessels also run during the ice-free midsummer months to Nome and Barrow and to the oil regions of the Arctic coast. A natural gas pipeline connects the Kenai gas fields and Anchorage, and the

Trans-Alaska Pipeline delivers North Slope oil to ice-free tanker terminals at Valdez.

In the mid-1950s the Alaska Communication Cable was installed between Seattle and Alaska. Radio telephones connect all interior communities.

Administration and social conditions. *Government.* The state constitution was adopted in 1956. The governor and lieutenant governor are the only executive officers and are elected for four-year terms. The 40-member House of Representatives and 20-member Senate are elected for terms of two and four years, respectively. The Supreme Court has a chief justice and four associate justices. A three-member court of appeals was established in 1980. There are four district courts. A single federal district court sits alternately in Juneau, Anchorage, Fairbanks, and Nome.

Public financing is implemented through various personal income, property, sales, and business taxes, including petroleum-based severance taxes and mining rents and royalties. As a part of the Act of Admission, Congress granted Alaska certain revenues from the sale of furs and of federal lands.

State and borough governments have difficulty in providing the usual range of services because of the limited extent of the economy and a high unemployment rate. The vast area and the difficult terrain increase these problems.

Native welfare

The U.S. Bureau of Indian Affairs (BIA) assists Alaska's natives in achieving economic and social self-sufficiency. Despite a number of helpful programs, many of Alaska's natives suffer from unemployment, low income, and poverty. The native peoples were educated first by missionary groups, though by the time of statehood the BIA had assumed most of the responsibility for education. Funds are provided for vocational training and the development of job opportunities and for welfare, social work, and medical and health needs. The BIA also assists natives in organizing their villages under federal and state laws. Some oil revenues from native lands have been applied in self-help programs. Settlement of the native land claims in 1971 improved their economic plight by placing 44,000,000 acres of federal land into the native entitlement.

Education. Education is compulsory through the eighth grade or until age 16 and is administered by a state board and a commissioner of education. Correspondence study is available for high school work through the State Department of Education. There are several federal schools on military bases. The University of Alaska, founded as a land-grant institution in 1917, operates campuses at Fairbanks, Anchorage, and Juneau. There are several community colleges, including those at Sitka, Ketchikan, Kenai, and Valdez. Alaska Pacific University in Anchorage, Alaska Bible College in Glennallen, and Sheldon Jackson College in Sitka are private institutions.

Health and welfare. The elderly, dependent children, and the blind are aided by the state, and a special fund benefits sick and disabled fishermen. The state also operates a psychiatric hospital, a tuberculosis treatment centre, a youth camp, and a prison.

Medical and health clinics and hospitals available to the general public are provided by municipal and borough governments or private agencies, or are run as church-operated facilities. Health standards have been raised markedly since 1950 through visits by U.S. Public Health Service nurses and doctors to the remote villages. The large number of airfields, the radio communications network, and the extensive use of bush pilots operating throughout the state make it possible for most persons, even in the remote villages, to reach medical facilities when there is serious need. There are modern hospitals located in Fairbanks, Anchorage, Juneau, and Ketchikan.

Cultural life. Alaska's past, including the arts and crafts of its native peoples, is a major influence in Alaskan culture. Juneau is the site of the state's historical library and state museum. The university has a large museum, as do other communities, including Sitka, Haines, Valdez, and Nome. Eminent Alaskan artists have included both whites and Eskimos. Native ivory and wood carvings are well known, and the nearly lost art of totem carving has been revived in part through private and public stimulus.

Wildlife refuges and ranges abound throughout Alaska, with more than 77,000,000 acres managed by the U.S. Fish and Wildlife Service. The federal Bureau of Land Management also holds about 25,000,000 acres for water-power development.

In 1980, more than 104,000,000 acres were designated for national parks, preserves, wildlife refuges, and wilderness areas, adding to the 7,500,000 already so established. The Alaskan national parks are notably spectacular. Denali (formerly Mount McKinley) National Park and Preserve (1917) has an abundance of wildlife, including brown and grizzly bears, caribou, and moose. Katmai National Park and Preserve (1918), on the Alaska Peninsula, includes the Valley of Ten Thousand Smokes, an area of active volcanoes that in 1912 produced one of the world's most violent eruptions. Glacier Bay National Park and Preserve (1925) has magnificent fjords, as well as glaciers that have retreated extensively in the 20th century. Sitka National Historic Park (1910), with a large totem pole collection, commemorates the stand of the Tlingits against early Russian settlers. The Tongass and Chugach national forests in the southeast and south central regions, respectively, are also federal public land reserves. The U.S. Department of the Interior has continued to study the need for withdrawing further regions from public domain into reserves.

National parks and wilderness preserves

The sporting industry, including guide and outfitter services and boat charters, continues to be a colourful activity. Alaska provides the nation's only significant Arctic wilderness, and much research is done in glacier, mountain, tundra, atmospheric, ionospheric, and polar oceanography fields by federal, state, university, and private agencies. For example, the University of Alaska carries out extensive research on Arctic problems through its Geophysical Institute, Institute of Marine Science, Institute of Arctic Biology, and other groups. Since 1946 the Juneau affiliate of the Foundation for Glacier and Environmental Research, in cooperation with the National Science Foundation, the University of Idaho, and the University of Alaska, has sponsored a glaciologic and environmental research and field sciences training program on the Juneau Icefield.

HISTORY

Explorations. As early as 1700, native peoples of Siberia reported the existence of a huge piece of land lying due east. An expedition appointed by the Russian tsar and led by a Danish mariner, Vitus Bering, in 1728 determined that the new land was not linked to the Russian mainland, but because of fog it failed to locate North America. On Bering's second voyage, in 1741, the peak of Mount St. Elias was sighted, and men were sent ashore. Sea otter furs taken back to Russia opened a rich fur commerce between Europe, Asia, and the North American Pacific Coast during the ensuing century.

Russian pioneering

Early settlement. The first European settlement was established in 1784 by Russians at Three Saints Bay, near present-day Kodiak. It served as Alaska's capital until 1806, when the Russian-American Company, organized in 1799 under charter from the emperor Paul I, moved its headquarters to richer sea otter grounds in the Alexander Archipelago at Sitka. The company governed Alaska until its purchase by the United States in 1867. Alaska's first governor (then termed chief manager), Aleksandr Baranov, was an aggressive administrator whose severe treatment of the native Indians and Eskimos led in 1802 to a massacre at Sitka.

A period of bitter competition among Russian, British, and American fur traders was resolved in 1824 when Russia granted equal trade rights for all. The near extinction of the sea otter and the political consequences of the Crimean War (1853–56) were factors in Russia's willingness to sell Alaska to the United States. The Russian minister made a formal proposal in 1867, and, after much public opposition, the purchase was approved by the U.S. Congress and the U.S. flag was flown at Sitka on Oct. 18, 1867.

Political growth. As a U.S. possession Alaska was governed by military commanders for the War Department until 1877. During these years there was little internal development, but a salmon cannery built in 1878 was the beginning of what became the largest salmon industry in the world. In 1884 Congress established Alaska as a judi-

cial land district, federal district courts were established, and a school system was initiated.

Establishment of the territory

In 1906 the first representative to Congress, a nonvoting delegate, was elected, and in 1912 Congress established the Territory of Alaska, with an elected territorial legislature. Alaskans voted in favour of statehood in 1946 and adopted a constitution in 1955. Congressional approval of the Alaska statehood bill in 1958 was followed by formal entry into the Union in 1959.

Mining booms. Other significant events in Alaska's history included early gold discoveries on the Stikine River in 1861, at Juneau in 1880, and on Fortymile Creek in 1886, and later the stampede to the Atlin and Klondike placer goldfields of adjoining British Columbia and Yukon Territory in 1897–1900. Gold discoveries followed at Nome in 1898 and at Fairbanks in 1903. The gold rush made Americans aware of the economic potential of this previously neglected land. The great hard-rock gold mines in the panhandle were developed, and in 1898 copper was discovered at McCarthy. Gold dredging in the Tanana River valley was begun in 1903 and continued until 1967.

Economic growth. A dispute between the United States and Canada over the boundary between British Columbia and the Alaska panhandle was decided by an Alaska Boundary Tribunal in 1903. The U.S. view that the border should lie along the crest of the Boundary Ranges was accepted and boundary mapping was completed in 1913. Between 1898 and 1900 a narrow-gauge railroad was built across White Pass to link Skagway and Whitehorse in the Yukon, and shortly afterward the Cordova-to-McCarthy line was laid up the Copper River. Another railway milestone, and the only one of these still operating, was the 538-mile Alaska Railroad connecting Seward with Anchorage and Fairbanks in 1923. In 1935 the government encouraged a farming program in the Matanuska valley near Anchorage, and dairy herds and crop farming became established there, as well as in the Tanana and Homer regions.

In 1942, during World War II, Japanese forces invaded Agattu, Attu, and Kiska islands in the Aleutian chain and bombed Dutch Harbor on Unalaska. This aggression prompted the construction of large airfields as well as the Alaska Highway linking Dawson Creek, British Columbia, and Fairbanks with more than 1,500 miles (2,400 kilometres) of road. Both proved later to be of immense value in the commercial development of the state.

During the 20th century nearly 40 earthquakes measuring at least 7.25 on the Richter scale have been recorded in Alaska. The devastating earthquake on March 27, 1964 (8.4 on the Richter scale), affected the northwestern panhandle and the Cook Inlet areas, destroying parts of Anchorage; a tsunami that followed wiped out Valdez; the coast sank 32 feet (9.75 metres) at Kodiak and Seward; and a 16-foot coastal rise destroyed the harbour at Cordova.

Discovery of oil

Oil and natural gas discoveries in the Kenai Peninsula and offshore drilling in Cook Inlet in the 1950s created an industry that by the 1970s ranked first in the state's mineral production. In the early 1960s a pulp industry began to utilize the forest resources of the panhandle. Major paper-pulp mills were constructed at Ketchikan and Sitka, largely to serve the Japanese market. The discoveries in 1968 of petroleum on lands fronting the Arctic Ocean gave promise of relief for Alaska's economic lag, but problems of transportation across the state and to the South 48 held up exploitation of the finds. In 1969 a group of petroleum companies paid the state nearly $1,000,000,000 in oil-land revenues, but the proposed pipeline across the eastern Brooks Range, interior plains, and southern ranges to Valdez created heated controversies among industry, government, and conservationists. In November 1973 a bill passed the U.S. Congress that made possible construction of the pipeline, which began in the following year. The completed 48-inch (122-centimetre) pipeline, 789 miles (1,262 kilometres) long, came into operation on June 20, 1977. As a result, oil flows freely from the Prudhoe Bay oil field on the Arctic coast to the ice-free harbour at Valdez, whence tankers transport it to U.S. West Coast ports. Further development of Alaska's petroleum reserves depends upon economic factors and the issue of high production costs in the hostile Arctic environment. In 1989 the oil tanker *Exxon Valdez* ran off course in Prince William Sound, causing the most disastrous oil spill in North American history and inflicting incalculable damage on the area's marine ecology and local economy. (M.M.M.)

California

California is the most populous state in the Union, and its personal income per capita is one of the highest in the world. The fluid nature of the state's social, economic, and political life, shaped so largely by the influx of people from other states, gives California the aura of a laboratory for testing new modes of living. Yet, by the 1980s, a profound disaffection with growth had set in throughout the state's urban areas.

Californians make up the most urban population in the nation, centred mainly along the coast, with more than three-fourths of its people living in the Los Angeles, San Francisco, and San Diego metropolitan areas. As in most of the nation's larger states, the capital, Sacramento, is not a major population or economic centre.

California has an area of 158,706 square miles (411,049 square kilometres), exceeded only by Alaska and Texas. The state is bounded on the north by Oregon, on the east by Nevada and Arizona, on the south by the Mexican state of Baja ("Lower") California, and on the west by the Pacific Ocean.

No version of the origin of California's name has been fully accepted, but there is wide support for the contention that it derived from a Spanish novel that described a paradisiacal island called California. Admitted to the Union on Sept. 9, 1850, as the 31st state, California is a land of stunning physical contrasts: from the rainy northern coast to the parched Colorado Desert of the south. The Sierra Nevada exceeds the Rocky Mountains in height. Within 85 miles (137 kilometres) of each other lie Mount Whitney and Death Valley, respectively 14,494 and 282 feet (4,418 and 86 metres) above and below sea level, the highest and lowest points in the 48 coterminous states. Despite its urbanization, California is also the principal agricultural state of the nation, though only about 15 percent of its area is cultivated. Almost half of its land is federally owned, with national parks and monuments in every part of the state devoted to irreplaceable forest, desert, mountain, and other natural resources.

The state's physical contrasts

Fred Lyon

Vineyards in the Napa Valley, northern California.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Coast ranges.* The long coastline of California is mountainous, most dramatically in the Santa Lucia Range south of San Francisco, where the homes of Big Sur perch on cliffs 800 feet above the sea. Hills of lesser height flank entrances to the coast's three major natural harbours, at San Diego, San Francisco, and Eureka. Coastal mountains, made up of many indistinct chains, are from 20 to 40 miles in width and from 2,000 to 8,000 feet in height.

The mountains of California

Sierra Nevada. The eastern portions of the state are occupied by sparsely settled desert. The Sierra Nevada rises to the west of this desert. The eastern slope is sheer, dropping 10,000 feet within 10 miles near Owens Lake. On the west the range slopes to the Central Valley, comprising the San Joaquin and Sacramento valleys, in gradually declining foothills. From the wall that rises near Lassen Peak in the north, the Sierra Nevada extends south for 430 miles to the fringes of Los Angeles. It is 50 to 80 miles in width and 27,000 square miles in area. Aside from Mount Whitney, 10 other peaks exceed 14,000 feet in altitude. East–west passes are few and high, some at more than 9,000 feet.

The largest lake of the Sierra Nevada is Lake Tahoe, astride the California–Nevada border at 6,229 feet (1,899 metres). A mountain-ringed alpine lake about 193 square miles in area, it ranks 11th in the world in average depth: the 1,200-foot line runs near shore, and the maximum depth exceeds 1,600 feet (488 metres). Elsewhere in the Sierra lie hundreds of smaller lakes, some above the timberline in regions of tumbled granite and smooth-walled canyons. There are three national parks in these highlands: Kings Canyon, Sequoia, and Yosemite—the latter rising from the purplish foothills of the Mother Lode gold country through ice-carved valleys of the Merced and Tuolumne rivers, with their waterfalls and granite domes.

Central Valley. The Central Valley runs for 450 miles (720 kilometres) through the centre of California, forming a trough between the Coast Ranges to its west and the Sierra Nevada to its east. The valley is the state's agricultural heartland. Its single opening is the delta through which the Sacramento and San Joaquin rivers drain into San Francisco Bay. The valley is sealed off at the northeast by the Cascade Range and at the northwest by the Klamath Mountains. This far north the terrain is rugged and sparsely populated, heavily timbered and wet on its coastal side and drier and barren in the higher northeast. In the south the Central Valley is closed off by the transverse ranges, notably the Tehachapi Mountains, which are regarded as a dividing wall between southern and central California.

Deserts. Most of eastern California is desert. The northeastern corner of the state is high desert, a jumble of barren plains and mountains and lava country, sparsely settled and sustaining little but open-range ranching. In the east central region is the Trans-Sierra desert, which extends along the sheer east escarpment of the Sierra Nevada range and comprises part of the vast interstate Great Basin of the Basin and Range Province. The Trans-Sierra desert ranges from 2,000 to 7,400 feet above sea level. Its largest towns are in the Owens Valley, a fertile farmland until its groundwater flow was diverted to Los Angeles through a mammoth series of conduits built in 1908–13.

The Mojave Desert

In the southeast lies the Mojave Desert, which occupies one-sixth of the land area of California, more than 25,000 square miles. Its landmarks are broad basins and eroded mountains, fault blocks, and alluvial surfaces, most lying more than 2,000 feet above sea level. Vegetation includes the evergreen creosote bush, yucca, saltbush, burroweed, encelia, cottonwood, and mesquite. Higher up are juniper and piñon pine. The clear skies and sparse settlement of the Mojave Desert have made it an ideal setting for aviation and ordnance testing. Long airstrips at Palmdale and Edwards Air Force Base are important for the testing of new aircraft and for projects of the U.S. space program. Most of the Mojave's population lives in small towns and military bases on the desert's southwestern edge.

The lower Colorado Desert begins in the Coachella Valley, where Palm Springs has spawned a series of resorts. This desert descends to the Imperial Valley; adjacent to the Mexican border, the valley is an agricultural area known for its winter crops. More than 4,000 square miles of this desert lie below sea level, including the 300-square-mile Salton Sea, a lake with no outlet, created in 1905–07 when the nearby Colorado River broke out of its channel.

Coastal settlement and forests. Southern California's dense settlement lies along a coastal plateau and in valleys ranging from 40 to 60 miles inland. North of the Tehachapis population becomes sparser along the coast. The populous coastal area around San Francisco Bay gives way to the less developed northern coast, where lumbering and fishing villages lie beside creeks and rivers flowing from the Coast Ranges. This is the area of coastal redwood forests and Redwood National Park. These trees, among the tallest in the world, may reach 300 feet (90 metres) in height and are 15 to 35 feet in diameter; some are 4,000 years old. Before European settlement the redwoods covered an estimated 1,500,000 acres (607,000 hectares) of California. Most redwood forests have been cut, but 109,000 acres of redwoods are protected in state and national parks.

The redwood forests

Drainage and water resources. Water is chronically scarce in southern California and the desert regions, but excesses of rain and snowmelt cause winter flooding along the rivers of the northern coast. More than 70 inches of rain falls at Crescent City, near the Oregon border; and the village of Honeydew, in the redwoods of Humboldt county, has received as much as 174 inches in a year. Southern California, in contrast, has most of the state's population but only about 2 percent of its water, with an average annual rainfall of 15 inches, entirely in the winter months. Complex systems of dams and aqueducts move water from north to south, but not without the protests of those who regard the export of water from their regions as a bar to future growth or as a threat to environmental balance. The Colorado River Aqueduct moves water from the river, at the Arizona border, across the southern California desert and mountains to serve the Los Angeles metropolitan area. The California State Water Project, launched in 1960, is the largest water-transfer system ever undertaken. It is designed to deliver water daily from the Feather River to communities as far south as the Mexican border.

Climate. The climate of the entire Pacific Coast is milder and more uniform in temperature than that of the states in corresponding latitudes east of the mountains. A mean annual temperature as low as that of Halifax, N.S. (latitude 44° 39′ N), is not found at any Pacific Coast point south of Sitka, Alaska (latitude 57° N), while the mean at San Diego is 6° to 7° F (3.3° to 3.9° C) less than that at Vicksburg, Miss., and Charleston, S.C., in roughly the same latitude. The means of winter and summer are very near the yearly mean. This condition is not so marked inward from the coast; yet everywhere, save in the high mountains, the winters are comparatively mild.

Division of the year into two seasons—a wet one and a dry one—marks this portion of the Pacific Coast in the most decided manner, being truly characteristic neither of Baja California nor of the greater part of Oregon, though more so of Nevada and Arizona. Except on the coast, the dryness of the air and the consequent rapidity of evaporation greatly lessen the disagreeableness of summer heat.

Range of temperatures and rainfall

Summer temperatures in the low-lying Colorado Desert, reaching as high as 130° F (54° C), compare with the highest of Africa's Sahara. Annual rainfall in this desert averages three to four inches (75 to 1,000 millimetres). In the higher eastern deserts of California, however, summer temperatures are moderate. Winter temperatures atop the Sierra Nevada are arctic. Rainfall ranges from 174 inches (4,420 millimetres) in the northwest to traces in the southeastern desert, but moderate temperatures and rainfall prevail along the coast. The average annual temperature is about 65° F (18° C) in Los Angeles and 57° F (14° C) in San Francisco. Annual precipitation averages about 14 inches in Los Angeles and 21 inches in San Francisco.

The moderate climate has been a major factor in the concentration of settlement along the coast, where cool ocean breezes hold off heat, and temperatures seldom

exceed 90° F (32° C) or drop to freezing. Low humidity usually prevails. Climate changes rapidly with the altitude extremes of California, and the coastal cities are only hours away from mountain skiing or desert sports.

Animal life. California's animal life is as varied as its geography; about 400 species of mammals and 600 species of birds are identified. The California grizzly bear is extinct, however, and the rare bighorn sheep are found mostly in remote desert mountains. The endangered California condor is being nurtured with the help of zoo hatching programs. Coyotes and pumas (cougars) prowl remote mountain areas, while the more common deer, bobcats, and black bears are found in various habitats. In drier areas and deserts there are squirrels, jackrabbits, and chipmunks. Even the hottest deserts support desert tortoises, horned toads, and rattlesnakes. Among common birds are California jays and thrashers, juncos, mountain bluebirds, and hermit thrushes. Trout and salmon are important species that run in California streams and rivers, and the yield from the state's shores include bass, perch, rockfish, and tuna.

The people. *Ethnic distribution.* The influence of Spanish settlers of the 18th and 19th centuries is evident in California's architecture and place-names. The many Californians with Spanish surnames largely reflect the 20th-century immigration from Mexico—to escape that nation's revolution (1910–17) or to find agricultural jobs.

It was the Spanish missionaries who both converted and subjugated the California Indians in the construction of the California mission chain. When the missions were secularized in 1833, some 30,000 Indians were farming under the direction of priests and soldiers at 21 missions. Disease decimated the Indian population for decades after the Spaniards' arrival. Only in the mid-20th century did the California Indian population again begin to increase. Of the Indian population of some 200,000, about 6 percent live on reservations.

The first settlers from the United States were mostly Midwestern farmers of Anglo-Saxon descent. With the gold rush a more cosmopolitan mix appeared. Ships sailed into San Francisco from the Atlantic Seaboard, Europe, and the Orient. In 1850 more than half of the Californians were in their 20s, typically male and single. Only a few hundred Chinese lived in the state in 1850, but two years later one resident out of 10 was Chinese; most performed menial labour. Irish labourers came with the railroad construction boom during the 1860s. The Irish, French, and Italians tended to settle in San Francisco. As Los Angeles began to grow at the end of the 19th century, it lured Mexicans, Russians, and Japanese, but primarily an additional influx of Anglo-Saxons from the Midwest.

Discrimination against Asians

Discrimination grew strong, especially against Asians. An alien land law intended to discourage ownership of land by Asians was not ruled unconstitutional until 1952. At one time the testimony of Chinese in courts was declared void. Separate schools for Asians were authorized by law until 1936, and not until 1943 was the Chinese Exclusion Act repealed by Congress. As discrimination against the Chinese flared, Japanese were encouraged to immigrate, and in 1900 alone more than 12,000 entered California. Prospering as farmers, they came to control more than 10 percent of the farmland by 1920, while constituting only 2 percent of the population. Los Angeles became the centre of the nation's Japanese community, while San Francisco's Chinatown became the nation's largest Chinese settlement.

Discrimination against the Japanese smoldered until World War II, when about 93,000 Japanese-Americans lived in the state. Some 60 percent were American-born citizens known as Nisei; most of the others were Issei, older adults who had immigrated before Congress halted their influx in 1924. Never eligible for naturalization, the Issei were classed as enemy aliens. During 1942 almost all of California's Japanese-Americans, both Nisei and Issei, were moved to isolated inland camps and held under guard until 1945. At the end of the war they found their property sold for taxes or storage fees and their enclaves overrun. After years of litigation some 26,000 claimants were reimbursed for their losses at about one-third of the claimed valuation. About 85 percent of the Japanese-Americans had been farmers, but with their land gone they became gardeners or went into businesses and professions. In 1988 the U.S. Congress voted grants of $20,000 each to all Japanese-Americans who had been interned.

Asian immigration to California surged in the 1970s and '80s, with Filipinos, Vietnamese, and South Koreans among the newcomers. By 1987 the Asian population of California was estimated at about 6 percent of the total.

Few blacks settled in California until World War II, but between 1940 and 1980 the black population in San Francisco rose from about 5,000 to about 86,000 and in Los Angeles from 64,000 to more than 900,000. California has had among the largest gains of any state in black population, with those leaving the Southern states attracted to such cities as Los Angeles and Oakland despite high unemployment rates there.

Hispanic population

About one-third of the nation's Mexican-Americans live in California. Hundreds of thousands of Mexicans entered southern California illegally in the years prior to 1987. In that year the U.S. Congress granted amnesty to those who could establish specific conditions of prior residence. By 1988 about 1,700,000 Hispanics had received temporary resident status under amnesty provisions, an estimated half of them within California.

The burgeoning minority populations have confronted police and other officials, especially in Los Angeles, to protest discrimination and unemployment. No confrontation has reached the proportion, however, of the black riots that leveled much of Los Angeles' Watts area in 1965.

Religion. About one-third of Californians list church affiliations, a proportion far below the national average. Judaism is strongest in the Fairfax and Beverly Hills areas of the Los Angeles Basin, Roman Catholicism in San Francisco, and fundamentalist Protestant sects in those parts of southern California inhabited by migrants from the South and Southwest.

Los Angeles has long been notorious for its exotic cults. Aimee Semple McPherson, whose Angelus Temple boasted 35,000 members, is the best remembered of such evangelists, and faith healers still are popular. Scientology, calling itself "the common people's science of life betterment," thrived in southern California under the leadership of L. Ron Hubbard. Zen Buddhism enjoyed popularity in San Francisco during the 1950s, with English-born Alan Watts serving as its Occidental interpreter to a following that included the "Beat Generation" of that era.

Demographic trends. Native-born Americans remained the dominant factor in California's growth phenomenon in the mid-20th century. Many workers who flooded the defense industries during World War II remained as residents, along with hundreds of thousands who first visited the state as military personnel. About three-fifths of the population is concentrated south of the Tehachapi Mountains in about one-fourth of the state's area, with the greatest concentration in the small coastal area.

Present-day migratory patterns

By 1970, however, the state's growth began to level off. Later migration took place from the crowded cities of California to rural areas and to cities of the Mountain states. Demographers predict continued population increases for California, which is likely to maintain its rank as the nation's most populous state, but these predictions have been scaled down from earlier years.

The economy. In economic terms California is more aptly compared with nations than with states. Its total personal income is surpassed only by that of the United States as a whole and of a few other industrialized nations.

Industry has triumphed over remoteness; lacking iron and coal deposits, California has developed light industry. Financiers have been imaginative in seeking and employing capital, and many of the nation's largest banks and corporations are California-based, the latter principally involved in aerospace, electronics, computers, and oil and gas. California supplanted New York in 1965 as the leading state in the export of manufactured goods. The state is dominant in aerospace, agriculture, wine making, and the film and television industries. Despite soaring taxes, California continues to attract high-income immigration and technologically oriented industry.

Agriculture. The foundation of California wealth lies in agriculture. Its fields and orchards yield more than 200 agricultural products of astonishing diversity from largely irrigated farmland. Its major cash products are cattle, milk and cream, cotton, and grapes. California produces about one-third of the nation's canned and frozen vegetables and fruits. About half of the farm output comes from the Central Valley, which is irrigated through a labyrinth of dams, canals, and power and pumping plants.

The state's agricultural supremacy dates from 1947, when its farm output first exceeded that of any other state. A growing season of nine to 10 months ranks Fresno, Kern, and Tulare counties among the top counties in the nation in value of farm produce. Most farms are huge, and most farm income is earned by only a small percentage of the farms. Many large landholdings have derived from federal land grants to railroads. Such farms have tended to become agricultural assembly lines with absentee owners, high mechanization and productivity, and persistent labour strife. Most farms specialize in one or two crops: almonds grow north of Sacramento; cotton and forage crops, figs, and grapes near Fresno; and in the wet delta, asparagus, tomatoes, rice, safflower, and sugar beets. Specialization has been enhanced by research at the University of California at Davis; this institution also counsels the California wine industry, which produces 90 percent of all the wine made in the United States. The citrus industry, almost destroyed in the 1940s by a virus, ranks second to that of Florida in production of oranges.

Farming as big business

Premium wine grapes grow in the Napa and Sonoma valleys north of San Francisco and in adjacent areas. The Imperial Valley in the Colorado Desert in the extreme south, though smaller in area than the Central Valley, has about 500,000 irrigated acres (202,000 hectares) of farmland. Other major farming areas include the Coachella Valley near Palm Springs, where dates and grapefruit grow, and the Salinas Valley and Monterey Bay region.

The farm labour pool is made up of low-income labourers, including the many migrants and Mexican nationals crossing the border in harvest seasons. Long abused, migrant labourers organized in the late 1960s under the leadership of César Chavez and began lengthy strikes that drew nationwide support in the form of consumer boycotts. Thereafter, however, Chavez' United Farm Workers lost much of its membership to the Teamster Union.

Mining. Petroleum production grew rapidly after 1895, with oil strikes in the Los Angeles–Long Beach area. California led all states in petroleum production from 1900 to 1936. Reserves now, however, are being depleted at a rapid rate, and fuel and natural gas are imported. Petroleum continues, however, to exceed the total of all other minerals in value of production. Gold mining is now insignificant. Other production includes natural gas, cement, sand and gravel, borate, soda, and salt.

Fisheries and forestry. California has a significant commercial fishing industry. Largely ocean fish, the yield includes tuna, mackerel, sole, squid, sardine, and salmon. Ownership of commercial forestland is almost equally divided between public agencies and private interests, with a total of almost 17,000,000 acres in use.

Industry. The aircraft plants and shipyards were supplemented after World War II by branch plants of many Eastern and Midwestern industries. Federal research-and-development funds allocated to California organizations also contributed to the postwar economy. By the 1970s and '80s, California's industries had diversified to include computer sciences, biotechnology, and health care. Construction has also become a major industry.

All federal military services have major facilities in California, affecting both the social and economic life of the state. San Diego, Long Beach, and San Francisco have home-port naval bases. Recruit training is the major role of naval and marine corps bases in San Diego, and infantry training takes place at Fort Ord near Monterey. Camp Pendleton, a marine base, encompasses the last large undeveloped area along the southern California coast. Air force activity centres around the Vandenberg base on the central coast and six other air commands, including the remote and esoteric test facilities in the Mojave Desert.

Military facilities

Tourism and recreation. Tourism is a consistent source of income in California. Inherent to this industry are such theme parks as Disneyland and Sea World; these sizable employers are part of the state's large service industry, which ranks second only to manufacturing among the state's industry payrolls.

No state offers more diverse recreation. There is good skiing along the Sierra Nevada as far south as Big Bear Mountain near San Bernardino. Squaw Valley near Lake Tahoe was the site of the Winter Olympic Games in 1960. The beaches of southern California, especially those from Santa Barbara to San Diego, are legendary; they give rise to a wide range of water sports, of which surfboarding is prominent. Hikers pursue the trails of the High Sierra, including the 212-mile John Muir Trail through the heartland of the Sierra Nevada, and the Pacific Crest Trail, which runs the length of the state. Fishing enthusiasts and hunters choose from extraordinary range and diversity in their sports.

More than one-fourth of the state's land area is set aside in recreation areas or in national seashores or wildlife refuges. Yosemite, Kings Canyon, and Sequoia are national parks in the Sierra Nevada. Along the 1,100-mile (1,770-kilometre) California coast about 40 percent of the shoreline is accessible and is visited by an estimated 50,000,000 people each year. Redwood National Park preserves 58,000 acres (23,500 hectares) of redwood trees extending for nearly 40 miles along the Redwood Highway near Crescent City. Among the 250 units of the state park system is Anza-Borrego Desert State Park, located in the Colorado Desert. Running 54 miles north–south and containing some 550,000 acres, it is the largest continuous state park in the United States. In addition there are more than 5,000 city, county, and special district parks in the state, including the four-mile-long Golden Gate Park in San Francisco.

Film industry. The industry for which California has been most popularly known is that of movies and television, centred in and around Hollywood. The pioneers of the motion-picture industry found southern California extremely well suited to their needs of maximum sunshine, mild temperatures, varied terrain, and a labour market. In 1908 one of the first "story" moving pictures, *The Count of Monte Cristo,* begun in Chicago by William N. Selig, was finished in the Hollywood area by Francis Boggs.

Development of Hollywood

The 1920s, '30s, and '40s saw Hollywood as the centre of a movie industry with a worldwide market. Real estate boomed, and riches were extravagantly displayed. The studios were ill prepared, however, for the revolution that they faced as a result of competition with television beginning after World War II. Grown soft from the ready demand for any product, they found that millions were staying home to see anything on television in preference to going out to the motion-picture house. At about the same time, a series of court decisions judged the major producing companies to be trusts in restraint of trade. Although new techniques such as the wide screen, richer colour, new lenses, and stereophonic sound were introduced, serious losses were suffered by the industry. Major studios began to sell their film backlogs and to sell or lease their facilities to television concerns. Some studios, such as Universal, became mammoth television producers. The presence of thousands of technically skilled artisans in the Hollywood area, as well as vast amounts of equipment, make it unlikely that the entertainment industry will ever be completely uprooted.

Transportation. Transportation, primarily by automobile and airplane, is in part both the cause and the product of the restless mobility of Californians, who move their residences more often than the average American and travel considerably more both for business and pleasure. California has the greatest concentration of motor vehicles on Earth and the most extensive system of multilane divided freeways. As in most American cities, light rail transit systems were largely discontinued in California cities after World War II. In the face of increasing traffic congestion, however, they have begun to return. The Bay Area Rapid Transit (BART) system, in San Francisco, was completed in the early 1970s. A San Diego trolley system, first built

in the late 1970s as a link to the Mexican border, was expanded in the late 1980s.

The freeway system

The freeway system is so extensive that one can drive on arterials from San Diego almost 500 miles northward through Los Angeles and the Central Valley without encountering any traffic signals or stop signs. Freeway construction has declined since the 1970s because of public opposition based on ugliness, pollution, and usurpation of private and community property rights. The rise of the freeway system after World War II coincided in Los Angeles with the demise of a 1,200-mile (1,920-kilometre) interurban rail system that had once been the longest such system in the nation. The lack of a conventional urban core in Los Angeles, along with low population densities, has made it difficult to construct modern rapid-transit systems there. Work on a rail link between downtown Los Angeles and the San Fernando Valley began in 1986, however, and construction of rail transit lines between Long Beach and Los Angeles and between Norwalk and El Segundo, near Los Angeles International Airport, began in 1985.

Transport of goods in California is predominantly by trucks. The intricate canals and waterways of the Sacramento River delta carry some waterborne freight traffic, and there is some coastal freight traffic. Maritime shipping across the Pacific Basin is centred at the Long Beach–Los Angeles ports, whose combined volume of cargo is several times greater than that handled by the Oakland, San Francisco, and Richmond ports in northern California.

Air commuting has increased phenomenally. The air corridor connecting San Francisco, Los Angeles, and San Diego has a greater volume than that linking Washington, D.C., New York City, and Boston. Air traffic congestion has become critical, but not so dire as that of the ground traffic around airports.

Administration and social conditions. *Government.* California is governed under a constitution framed in 1878–79, its detail reflecting the disillusionment of the period with rampant graft. Before a series of deletions began in 1966, it had grown longer than any governmental constitution except those of Louisiana and India. Reform has often been undertaken in California through constitutional amendment. Those instituted by Governor Hiram Johnson in 1911 included provisions for voter initiative of and referendum on legislation, recall of elected officials, the direct primary, woman suffrage, and a unique system that allowed candidates to run in primaries of opposing political parties. Since 1962 constitutional revision may be made by voters without calling a convention.

State executive officers are elected for four-year terms, with members of more than 30 boards and commissions being appointed by the governor. The legislature comprises the Senate, with 40 members, and the Assembly, with 80 members. Legislative dominance is held by populous southern California at the expense of rural areas.

California's judicial system

The judicial system has five levels, including the seven-member Supreme Court, district courts of appeal, and superior, municipal, and justice courts. Superior courts are the major trial courts, whereas the more numerous municipal districts hear lesser matters.

Local government is conducted through almost 4,000 agencies, including 58 counties and a few hundred incorporated cities. Counties and cities may establish charters or accept general-law provisions and statutory laws. Cities operate under variations of mayor–council–manager control. Los Angeles and San Francisco operate under mayor–council, while San Diego and San Jose employ city managers, who assume a large share of administrative duty.

Volunteer party organizations often have usurped roles ordinarily fulfilled by the Democratic and Republican party structure. The parties are forbidden to endorse any candidate prior to the primary, but unofficial organizations do so and are often better funded and organized than the party structure. To overcome this party ineffectiveness, candidates turn to professional campaign managers to enhance their public images.

Attempts at machine politics have proved ineffectual because of voter mobility, lack of party entrenchment, and the prime role of civil service in bestowing jobs. The vastness of the state and the political cleavages between the liberal north and the conservative south make it difficult for one party to sweep statewide offices, even with majority registration. Traditional party alignments seem of minor significance to many Californians, and crossovers are common despite heavy Democratic pluralities in registration.

Finances. After state and federal aid, property taxes provide the chief source of local revenue. Rising income, sales, and gasoline taxes support state expenditures dominated by highway building, education, and welfare costs.

Education. California is oriented toward tax-supported public education. The two-year junior or community college was introduced in California in 1907, and there are now about 100 such colleges. Several four-year state colleges and the University of California system complete the public higher-education structure. The University Extension system operates throughout the state. About 10 percent of California schoolchildren and a slightly higher percentage of college-age students attend private schools.

Higher education

According to a master plan that attempts to avoid overlapping roles in the complex system of public colleges and universities, the top one-third of high school graduates is eligible to enroll at one of the state university campuses, which retain supervision over doctoral degrees. Four-year state colleges also draw from among the top one-third of high school graduates. High school graduates from the lower two-thirds of their classes attend two-year colleges and often are able to transfer at the end of that period to one of the four-year campuses.

The University of California has campuses at Berkeley, Los Angeles, Davis, Riverside, Santa Barbara, San Francisco, Irvine, Santa Cruz, and San Diego. The campuses at Santa Cruz and San Diego were established on variations of the Oxford University system of numerous small independent colleges sharing limited central facilities or services. The original campus at Berkeley was founded in 1855 and has remained one of the most prestigious academic communities in the nation.

Health and welfare. California long has been considered a liberal state in the extent of its health and welfare statutes. State funds are dispersed through systems such as aid to families with dependent children, aid to the totally disabled, aid to the blind, and old-age security.

Cultural life. California's culture is marked by widespread public involvement with the arts and enthusiasm for cultural trappings as symbols of achievement, often in the form of lavish expenditures to erect galleries, museums, and concert halls.

As a state California has harboured—and given birth to—a notable procession of creative people in all the arts. Early writers associated with California came from outside the state: Bret Harte, born in New York; Mark Twain, in Missouri; Joaquin Miller, in Indiana; and Ambrose Bierce, in Ohio. But the San Francisco of the gold rush days provided an eager audience for their writing, as it did for theatre and music. There followed a line of writers who came as close to establishing a regional tradition as have artists in any medium. Jack London, chronicler of men amidst frontier violence, was born in San Francisco. Frank Norris and Upton Sinclair, who opposed the social ills of their times in a foreshadowing of the later work of John Steinbeck and, to a lesser degree, of William Saroyan, were California-born but left the state in the 1940s. The naturalist John Muir, the progenitor of a school of environmental writers, extolled the state's natural wonders. Robinson Jeffers, who lived in California much of his life, was the state's most renowned poet. An influx of literary figures as screenwriters into Hollywood in the 1930s and '40s established little in the way of regional cultural tradition, and the California milieu became instead a favourite target of satire in such novels as Nathanael West's *The Day of the Locust* and Evelyn Waugh's *The Loved One,* and in works by F. Scott Fitzgerald, Budd Schulberg, Raymond Chandler, and Ross Macdonald.

California in literature

San Francisco has produced such painters as David Park, Elmer Bischoff, and Richard Diebenkorn. Los Angeles seems more successful as a marketplace for art, with a thriving colony of galleries along La Cienega Boulevard. The numerous wealthy art collectors in southern Califor-

nia are prominent in funding such institutions as the Getty Museum (1953), the Los Angeles County Museum of Art (1965), and the Museum of Contemporary Art (1979). The Music Center of Los Angeles County is a concert and theatre complex that was constructed during the 1960s by private contributions.

The California Arts Commission was created in 1963 to establish the state as a major cultural centre. Its low budget, however, has limited it largely to inventorying the arts and to providing token sponsorships. Tax-supported state institutions, most prominently the University of California and its extension program, are active in presenting dance recitals, plays and films, concerts, and lectures. The Theater Group of the University of California at Los Angeles is one of the most innovative in the nation. Experimental theatre in San Francisco has bloomed from time to time, and an often distinguished mixture of light and avant-garde theatre is offered throughout the year at several theatres, including the community-sponsored Old Globe Theater and La Jolla Playhouse in San Diego and the Magic Theatre in San Francisco. Amateur theatrical groups are widespread, as are community orchestras, chamber-music societies, and weekend artists. Carmel, Big Sur, and Sausalito have harboured communities of workers in diverse arts. The symphony orchestras of San Francisco and Los Angeles have achieved international recognition, as has the San Francisco Opera Company. The San Diego Symphony began performing in its own downtown hall in 1985.

Hollywood still is responsible for the bulk of the national movie and television output, and as such it remains an international symbol of glamour. An increasing number of national magazines and periodicals emanate from editorial offices that are located in California. Metropolitan California newspapers have decreased in number, but their total circulation has grown, led by the *Los Angeles Times,* with the largest number of readers in the state.

HISTORY

Exploration. Modern California derives from the discovery of gold at Sutter's Mill in 1848, just nine days before Mexico signed the Treaty of Guadalupe Hidalgo, ceding to the United States a vast area of the Southwest that included all of present-day California. The region was neglected by the Europeans for more than three centuries after its first sighting in 1542 by the Spanish navigator Juan Rodríguez Cabrillo. The merchant Sebastián Vizcaíno sailed the southern California coast in 1602, naming San Diego, Santa Catalina Island, Santa Barbara, and Monterey Bay. Despite these early explorations, California was left to its Indian population—estimated at 130,000 when Spanish explorers reached California in 1542.

Settlement. Pressure for settlement came from missionaries eager to convert the Indians and from the intrusion of Russian and British traders, primarily in search of sea otter pelts. In 1769 the Spanish viceroy dispatched land and sea expeditions from Baja California, and the Franciscan friar Junípero Serra established the first mission at San Diego. Gaspar de Portolá set up a military outpost in 1770 at Monterey. Colonization began after 1773 with the opening of an overland supply route across the southwestern deserts.

Establishment of the Spanish missions

The 21 missions established by Serra and his successors were the strongest factors in developing California. While attempting to Christianize the Indians, the padres taught them farming and crafts. With the labour of the Indians, the padres irrigated vast ranches and traded hides, tallow, wine, brandy, olive oil, grain, and leatherwork for the manufactured goods brought by Yankee trading vessels around Cape Horn.

U.S. colonization and acquisition. Secularization of the missions was sought by Spanish-Mexican settlers known as Californios when Mexico became independent of Spain in 1821. Between 1833 and 1840 the mission ranches were parceled out to political favourites by the Mexican government. The padres withdrew, and the Indians were cruelly exploited and diminished. In 1841 the first wagon train of settlers left Missouri for California. The colony grew slowly, but in 1846 the Northwest became a part of the United States, and settlers at Sonoma proclaimed an independent California republic. In May the United States declared war on Mexico, and in July the U.S. flag was raised at Monterey. Only minor skirmishes occurred before the Californios surrendered to troops under John C. Frémont near Los Angeles in January 1847.

The gold rush. Early in 1848 James Wilson Marshall, a carpenter from New Jersey, picked up nuggets of gold from the American River at the site of a sawmill he was building near Coloma. By August the hillsides above the river were strewn with the tents and wood huts of the first 4,000 miners. From the East, prospectors sailed around Cape Horn or risked disease in hiking across the Isthmus of Panama. The hardiest took the 2,000-mile overland route, where cholera proved a greater killer than Indians. About 40,000 came to San Francisco by sea in 1849. Some 6,000 wagons, carrying about 40,000 more, moved west that year over the California Trail. Few of the prospectors struck it rich. The work was hard, prices were high, and living conditions were primitive. The wiser immigrants became farmers and storekeepers.

Prospectors for gold

Gold hastened statehood in 1850 (as a part of the Compromise of 1850); and, though the gold rush peaked in 1852, the momentum of settlement did not subside. Nearly $2,000,000,000 in gold was taken from the earth before mining became almost totally dormant.

The Civil War and after. The Compromise of 1850 did not settle the slavery issue in California. Political parties were divided according to whether they believed that California should be a free or a slave state, and one movement, led by the backers of California senator William M. Gwin, sought to divide California into two states, one slave and one free. The same group also attempted to promote a Pacific Coast republic. At the opening of the war, however, California sided with the North.

After the war, control of the governor's office passed back and forth between Democrats and Republicans to the end of the century. The political climate after 1876 was distinguished by labour problems and the activity of those seeking to control mining, irrigation, and fruit growing through state funding. Economic problems were particularly intensified by the forces seeking the exclusion of the Chinese, who provided cheap labour. A slump in the 1870s brought increased discontent among the unions.

The problems and agitation of the period resulted in the constitution of 1879, which carried reforms but discriminated against the Chinese. An exclusion law passed by the U.S. Congress that year was killed by presidential veto, but in the next year a treaty agreement with China allowed U.S. regulation of Chinese immigration. This was followed by the Chinese Exclusion Act in 1882, which suspended Chinese immigration for 10 years. In 1902 Congress reenacted exclusion legislation against the Chinese. By cutting off cheap labour, exclusion helped make the huge single-crop ranches unprofitable and led to the proliferation of smaller farms growing varied crops.

Japanese farm workers were brought in to replace the Chinese, but as they grew successful the "yellow peril" outcry rose once again. Japanese agitation, focused largely in San Francisco, affected domestic and international policies. In 1913 the Webb Alien Land Law, designed to keep the Japanese from owning land, was the culmination of anti-Japanese lobbying. Japan and the United States then concluded a gentleman's agreement in which Japan agreed to halt further emigration to the United States.

The 20th century. Reform movements of the early 20th century promoted, among other things, greater influence of the people in government. The effects of the Great Depression were not as devastating in California as in most other parts of the country but were felt nonetheless. Migrant farm workers from the Dust Bowl of the Great Plains flocked into the state to seek work, a situation that caused widespread social unrest. Depression conditions gave rise to a number of social welfare schemes, including the End Poverty in California (EPIC) reform movement, presented by the noted author Upton Sinclair. The Democratic Party also grew strong during the Depression era. Nevertheless, Republicans dominated the statehouse during the first half of the 20th century, the Democrats

gaining control only during 1939–43. Notable among the Republican governors was Earl Warren, who resigned in 1953 to become chief justice of the United States, the first person of his state to hold the office.

In 1958 a Democratic victory installed Edmund Brown as governor. The Republican defeat reflected a national trend, and Democrats not only won gubernatorial and U.S. senatorial races but for the first time in the 20th century received a majority of seats in both houses of the state legislature. In 1959 the Democrats ended the California system of cross-filing, whereby a candidate could file for both parties in primary elections. Governor Brown was reelected in 1962, but in 1966 Republican Ronald Reagan, a former actor who in 1981 became president of the United States, replaced him. From the 1960s on, the state seesawed between the Democrats and Republicans.

Meanwhile, the economy and population advanced. The tide of immigration first moved toward southern California around 1900, spurred by citrus, oil, and some wariness of San Francisco after the earthquake and fire of 1906. Land booms came and went. Agriculture in inland valleys and industry in the cities boomed. During World War II aircraft plants and shipyards expanded, and in the 1950s and '60s research and educational facilities burgeoned as the movement of people to the West Coast came to include an unusual share of scientists and academicians.

20th-century growth

The 1970s and '80s were marked by shifting demographics. New urban centres emerged, and smaller cities experienced the most rapid rate of growth. The population of San Diego's metropolitan area rose past 2,000,000 and included the highest proportion of college graduates of any of the 10 most populous American cities. There was also rapid growth in new suburban areas near San Jose, Sacramento, and Riverside. For a century San Francisco had served as a financial and corporate centre, but most major corporations had relocated their headquarters to more populous southern California by the late 1980s.

California's population increased rapidly until the 1960s, when the rate of population growth began a slow decline. Also on the decline was the so-called dream factory that had manufactured the state's least tangible but most provocative product—motion pictures. As California continued to absorb migrants from all over the nation, it became, in most measurable specifics, more like America as a whole. (N.Mo.)

Hawaii

A group of volcanic islands in the central Pacific Ocean, Hawaii was characterized by Mark Twain as "the loveliest fleet of islands that lies anchored in any ocean." The name is thought to derive from Hawaiki, the former name of Raiatea, one of the Society Islands, from where Polynesians sailed in voyaging canoes to settle after AD 1000. It became, on Aug. 21, 1959, the 50th state of the United States.

Hawaii is economically vigorous, with diversified agriculture and manufacturing; strategically important to the global defense system of the United States; a Pacific Basin transportation and cultural centre, often called the Crossroads of the Pacific; and a major tourist mecca. Hawaiian activities of national and international importance include research and development in oceanography, geophysics, astronomy, satellite communications, and biomedicine.

The capital city of Honolulu, on the island of Oahu, is 2,397 miles (3,857 kilometres) from San Francisco to the east, and 5,293 miles (8,516 kilometres) from Manila, in the Philippines, to the west.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The land area of the state of Hawaii consists of the tops of a chain of emerged volcanic mountains that form eight major islands and 124 islets, stretching in a 1,500-mile crescent from Kure Island in the west to the island of Hawaii in the east, with a combined land area of 6,471 square miles (16,759 square kilometres). With the exception of Midway, a U.S. naval reservation near the western end of the archipelago, the leeward coral atolls and central lava islets, forming a total of only 3¼ square miles, are in the Hawaiian Islands National Wildlife Refuge. The eight major islands at the eastern end of the chain are, from west to east, Niihau, Kauai, Oahu, Molokai, Lanai, Kahoolawe, Maui, and Hawaii. Volcanic activity has become dormant, with the exception of the volcanoes of Mauna Loa and Kilauea on the easternmost and largest island, Hawaii, where spectacular eruptions and lava flows take place from time to time. The highest Hawaiian mountains are Mauna Kea and Mauna Loa, reaching 13,796 feet (4,205 metres) and 13,677 feet (4,169 metres) above sea level, respectively.

The eight major islands

There is little erosion in the geologically young areas, where the terrain is domelike and the volcanic craters are clearly defined. In the older areas the mountains have been shaped and eroded by the action of sea, rain, and wind. Their aspects thus include sharp and craggy silhouettes; abrupt, vertically grooved cliffs pocked with caves; deep valleys; collapsed craters (calderas); and coastal plains. The powerful Pacific surf, churning and crashing against the fringing coral shelves and the lava shorelines, has carried minute shells onto the shore and reduced coral and large shells to sand, creating the state's famous expanses of beach.

Volcanic ash, gravel, rotted vegetation, crumbling lava, and windblown sand and dust all help to make up the alluvial, residual, and organic soils found in various depths and densities in valley floors, the regions between mountain ranges, and along the shores. Oxidation of iron causes a ubiquitous bright red soil and rock strata. The iron content is, however, insufficient for smelting, and there are no coal or petroleum deposits.

Because the topography is generally abruptly descending or sloping, there are few surface collecting basins or lakes. Excess rainfall seeps through porous mountain areas to collect in subterranean chambers and layers retained by less permeable lava and ash beds, or it is prevented by underlying salt water from seeping to the sea. The resultant artesian water supply is tapped for use in irrigation and also for human consumption.

Heavy rainfall in mountainous areas produces an extremely voluminous runoff, which is responsible for the erosion that forms the numerous grooves, ridges, and V-shaped valleys characteristic of the older volcanic islands such as Kauai and Oahu. The action of rain combined with waves has had a particularly dramatic effect on the more exposed windward sections of the islands.

Climate. Hawaii lies just below the tropic of Cancer, and its mild tropical climate is considered by many people to be the world's ideal. Although often humid by U.S. mainland standards, temperatures are conditioned by the northeast trade winds, which prevail most of the year. Blowing for many miles over the open Pacific, the trades pass along the great reservoir's stabilizing influence, to make living in the islands delightfully comfortable.

Mild tropical climate

The average temperature in downtown Honolulu is 72° F (22° C) in the coolest month and 78° F (26° C) in the warmest, with extremes from 57° F (14° C) to 88° F (31° C) having been recorded there. The average water temperatures off Waikiki Beach, near Honolulu, range from 75° F (24° C) in late February to 79° F (26° C) in late September. Mountainous regions are considerably cooler, especially during the winter months, when there can be frost; a temperature of 1.4° F (−17° C) has been recorded on the summit of Mauna Kea, and winter snows frequently blanket the crests of Mauna Kea and Mauna Loa.

Rainfall variations throughout the state are dramatic. Mount Waialeale, on the island of Kauai, is often called the wettest spot on Earth, with an annual average rainfall of 444 inches (11,280 millimetres) over a 60-year period, the highest long-term median on record. The driest area in the islands is at Kawaihae, on the island of Hawaii, where the average annual rainfall is only 8.7 inches (220 millimetres). The average yearly rainfall in Honolulu is 23 inches, and, in Hilo, it is 129 inches.

As moisture-laden air is carried over the islands, most frequently by the trade winds, it is apt to condense, form cap clouds, and dissipate against the shores and mountains of the windward coasts, which are therefore more lush in foliage than the leeward coasts.

Plant and animal life. The seeds of endemic plant species were carried to Hawaii by birds, winds, or currents and tides, bringing about extensive forestation, shrubbery, and grasslands, where soil and precipitation were favourable. Since the first Polynesian settlement a tremendous variety of food and ornamental plant life from many parts of the world has been introduced. Food plants grown commercially or in backyards for home consumption include sugarcane, pineapples, papayas, bananas, mangoes, guavas, lichee, coconuts, avocados, breadfruit, macadamia nuts, limes, passion fruit, taros, and tamarinds. Nearly all varieties of common garden vegetables are raised in the islands, and flowers abound all year.

The effects of isolation on natural life

Endemic birds, long isolated from others of their kind, have taken on certain characteristics of their own. These include the nene (Hawaiian goose), the Hawaiian stilt, and a variety of small forest birds. Some species have become extremely rare, but as the result of an increased environmental awareness, great strides have been taken to preclude their extinction. Seabirds nest in profusion on the western islands of the archipelago and to a far lesser extent among the major eastern islands. There has been considerable importation of birdlife. Quantities of mynas, sparrows, cardinals, and doves live in the trees in both urban and country areas. Every fall the small golden plover make an awe-inspiring, nonstop 3,000-mile (4,800-kilometre) flight from Alaska to Hawaii, where they spend the winter, together with ducks from Alaska, Canada, and the northwestern United States.

Wild animal life includes mongooses, rats, frogs, toads, and, in the more remote regions of some of the islands, deer, sheep, pigs, and goats. The insect population is multitudinous, and marine life abounds in Hawaiian waters.

Settlement patterns. Agricultural and fishing activities bring about extensive and scattered rural settlement, ranging from tiny fishing villages far off the main roads, scant clusters of small houses in isolated valleys, solitary farm and ranch houses, to large coastal and upland villages and plantation and ranch towns.

The older houses in the smaller villages are largely single-family, raised, frame structures, with corrugated-iron roofs. Plants of native origin skirt the foundations of houses, and the yards are informally planted with fruit and flower trees. In all but the very small villages, there are a school, markets, a post office, a fire station, and at least one church. The day's activities traditionally begin early and end early, following the sun. The life-style of the rural people is simpler and less sophisticated than that of the urban populations, and the country dwellers tend to retain more of the speech patterns and customs of their distinctive ethnic backgrounds.

Village and urban life

During the 1950s and '60s there was a building boom in Hawaii of such magnitude that the configuration of entire towns was altered. The most graphic example of this was in the city of Honolulu, where construction of 20- and 30-story buildings gave the city, once sprawling and low, a thrusting, multileveled skyline. On Oahu, erstwhile vacation or agricultural towns have become expansive residential areas for commuters to Honolulu and Pearl Harbor.

Urban settlement once consisted almost entirely of single-family dwellings, individual business houses and shops, small markets, and three- or four-story hotels. With the increase of residents and tourists since 1950, however, Hawaiian towns and cities have built more and more high-rise apartment houses, hotels, and business establishments, with the traditional individual shopkeepers becoming absorbed into the complexes of shopping centres and supermarkets. The concept of the planned city has been developed in areas that were previously open spaces or given over to agriculture.

The people. Most anthropologists believe that the original settlement of Hawaii was by Polynesians who migrated northwest from the Marquesas Islands perhaps as early as AD 400, to be followed by a second wave of immigration that sailed from Tahiti during the 9th or 10th century. Once they had established themselves in Hawaii, the Hawaiians had no further need to obtain supplies from their old homeland and underwent centuries of isolation. Although there are still rather close resemblances in linguistics, physical characteristics, and general customs and life-styles between the Hawaiians and their Polynesian relatives, a degree of racial individuality evolved.

The original Hawaiians were a brown-skinned people of large stature, highly skilled in fishing and farming, who adhered to an extremely rigid and strict system of laws that was set down by their chiefs and their priests. They worshiped and feared a group of gods not unlike, in character and power, the ancient Greek deities of Mount Olympus.

The first recorded contact between the Hawaiians and Europeans took place in 1778, when Captain James Cook came upon the islands. During the ensuing four decades the influence of European and American explorers, adventurers, trappers, and whalers stopping for fresh supplies at Hawaiian islands was to have a profound effect.

Impact of the cultural mix

Contact with people of different cultures who believed in only one god eventually brought about a spiritual revolution among the Hawaiians. In a series of defiant acts led by members of the royal family, the basic beliefs of the Hawaiian religion were undermined, and the priests were overthrown. Loss of faith in the old gods, intense interest and curiosity about the ways of the people of the United States and Europe, avid interest in learning to read and write, and a desire for spiritual identity brought about a swift adoption of Christianity on the part of the Hawaiians. The first group of Christian missionaries arrived from the United States in 1820, and by the mid-19th century the Hawaiian kingdom was largely a Christian nation.

It has been estimated that the population of the Hawaiian Islands at the time of Captain Cook's discovery was approximately 300,000. Virtually disease free, this population had no natural immunity to the diseases introduced from both West and East and fell easy prey to venereal disease, cholera, measles, bubonic plague, and leprosy, all of which contributed to the decimation of the native peoples. In 1853 the native population of the Hawaiian kingdom numbered 70,036.

The racial and religious makeup of Hawaii has undergone quite dramatic change since that time. Thousands of settlers from the Pacific Basin—primarily from Japan, the Philippines, and China—as well as immigrants from Europe and from the U.S. mainland carried their own customs, languages, and religions into the Hawaiian way of life. The descendants of these later settlers now far outnumber the descendants of the original Hawaiians. There is also a continuous influx and outflow of military and naval personnel and their dependents, connected closely to the continuing American presence in the Pacific.

Most of the state's residents live on the island of Oahu, 60 percent in the Honolulu urban area and another 20 percent in outlying districts. Because there are vast areas of Oahu devoted to agriculture and forest reserves, the majority of the population actually resides in high-density clusters. Honolulu is the only legally incorporated town or city in the state.

Linguistic diversity

Hawaii is English speaking. Although Hawaiian, formerly a major means of communication, is all but extinct, it remains in place-names and street names and in songs, and the local residents liberally sprinkle their speech with words and phrases from the traditional language. A pidgin English is spoken throughout the state in varying degrees of richness, while some of the older immigrants from Japan and China continue to speak their native tongues. As Filipinos continue to move to Hawaii, their language, too, is frequently heard in the state.

The largest religious groups are Roman Catholics and Protestants. There are, however, small but important groups of Buddhists and of adherents of other Asian religions.

The economy. Hawaii ranks relatively low among the states in terms of personal income, farm products sold, value of manufacturing shipments, retail sales, and bank deposits.

A major problem in Hawaii is the high cost of living, due in large part to Hawaii's insularity and dependence on imports. Transportation costs are included in the prices of nearly all consumer goods. As the population increases, housing grows increasingly difficult to acquire, and it is disproportionately expensive when compared with housing

costs in many of the mainland states. Building materials, most of which are imported, are expensive. Residential land is limited and highly priced, since much of the property, notably on Oahu, is owned by corporations and trusts. Efforts have been made through legislation to remedy this situation. Carefully planned housing located in communities in which the single-family home gives way to high-rise, high-density dwellings as well as townhouses and apartment complexes has become one solution to the shortages and expense associated with urban housing.

State and private interests

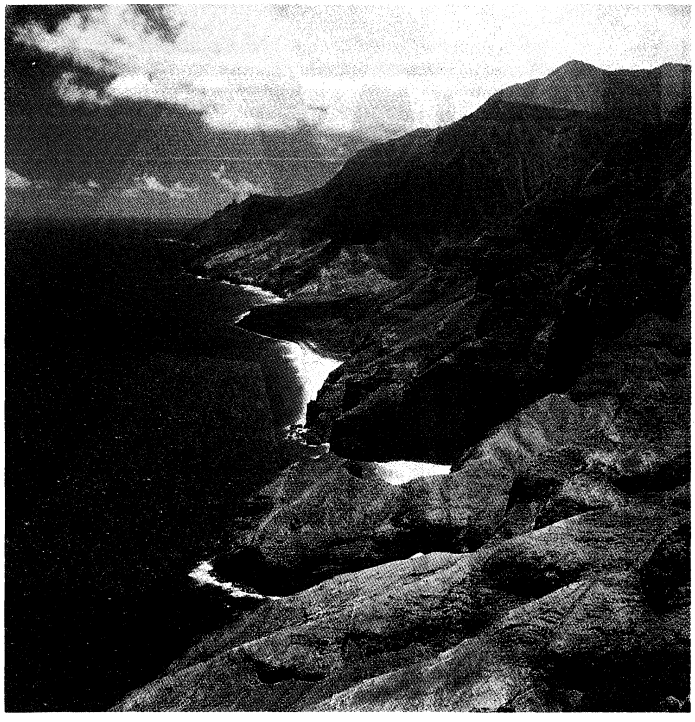
More than half the land in the state is owned by private individuals or corporations, although the state itself, holding more than one-third of the land, is the largest single landowner. State and county governments are major employers. Honolulu is the regional headquarters of the federal government, which owns one-sixth of the land.

Resources. Hawaii has no important mineral deposits; its only natural resources are its climate, water supply, soil, vegetation, and surrounding ocean and rock, gravel, sand, and earth quarried for use in construction and landscaping. Electric power is supplied by a small number of power companies operating oil-powered steam and diesel generators. Several military installations and some private institutions generate their own power. A small amount of hydroelectric power is generated on several of the islands, and in the mid-1980s a geothermal plant began producing electricity on the island of Hawaii.

Tourism. Tourism is Hawaii's largest industry. Expansion has been particularly rapid since World War II, and the growth has resulted in part from continued improvements in transportation and the stimulus provided by the state government and local businesses. The majority of visitors come from the U.S. mainland, Canada, Australia, and countries of the Far East, particularly Japan. About 60 percent of the hotel units are on Oahu, chiefly in Waikiki and the adjacent Ala Moana area. Visitors have access to a wide range of recreational and cultural facilities, such as golf courses, tennis courts, parks, surfing sites, beaches, restaurants, theatres, musical attractions, and sporting events. Tourism has helped Hawaii to become the centre of the international market of the Pacific Basin. Capital investment by U.S. mainland and foreign companies has increased tremendously.

Agriculture. Although the second largest source of income in Hawaii is the federal government, primarily through defense expenditures, agriculture remains the basis of the local economy. Hawaii is the second largest sugarcane-producing state in the nation and leads in the production of pineapples. Hawaii's dominance of the world pineapple market is challenged by the lower labour costs in pineapple-producing countries such as the Philippines. There has been a slow but steady growth of diversified agriculture, including grain sorghum, corn (maize), flowers, and nursery products. Livestock, poultry, and dairy production, together with some lumbering and commercial fishing, are other important sources of income. Nearly half of the commercial fish catch is aku (skipjack tuna).

© Joe Carini—Bear Productions



Sea cliffs on the northwest coast of the island of Kauai, Hawaii.

Industry. Hawaii has several hundred companies engaged in diversified manufacturing. Heavy-manufacturing plants, using raw materials for the most part imported from the U.S. mainland, include an oil refinery that produces a variety of petroleum products and chemical compounds, a steel mill manufacturing reinforcing bars, several cement plants, a concrete-pipe plant, and an aluminum-extrusion plant. Heavy manufacturing is confined mainly to the island of Oahu. Most building lumber is imported from the mainland. A number of garment manufacturers, largely situated in Honolulu, produce printed fabrics and apparel marketed locally, nationally, and abroad.

A wide variety of Hawaii-grown foodstuffs, sold locally and exported to the mainland, are processed in the state. These include Oriental and Hawaiian food specialties, such as tropical fruit juices, jams and jellies, candies, coffee, macadamia nuts, and various alcoholic beverages.

Major Hawaiian industries are unionized, as are many of the service and construction industries. The largest union in the state, and one with a turbulent history, is the International Longshoremen's and Warehousemen's Union.

Trade and finance. Exports are largely in the form of sugar, canned pineapple, garments, flowers, and canned fish. Major imports are fuel, vehicles, food, and clothing.

State taxes are collected under a centralized tax system. The chief sources of the state's revenue are a general excise tax, individual income taxes, and federal grants-in-aid.

Transportation. Ocean-surface transportation is Hawaii's lifeline, and Honolulu Harbor, with its extensive docks, warehouses, and storage sheds, is the centre of Hawaiian shipping. A large percentage of the cargo ships ply between Hawaii and California ports, a few between Hawaii and the East Coast of the United States via the Panama Canal, and others from western Pacific ports. Around-the-world passenger ships carry visitors through Honolulu, and there is an interisland luxury cruise line. Tug-pulled barges and small freighters transport goods from Honolulu to the outer islands, returning with agricultural crops and livestock.

The importance of Honolulu Harbor

The majority of voyagers to and from Hawaii travel by air, as do nearly all interisland passengers. The Honolulu International Airport, on Oahu; General Lyman Field at Hilo, on Hawaii; and the Kahului Airport, on Maui, are the major civilian airports capable of serving large-jet traffic. There are several smaller airports among the islands and a number of small private airfields. Military authorities maintain a number of airports throughout the state.

Most of the roads follow lowland contours, circling the islands along or near the shorelines and crossing islands only between mountain ranges. There are many spectacular mountain roads providing dramatic vistas. On Oahu two tunnels bring traffic from the heads of two valleys behind Honolulu through the Koolau Range and out into the windward, or northeastern, side of the island. Hawaiian roads range from narrow country tracks to an eight-lane freeway, which crosses the city of Honolulu.

Administration and social conditions. *Government.* Hawaii is governed by a state constitution that was originally adopted in 1950; it was amended in 1959, at the time of admission to statehood, and further amended at the constitutional convention of 1968. The governor and lieutenant governor, elected for concurrent terms of four years, must be members of the same political party. They are not permitted to serve more than two consecutive terms. The only other elected members in the 17 departments of the executive branch are the members of the Board of Education. Hawaii's bicameral legislature con-

Executive, legislative, and judiciary

sists of the Senate, with 25 elected representatives from 25 senatorial districts, serving four-year terms, and the House of Representatives, consisting of 51 members elected from 51 districts for two-year terms. The state judicial system consists of the Supreme Court, an intermediate appellate court, four circuit courts, and four district courts, as well as a family court, a land court, and a tax appeal court. Judges in the higher courts are appointed by the governor, subject to approval of the Senate.

Hawaii's governmental structure is unique among the states in that it is limited to two levels of government: the state and the four counties, each with a mayor and a council. There are no municipal governments.

Primary elections are held in October and general elections in November. Party competition is intense in Hawaiian politics. During the first half of the century, the Republican Party remained dominant, but party success at the polls began to seesaw somewhat, and the Democratic Party has captured a majority of House, Senate, and council seats on several occasions.

Military. Hawaii holds a strategic position in the defense system of the United States. Pearl Harbor, a vast shipyard for the repair and overhaul of U.S. fleet units, is the home port for many U.S. naval ships. It serves as a training base for submarine and antisubmarine warfare forces. The headquarters of the commander in chief, Pacific, and of the Fleet Marine Force, Pacific, are at Camp H.M. Smith. The major army, marine, and air force bases are Schofield Barracks, Fort Shafter, Fort De Russy, Hickam and Wheeler air force bases, and the Kaneohe Bay Marine Corps Air Station. In addition to these there are military installations, camps, and airfields of varying sizes throughout the state. More than 100,000 U.S. military personnel and their dependents are stationed in or have their home port in Hawaii, and their presence has an important influence on the local economy and social life.

Education. Hawaii's school system provides educational facilities from nursery school through the graduate school level. Institutions of higher learning include the University of Hawaii, several smaller private colleges, and a state-established system of two-year community colleges. Private business, technical, and specialized schools provide additional educational facilities and opportunities.

The Center for Cultural and Technical Interchange Between East and West, commonly referred to as the East–West Center, is a project of the federal government housed at the Manoa campus of the University of Hawaii. It provides specialized and advanced academic programs and technological training to students from the United States and from countries in Asia and the Pacific.

Health and welfare. The Department of Health maintains hospitals, health centres, clinics, care centres, and nursing services. The Hawaiian Home Lands Commission controls the transfer of land use to qualified persons of Hawaiian ancestry for homesteading.

Cultural life. Hawaii's cultural milieu is the result of overlay after overlay of varied cultural groups. The force of the original culture remains evident in the islands, although the Hawaiian race has become diminished and diluted over the years through death and intermarriage.

Uniqueness of Hawaii's cultural heritage

Vestiges of New England culture remain, as do the cultures of the early Asian immigrants. With the advent of fifth-, sixth-, and seventh-generation descendants of Asian and Caucasian immigrants and the massive influx of Americans from all parts of the country, the cultural overlays have melded to form a uniquely Hawaiian culture.

Interest in the arts is high, and many distinguished artists, photographers, and performers have been native residents. Appreciation of classical, modern, and experimental art forms is manifest in attendance figures at galleries, concerts, legitimate theatre performances, and museums. Many ethnic groups preserve the traditions of their ancestors or combine or modify music and dance forms.

An assortment of cultural and scientific institutions in Hawaii provides a wide variety of opportunity for the appreciation and understanding of the fine arts, history, traditions, and sciences. The Bernice P. Bishop Museum, founded in 1889 in Honolulu, is a research centre and museum dedicated to the study, preservation, and display of the history, sciences, and cultures of the Pacific and its people. The Honolulu Academy of Arts, often called the most beautiful museum in the world, houses a splendid collection of Western art, including works by the late 19th- and early 20th-century masters Monet, van Gogh, Matisse, Gauguin, and Picasso. Its collection of Asian art is also one of the finest in the Western world. The active art, music, and drama departments in Hawaiian schools and colleges and at the University of Hawaii contribute to the expanding cultural life of Hawaii, while the state has several theatre organizations, professional and amateur. The Honolulu Symphony Orchestra performs concerts in Honolulu and on the other major islands. Its home is the Neal Blaisdell Center, a municipal theatre–concert-hall–arena complex, where touring opera companies and ballet troupes and musical artists of international renown also perform. Honolulu's Chamber Music Society gives a concert series each year.

Hawaii has two national parks—Hawaii Volcanoes, on the island of Hawaii, and Haleakala, on Maui, as well as the much-visited U.S.S. Arizona Memorial in Pearl Harbor. There are also many state and county parks, including the Waimea Canyon State Park on Kauai. Surfing originated in ancient Hawaii and is now practiced at some 1,600 recognized surf spots throughout the islands. There are also some 25 miles of beaches accessible to the public.

HISTORY

The early Hawaiians

The first inhabitants of Hawaii may have reached the islands as early as AD 400 from the Marquesas. Contact with and settlement by Tahitians began about AD 1000. Powerful classes of chiefs and priests arrived and established themselves, followed by conflicts, similar to the feudal struggles in Europe, with complicated land rights contributing to the disputes. The early Hawaiians lacked a written language, and their culture was entirely oral and rich in myth, legend, and practical knowledge, especially of animals and plant life. The material life of the islands was hampered by the lack of metal, pottery, or beasts of burden, but there was great skill in the use of wood, shell, stone, and bone, and the huge double and outrigger canoes were technical marvels. Navigational methods were well developed, and there was an elaborate calendar. Athletic contests encouraged warrior skills.

European discovery. Captain James Cook, the English explorer and navigator, is generally credited with having made the first European discovery of Hawaii; he first landed at Waimea, Kauai Island, on Jan. 20, 1778. Upon his return in the following year, he was killed during an affray with a number of Hawaiians at Kealakekua Bay.

The initial discovery by Cook was followed by a period of intermittent contact with the West. During this period Kamehameha I used European military technology and weapons to emerge as an outstanding Hawaiian leader, seizing and consolidating control over most of the island group. For 85 years thereafter, monarchs ruled over the Hawaiian kingdom. In the early 19th century the American whaling fleet began wintering in Hawaii, and the islands were visited with mounting frequency by explorers, traders, and adventurers. Captain George Vancouver introduced livestock to the islands in 1792. In 1820 the first of 15 companies of New England missionaries arrived. By the middle of the century there were frame houses, horse-drawn vehicles, schools, churches, taverns, and mercantile establishments. A written language had been introduced, and European and American skills and religious beliefs—Protestant and Roman Catholic—had been imported. Hawaiian culture was irrevocably changed.

Establishment of U.S. dominance. Political maneuvering between U.S., British, and French consuls and naval forces brought about uncertainty in the governmental situation. The foundations of constitutional government were nevertheless laid down with the promulgation, by Kamehameha III, of the Declaration of Rights (June 7, 1839), the Edict of Toleration (June 17, 1839), and a written constitution (Oct. 8, 1840). These progressive steps—made under missionary influence—were followed by formal avowals of Hawaiian independence by the United States, Great Britain, and France. The ambitions of these

powers continued unabated, however, with a succession of overt and covert diplomatic moves, culminating in the signing of a reciprocity treaty with the United States in 1875. The kingdom was overthrown in 1893, and a republic was formed with U.S. support. This was followed by the joint annexation resolution of Congress in 1898, the final stamp of U.S. domination. This status was confirmed by the establishment of a territory on June 14, 1900.

Overthrow of the kingdom

The period until 1940 was distinguished by a rapid growth in population, the development of a modern economy based on the production of sugar and pineapples for consumption on the U.S. mainland, and the growth of transport and military links. Movements for statehood, based in part on Hawaii's obligation to pay U.S. taxes without having corresponding legislative representation, began to emerge. The Japanese attack on Pearl Harbor, on Dec. 7, 1941, precipitated not only Hawaii but the United States as a whole into World War II, and the islands were beset by an upsurge of military activity and a sometimes controversial curtailment of civil liberties. The post-1945 period was marked by further economic consolidation and a long constitutional path to statehood, a status finally achieved in 1959.

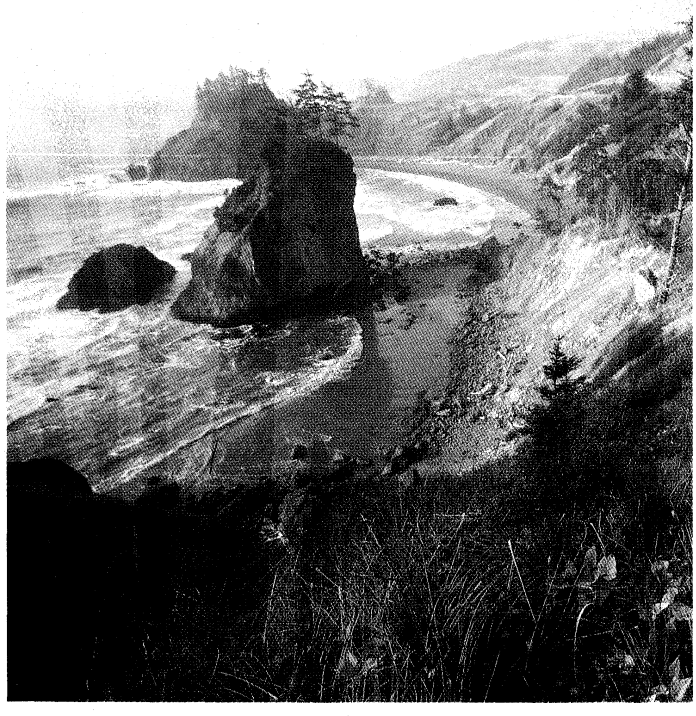
Since statehood tourism has grown in Hawaii, with ever-increasing numbers of visitors, especially from Japan and the U.S. mainland. They are lured not only by the warm climate and exotic beauty of the islands but also by a growing number of world-class resorts, built on such a grand scale that they are destinations in themselves. In addition, new telescopes atop Mauna Kea are helping Hawaii become a major world centre of astronomy.

(J.P.M.S./L.S.M.)

Oregon

Admitted to the Union as the 33rd state on Feb. 14, 1859, Oregon comprises an area of startling physical diversity, from the moist rain forests, mountains, and fertile valleys of its western third to the naturally arid and climatically harsh eastern deserts. Mountains, plateaus, plains, and valleys of different geologic ages and materials are arrayed in countless combinations, including such natural wonders as the Columbia River Gorge, Oregon Caves National Monument, Crater Lake National Park, the majestic snow-covered peaks of the Cascade Range, and the "moon country" of central Oregon. The name Oregon is thought to be Indian in origin.

© Tim Thompson



Samuel H. Boardman State Park, in the southwestern corner of Oregon.

To the north of the state's 97,073 square miles (251,419 square kilometres) of land and inland water lies Washington, from which Oregon receives the waters of the Columbia River; to the east, Idaho, more than half of the border with which is formed by the winding Snake River and its Hells Canyon, the deepest gorge on the North American continent; to the south, Nevada and California, with which Oregon shares its mountain and desert systems; and, to the west, the Pacific Ocean, which produces the moderate climate of Oregon's western lands.

The forested mountains of western and northeastern Oregon have supplied the traditional core of the state's economy. Its many forest-product plants produce more than one-fith of the nation's softwood lumber, much of its soft plywood, and large quantities of hardboard, pulp, and paper. Nationally, Oregon ranks first in the production of wood products. In addition, the multipurpose development of the Columbia River system provides huge quantities of electricity, water for irrigation and industry, shipping channels, and water for recreation. The heartland of Oregon, however, is the Willamette valley, containing the major cities of Portland, Eugene, and Salem (the capital) and a rich and diversified agriculture.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Oregon has nine major landform regions. Of them the forest-blanketed Coast Range, which borders the Pacific Ocean from the Coquille River northward, is the lowest. Its elevations are usually below 2,000 feet, but Marys Peak, southwest of Corvallis, reaches 4,098 feet (1,249 metres).

The Klamath Mountains, which extend from California, lie south of the Coast Range and west of the Cascades. Of ancient resistant rocks, they have had a complicated geologic history. They are higher and more rugged than the Coast Range and lack the north–south orientation. The Rogue River, bisecting the area, provides the major drainage. Thick forests grow on these mountains, which also contain rich mineral deposits.

The Willamette valley is essentially an alluvial plain produced by burying stream-modified lowland with enormous quantities of sediments brought down by tributary streams from the bordering mountains. The low, hilly areas in the central and northern portions are composed of resistant rocks. This valley contains the prime land of the state, and its soils support intensive agriculture.

Willamette valley

The Cascade Range forms a broad lava plateau. The wider western section is deeply eroded by numerous streams fed by heavy precipitation. The eastern section, less dissected, is crowned with a chain of volcanic peaks. Mount Hood, reaching 11,239 feet (3,428 metres) above sea level, is the highest peak in Oregon, and Mount Jefferson, rising to 10,497 feet (3,199 metres), is the second highest.

In the north central Oregon plateau, known as the Deschutes–Umatilla Plateau, a portion of the Columbia River basin, streams are entrenched and provide some bold relief. The areas lying between the streams are broad, little-dissected, smoothly rolling surfaces that provide the land for Oregon's large wheat ranches.

The Blue–Wallowa mountains comprise two highland masses in the northeastern part of the state. The name Blue Mountains refers to the eroded plateaus and ranges extending westward from the agriculturally important La Grande and Baker valleys. Basins and valleys, headquarters for large cattle ranches, are scattered through the Blue Mountains. The Wallowa Mountains, east of the La Grande and Baker valleys and near the Idaho border, contain the highest elevations in northeastern Oregon. They were heavily glaciated and display spectacular scenery.

The area of the High Lava Plains, or High Desert, is located south of the Blue Mountains and eastward from the Cascade Range. It is the youngest and least eroded of the landform regions of Oregon, but the smoothness of the surface is broken by cinder cones, buttes, and craters; other features include immaturity of erosion and localized interior drainage. Low precipitation, short and erratic growing seasons, and the absence of soil in many places

result in an arid landscape of skimpy vegetation, with the details of the surface features commonly visible.

The Great Basin of the Basin and Range Province to the south, which merges with the High Lava Plains, has long, narrow, asymmetrical fault block ranges that alternate with wide basins. Small volcanoes are numerous in the western portion, where pumice modifies surface runoff, vegetation, and land use. Irrigation agriculture is practiced in the Upper Klamath Lake area, and hay is grown with irrigation in a number of other basins and valleys, but most of this region is used by range livestock.

The Malheur–Owyhee Upland of southeastern Oregon is generally a high, warped plateau. It contains older lava and has been more eroded than the High Lava Plains. The major drainage system, the Owyhee River, has incised several notable canyons in an area locally called the Rimrock Country. Along the Snake River in the east central portion of the state there is highly productive irrigation agriculture to supplement livestock grazing.

Climatic variation

Climate. Oregon's climates range from equable, mild, marine conditions on the coast to continental conditions of dryness and extreme temperature in the interior. Location with respect to the ocean, prevailing wind and storm paths, and topography and elevation are the principal climatic control factors.

The narrow coastal area and the bordering mountain slopes are marine influenced. Temperatures are mild and equable: July temperatures average 57° to 60° F (14° to 16° C), January temperatures about 40° to 46° F (4° to 8° C). Summers are relatively dry but receive only half the sunshine possible; other seasons are cloudy and wet. Annual precipitation ranges from 60 to 120 inches (1,500 to 3,000 millimetres) or more.

The lowlands of the Willamette, Umpqua, and middle Rogue rivers are warmer in summer, slightly cooler in winter, and have less precipitation than the coast. July averages 67° to 72° F (19° to 22° C), with 65 to 70 percent of the possible sunshine; January averages about 40° F (4° C). The rainy season extends from October through April, with precipitation averaging 35 to 40 inches, except in the middle Rogue valley, where 20 to 25 inches are common.

The Cascade Range has copious winter precipitation, including phenomenal snow depth, and short, dry, sunny summers. Above 3,000 feet, January average temperatures are below 32° F (0° C). Snow begins to fall in October and remains through April, with large patches persisting until July. The higher peaks support snowfields and small glaciers throughout the year. July average temperatures range from 50° to 60° F (10° to 16° C).

The north central Oregon plateau receives 10 to 20 inches of precipitation annually. Distribution is fairly even, but the majority of the rainy days occur in winter. Summers are sunny, with July temperatures averaging 70° to 75° F (21° to 24° C). The brisk winters have considerable sunny weather, and January temperatures average 31° to 33° F (−1° to 1° C). The plateau area of central and southeastern Oregon has climatic characteristics similar to the north central plateau except for somewhat less precipitation and lower temperatures at higher elevations.

The Blue–Wallowa mountains have climates that vary with location. The intermontane basins and valleys are similar to the north central plateau, with colder winters, while the higher, exposed elevations receive heavy precipitation, much of it in the form of snow during winter.

Forested areas

Plant and animal life. Forests cover about 30,000,000 acres (12,000,000 hectares) of Oregon. In the eastern two-thirds of the state, ponderosa pine, large sagebrush, and western juniper predominate, along with various annual grasses and wildflowers. On the Blue–Wallowa mountains and the eastern slopes of the Cascades occur great stands of ponderosa pine in association with ground coverings of bitter brush, green manzanita, and herbaceous plants. The western slopes of the Cascade, Klamath, and Coast ranges are heavily forested with stands of Douglas fir, with varying degrees of undercover vines and intrusions of other tree growths depending on the age of the stand. In cleared areas of the damp coastal region are found alder and noncommercial deciduous growth. In the alpine zones of the mountains, larch, mountain hemlock, and alpine firs occur in association, and mountain mahogany is found in the Blue Mountains.

Oregon's animal life is related to its climatic zones. Deer and elk flourish in less populated parts; antelope are found in the eastern high plateau; and bear and fox, in the mountain foothills. The lakes are breeding grounds for waterfowl and resting places for migratory birds.

Settlement patterns. At least five major patterns of land use emerge from the tangle of Oregon's natural landscapes and climates. The forested mountains—the Coast Range, the Cascades, the Klamath, and the Blue–Wallowas—show relatively little evidence of human habitation or modification except for the harvest pattern of clear-cutting in the Douglas fir region, the logging and forest-management roads, and scattered roadside homesites at lower elevations. Most of the few loggers live in the valley towns.

The western valleys, dominated by the Willamette, are Oregon's main centres of population, industry, and transportation. Most persons live close to well-populated centres. The nearly 1,300 wood-product plants that in 1947 were located in valley towns or up tributary valleys into the forested mountains had, by the late 20th century, dwindled to less than 300 large-scale sawmills, plywood plants, and pulp mills.

In the rolling, sparsely populated wheat country of north central Oregon, ranches commonly exceed 1,500 acres (600 hectares) in the eastern portion and double that size to the west, where wheat–fallow rotation is practiced. In regions of natural erosion alternate bands of crop and fallow occur. Farmsteads are widely separated, and owners often live in towns.

The growth of natural feed in open range country is relatively poor, and cattle scatter over enormous areas. There are fences, occasional watering places with metal tanks, and ranchsteads located at great distances from one another.

Most of the eastern Oregon towns except Pendleton lie in the area of irrigated agriculture, on the eastern slopes of the Cascades or near the Idaho border. Farming is highly mechanized.

The people. Oregonians are predominantly white and American-born. There are small populations of Hispanics, blacks, and Asians. American Indians make up about 1 percent of the population. Roman Catholics form the largest single religious denomination but make up only one-third of all religious adherents. Methodists, Baptists, Presbyterians, Disciples of Christ, Lutherans, and Mormons are other large religious denominations.

Metropolitan areas

The people are unevenly distributed, the great majority living west of the crest of the Cascade Range in the Willamette valley. Nearly 60 percent of all Oregonians live in three metropolitan areas, Portland, Eugene, and Salem. Portland, near the confluence of the Willamette and Columbia rivers and the largest city in the state, is a leading West Coast port and the major commercial, industrial, service, and cultural centre of the state. Eugene and Salem, the second and third largest cities, respectively, are important for trade and processing. Salem, the state capital, is among the nation's leading food-processing centres. The major cities outside the Willamette valley are Medford, in the Rogue valley; Klamath Falls, in south central Oregon; and Pendleton, in the north central plateau.

The economy. Traditionally, Oregon has had a resource-oriented economy, strongly dependent upon its forests and farms. Through diversification, however, various new industries have been established and tourism, recreation, and trade and service activities have grown.

Industry. Forest-product manufacturing is Oregon's leading industry. About one-half the land area of the state is forested, and nearly 40 percent produces commercial timber. Public agencies control about 60 percent of Oregon's commercial forest, and private owners the remaining 40 percent. Additional forest is reserved for wilderness preservation, recreation, and other exclusionary uses.

The forest industry began as a producer of lumber: since 1938 Oregon has ranked first in softwood lumber. Products have changed, however, and by the late 20th century only 40 percent of the forest income was from lumber. More than one-third of the logs harvested go into ply-

wood, which accounts for about one-third of the value of forest products. Pulp and paper plants and hardboard and particleboard plants contribute most of the remainder.

Metals-related industries—primary metals, fabricated metals, and transportation equipment—were the pacesetters after World War II. They have been replaced by high-technology industries—machinery, electrical equipment, and instruments—as the major growth factor. The greatest concentration of metals-related industries is in the Portland metropolitan area. The high-technology industries are in Portland and the Willamette valley.

Agriculture and fishing. The agricultural land base of Oregon includes both cropland and pastures and rangeland. Livestock products contribute one-third of the total commodity value, led by cattle and calves; dairy and poultry products are also significant. Wheat is the leading crop, but potatoes, barley, pears, apples, and grapes for wine are also important.

Major crops

Chinook, silver, chum, and pink salmon and shellfish are the most valuable fishery products. Other fish include flounder, tuna, ocean perch, and rockfish.

Mining. In mining, stone and construction sand and gravel make up the bulk of the value. Quarrying occurs in every county, but the greatest quantities are taken near urban areas. The only integrated nickel mine and smelter in the nation is located near Riddle. Studies have shown that the state likely has additional extractable reserves.

Tourism. Tourism has become a major sector of Oregon's overall economy. Those coming to the state enjoy its scenery and myriad opportunities for recreation, including hiking, skiing, fishing, beachcombing, and windsurfing. Tourism supports the many small businesses that provide food, lodging, fuel, and other supplies and services.

Transportation. In addition to an extensive network of highways and roads under the jurisdiction of the state, the federal government, and counties and municipalities, Oregon has forest development roads, national park roads, and military and Indian reservation roads that are controlled by federal agencies and various local governments. Railroads provide north–south and east–west routes. The largest airport is Portland International Airport; other significant commercial airfields are at Eugene, Medford, Pendleton, Klamath Falls, and Redmond.

Throughout the state's history water transportation has been important. Six of the port districts are located on the Columbia above the head of deep navigation, where barge traffic is composed principally of grain and petroleum downstream and cement and structural steel upstream. Portland, open to oceangoing vessels, is by far the most important port. The other districts stretch along the Oregon coast and up the Columbia on the deep-draft channel. Portland, Astoria, Newport, and Coos Bay have regular shipments to and from foreign countries.

Administration and social conditions. *Government.* The state constitution was adopted in 1857. Oregon has been in the vanguard of several innovative movements in U.S. government collectively known as the Oregon System. In 1902 the concepts of initiative and referendum were introduced, by which voters are able to initiate and vote upon statutes or constitutional revisions; these were supplemented in 1908 by the system of recall, under which the removal of elected officials can be initiated by the voters. The state was also one of the earliest to impose a state income tax, which it did in 1923.

The Oregon System

State government in Oregon follows the pattern of most states. Limited to two four-year terms within any 12-year period, the governor supervises the state budget, coordinates the activities of state agencies, boards, and commissions, initiates future planning, and is the focus of federal–state interaction. The governor may also veto individual items in appropriation bills. The legislature comprises the Senate, with 30 members serving four-year terms, and the House of Representatives, with 60 members serving two-year terms.

The court system is headed by the seven-justice Supreme Court, which has general administrative authority over all other courts. The justices, elected for six-year terms, elect one of their members as chief justice.

Oregon gives its towns and cities home rule—that is, the right to choose their own form of government. Most cities with populations of more than 2,500 have the council–manager form of government, whereas smaller cities usually are governed by a city council and a mayor. Portland is governed by four commissioners and a mayor.

In 1958 a constitutional amendment authorized counties to adopt home-rule charters, and in 1973 a state law granted all counties the power to exercise broad home-rule authority. In most counties a county judge and two commissioners or a board of commissioners exercise the powers of government. These officials usually are elected for terms of three years.

Oregon's budget includes general fund revenue derived from personal and corporate income, excise, inheritance, and insurance taxes and from liquor sales and other fund revenue derived from federal grants, use taxes, trust funds, licenses, and the sale of services and commodities. In 1971 the legislature passed a far-reaching program to deal with the problem of air and water pollution, and in 1973 a mandatory program of land and resource development and conservation was established.

Republicans dominated Oregon's politics through much of the state's history. With post-World War II industrial and population growth, however, Democrats came to outnumber Republicans in registration. An unusual number of Oregonians have made their mark in the U.S. Congress by their independent stances.

Education. The first free public school system was created by the territorial legislature in 1849. In 1951 the legislature established a board of education, appointed by the governor. The constitution provides for an elected superintendent of public instruction.

Opportunities for education after high school are provided by community colleges, a state system of higher education composed of three universities (the University of Oregon, Oregon State University, and Portland State University), three regional colleges, two specialized schools, and several private colleges. The community colleges are administered by lay boards, supported by local taxes, and responsive to local needs in their curricula. Private colleges include Reed College (1909) in Portland, Willamette University (1842) in Salem, and Lewis and Clark College (1867) in Portland.

Colleges and universities

Health and welfare. The Department of Human Resources coordinates the activities of the state's principal social service agencies. More than 250 programs provide service directly to citizens. The Oregon Health Sciences University, located in Portland, includes schools of medicine, dentistry, and nursing, hospitals and outpatient clinics, and other facilities. The university's Institute for Advanced Biomedical Research was one of the world's first centres to focus on study of the molecular biology of the brain.

Cultural life. As a relatively young state and one in which the human imprint is scarcely visible over vast stretches of land, Oregon has not developed a cultural identity equivalent to those of the longer settled or more heavily populated regions. Its people, however, no less in the sparsely settled areas of the east than in the population centres of the Willamette valley, take full part in the increasingly homogeneous character of American life. Portland has large auditoriums and a coliseum. Theatrical and musical groups are found in all of the cities and larger towns, and the Oregon Shakespearean Festival in Ashland draws thousands of viewers each summer. University and college communities have public offerings in the arts and other cultural activities.

The cultural milieu

In addition to sporting events, both spectator and participatory, Oregon has a number of attractions related to its history and location. These include the Pendleton Round-Up, which attracts participants from across the West; Albany's World Championship Timber Carnival, which takes place each July 4 and features logger events, carnivals, and a parade; and Portland's Rose Festival in early June.

The Multnomah County Library, in Portland, was the first to serve the public on a large scale; it began membership service in 1864 and free service in 1902. The Oregon State Library in Salem maintains a general reference service and loan collection.

The Oregon Historical Center in Portland and the Horner Museum at Oregon State University own large collections of items from pioneer days in the Oregon country. The Oregon Museum of Science and Industry in Portland features demonstrations of science at work in Oregon industries. The Portland Art Museum features Northwest Coast Indian art and pre-Columbian art in its collection. The Murray Warner Collection of Oriental Art at the University of Oregon has one of the largest Asian collections in the United States. The High Desert Museum, in Bend, has living exhibits of plants and animals native to the arid region of the Pacific Northwest.

HISTORY

The Indian cultures of Oregon

When the first Europeans arrived in the Oregon country—a region vaguely defined at the time but roughly comparable to the present Pacific Northwest—about 125 Indian tribes with a population estimated at 100,000 to 180,000 lived in and around the area. In what became the state of Oregon, the leading tribes were the salmon-eating Chinook along the lower Columbia River; the Tillamook, Yamel, Molala, Clackamus, and Multnomah in the northwest; the Santiam and Coos in the southwest; the Cayuse, Northern Paiute, Umatilla, Nez Percé, and Bannock in the dry lands east of the Cascade Range and in the Blue–Wallowa mountains; and the Klamath and Modoc in the south central area. Their mode of life resulted in a relatively small population: they had no form of agriculture and no domesticated animals other than the dog; and they used crude implements for gathering, hunting, and fishing. The tribes along the Columbia River, known as the Canoe Indians, fashioned excellent canoes from logs.

The explorers. The first Europeans to see the Oregon coast were Spanish sailors searching for a northwest passage. In 1579 English pirate Francis Drake, in quest of Spanish loot and a northwest passage in his *Golden Hind,* anchored in an inlet north of the Golden Gate and with a brass plate "took possession" of the country for Queen Elizabeth I. Until the third quarter of the 18th century, when the Spanish renewed exploration along the coast, the Oregon country remained unexplored. In 1778 the English sea captain James Cook visited and traded in Oregon.

In 1787 Boston merchants sent two ships to the Oregon country under Captains Robert Gray and John Kendrick. On his second voyage Gray entered the harbour that bears his name (in Washington), and in May 1792 he sailed over the bar of the Columbia River and named it for his ship, the *Columbia.* This was the first U.S. claim to the Pacific Northwest by right of discovery.

The Northwest was also approached by land. Two British fur companies, the Hudson's Bay Company and the North West Company, raced across the continent to open routes to the Pacific; the Americans were not far behind. Meriwether Lewis and William Clark reached the mouth of the Columbia in 1805, strengthening the U.S. claim to the region. John Jacob Astor, at the head of the Pacific Fur Company, began white settlement of the Oregon country with the establishment of a trading post at Astoria in 1811. In 1824 the Hudson's Bay Company established Fort Vancouver, and John McLoughlin was appointed to head this company's far-flung operations. For the next 22 years he was the dominating figure in the region.

Permanent settlement. Beginning in 1830, thousands of people from the Midwest migrated to the Pacific Northwest. Missionaries played a role in settlement. In 1834 the Methodists, headed by Jason Lee, established the first permanent settlement in the Willamette valley. The migrations that carved the deep wagon wheel ruts still visible in the Oregon Trail began in the early 1840s. After 1838 U.S. claims and rights to the region were constantly before Congress. Settlers in the Willamette valley made known their desire to become part of the United States. In 1843 representatives met at Champoeg to organize a provisional government; a set of laws patterned after those of Iowa was accepted. By 1844 the British government had concluded that the Columbia River boundary line would have to be abandoned, and the Hudson's Bay Company moved its chief Northwest depot to Fort Victoria. In spite of the "fifty-four forty or fight" slogan of the presidential campaign of 1844, the 49th parallel was accepted by both nations as the boundary, and the Oregon country was added to the United States in 1846.

Disputes over ownership

Statehood and growth. The influx of population led to political agitation, and in 1853 the Washington Territory was given independent status. Oregon became the 33rd state in 1859. By 1883, following several conflicts with whites, most of the Indians of Oregon had been moved to reservations. The same year a railroad linking Oregon with the rest of the nation was begun, which vastly improved the opportunity for economic growth. Agriculture and forestry were especially stimulated, and by the turn of the 20th century two-thirds of the people of Oregon lived in rural areas. During the 20th century, however, the cities grew rapidly, and by the late 20th century more than two-thirds of the people were living in urban areas. Land use planning and legislation have helped to preserve the state's environment. Since 1940 there has been diversification of the economy, including rapid growth in international trade, and a significant increase in the number of people emigrating to Oregon from other states. (R.M.Hi.)

Washington

Admitted on Nov. 11, 1889, as the 42nd member state of the United States, Washington lies at the northwest corner of the 48 coterminous states. It is bounded by the Canadian province of British Columbia on the north, Idaho on the east, Oregon on the south, and the Pacific Ocean on the west. A coastal location and excellent harbours give the state a leading role in trade with Alaska, Canada, and countries of the Pacific Rim. Washington cities have sister cities in several countries, and their professional and trade associations commonly include Canadian members.

Washington's area of 68,139 square miles (176,479 square kilometres) is smaller than those of its border states, Idaho and Oregon, but it has the largest population of the three. Most of the people live in the metropolitan areas of Seattle–Everett and Tacoma and other cities along Puget Sound, including the state capital, Olympia.

The terrain and climate of Washington divide the state into a rainy western third and a drier eastern two-thirds in the rain shadow of the Cascade Range. Western Washington industries depend on agriculture, forests, and fisheries and imported raw materials, whereas eastern Washington is mainly agricultural, producing wheat, irrigated crops, and livestock.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Washington has seven physiographic regions. In the northwest the Olympic Peninsula borders the Pacific Ocean south of the Strait of Juan de Fuca. Dense rain forests extend along the western slopes of the rugged Olympic Mountains, which rise to 7,965 feet (2,428 metres) on Mount Olympus.

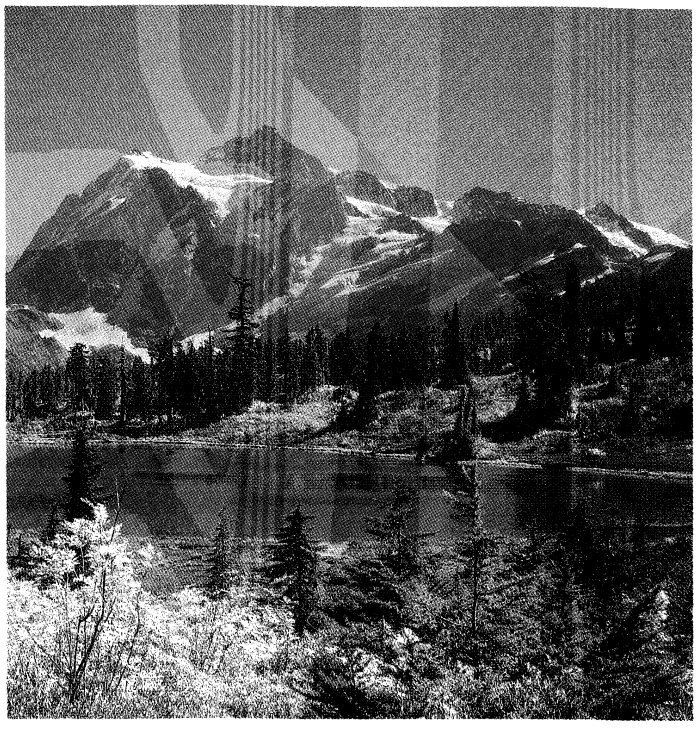
The Willapa Hills parallel the coast from Grays Harbor to the Columbia River in the southwest. Gentle, forested slopes descend to an indented Pacific coastline and to the Chehalis and Cowlitz valleys on the north and east.

The Puget Sound Lowland stretches southward from Canada between the Olympic Mountains and the Cascade Range to join the Chehalis and Cowlitz valleys, which form an extension to the Columbia River. Deep waters and fine harbours in Puget Sound, together with relatively flat terrain along its shores, favour the densest population and greatest commercial development in the state.

The Cascade Range

The Cascade Range, east of the Puget Sound Lowland, has the state's highest elevations. Its chain of volcanic peaks includes 14,410-foot (4,392-metre) Mount Rainier, the fifth highest peak in the coterminous United States. Mount St. Helens, located in the Cascades near the Oregon border, erupted violently in 1980. The highest peaks have permanent glaciers.

The Columbia Basin occupies most of central Washington, surrounded by the Cascades to the west, the Okanogan Highlands to the north, uplands to Idaho on the east, and the Blue Mountains to the southeast. A basalt plateau, lying at about 1,000 to 2,500 feet above sea level, it is drained by the Columbia River and its main tributary, the



Picture Lake and Mount Shuksan in North Cascades National Park, Washington.
Bob and Ira Spring

Snake. Glaciation, flooding, and wind have shaped diverse landforms, although the general appearance is that of a large interior plain.

The Okanogan Highlands, in the northeast, are an extension of the Rocky Mountains. Their north–south ranges, with summits that rise to more than 7,000 feet (2,100 metres), are separated by glaciated trenches. Most of the state's metallic ores are found in this region.

The Blue Mountains, which extend into Washington from Oregon, consist of uplifted plateaus and ranges in the southeast corner of the state. Gentle slopes and broad valleys descend from 6,000-foot (1,800-metre) heights to the Columbia Basin. Outliers to the west comprise the Horse Heaven Hills and Rattlesnake Hills.

Soils. The most productive soils in Washington are those of the river floodplains and the weathered basalts and windblown silts of the Columbia Basin. In wetter areas acidic soils support forests, but the driest regions east of the Cascades have sparse plant life and require irrigation for agriculture. The fine-textured soils of the Big Bend and Palouse areas are susceptible to erosion by wind and water.

Climate. Prevailing westerly winds and the influence of the Pacific Ocean dominate the climate of Washington, although the Cascades barrier creates significant differences between western and eastern regions. The west has milder conditions than any part of the United States at the same latitudes. Seattle has an average January temperature of 41° F (5° C) and a 66° F (19° C) July average. Annual precipitation on the Pacific slopes of the Olympic Peninsula exceeds 150 inches (3,810 millimetres), but places on the northwest of the peninsula receive less than 20 inches (508 millimetres) a year. From the Puget Sound Lowland, where 30–40 inches are typical annual totals, amounts increase again to more than 100 inches in the Cascades.

Differences between west and east

East of the Cascade Range seasonal temperature variations are greater, but the Rocky Mountains to some extent shield the region from cold Canadian air masses in winter. Maximum summer temperatures usually exceed 100° F (38° C) a few days each year. Spokane's January average temperature is 25° F (−4° C); the July average is 70° F (21° C). Annual precipitation is 17 inches (430 millimetres) at Spokane but less than eight inches (200 millimetres) in the lower Yakima valley.

Throughout the state precipitation is greatest in the cooler months, when a succession of cyclonic storms move inland from the North Pacific, sometimes with gale-force winds. Rain falls on a great number of days even in areas that are relatively arid, such as in the west. The occasional outbreaks of continental air from the north or northeast may reach the outer coast, bringing freezing conditions in winter or hot, dry air that increases the danger of forest fires in summer.

Plant and animal life. Washington's 23,000,000 acres (9,308,000 hectares) of forest are among the most extensive in the United States. Major tree species are Douglas fir, hemlock, western red cedar, and ponderosa pine, found mainly in the mountain regions. On the semiarid parts of the Columbia Basin, grasses prevail, merging into sagebrush and other scattered shrubs in the driest areas.

Deer, elk, bears, mountain goats, and pumas (cougars) are among the large mammals, and there are also several fur-bearing animals. The Pacific flyway, a major route of North American waterfowl migration, follows the Puget Sound Lowland. Freshwater game fish include trout, bass, grayling, and sturgeon. Five species of Pacific salmon ascend western Washington streams to spawn. The coastal bays and Puget Sound are habitats for shellfish.

Settlement patterns. About three-fourths of Washington's people live in urban areas, principally in the Puget Sound Lowland. More than 50 percent live in the Seattle and Tacoma metropolitan areas. Spokane is the largest city east of the Cascades and the focus of the "inland empire," a large economic region of agriculture, forestry, and mining that reaches to northeastern Oregon, northern Idaho, western Montana, and southern British Columbia, Can. Smaller cities of eastern Washington include agricultural trade centres such as Wenatchee, Yakima, and Walla Walla. The Tri Cities area (Richland–Kennewick–Pasco) at the confluence of the Snake and Columbia rivers forms a transportation centre for irrigated agriculture, manufacturing, and the Hanford Site (an atomic energy installation).

Cities of Washington

Typical towns of the eastern wheat lands are crowned by grain elevators, whereas food processing plants are common in the towns that serve irrigated farms. Lumber towns and small mining settlements are found along the upland margins of the Columbia Basin.

The people. The early settlers, from the 1830s through the 1850s, came primarily from the Midwest along the Oregon Trail. Growth was slow until the 1880s, when railroads began to link Puget Sound and the Columbia River to the East and to California, ending the frontier era of the Pacific Northwest. The population of Washington grew fivefold from 1881 to 1890, to almost 360,000—and by 1920 it reached almost 1,360,000.

Immigration continued, particularly from the Midwest, and, until national quotas on foreign immigration of the 1920s, large numbers of foreign-born people entered the state, especially from Canada and the Scandinavian countries. The Japanese came late and by 1930 numbered about 18,000. During World War II, citizens or not, they were moved from the coastal areas to relocation camps in inland regions. After the war only a few received back their homes and property, and many chose to live elsewhere.

Washington has a relatively small percentage of blacks. It ranks among the top 10 states, however, in numbers of American Indians and Asians.

For decades the western movement of the nation's population dominated Washington's growth. During the 1950s, however, for the first time and by a wide margin, natural increase overtook immigration. Immigration has regained some of its former importance, but it remains below natural increase as a growth factor.

The economy. Agriculture, forestry, and fisheries have been major contributors to the state's economy since early settlement by Europeans. The rapid increase in manufacturing and services that began in the 1940s led to concentration of the population in urban areas. Nearly one-fifth of the nonagricultural labour force is employed in manufacturing; another one-fifth works for state or federal government agencies.

Resources. Water is Washington's most valuable and most versatile natural resource. The leading freshwater

source is the series of dams on the Columbia River drainage system that impound water for irrigation, hydroelectric power, and flood control, while also providing for navigation, fisheries, recreation, and industrial uses. The Columbia and rivers of western Washington account for one-third of all hydroelectric production in the United States. Grand Coulee Dam, with a capacity of about 6,500,000 kilowatts, ranks among the largest power plants in the world. Groundwater resources are exploited for domestic use, industry, and limited irrigation in the Puget Sound Lowland and, to a lesser extent, along the main river valleys of the Columbia Basin.

Grand Coulee Dam

Forests support both wood-product industries and wildlife and recreation. Multiple use and sustained yield have been primary management objectives on both private and public forestlands since early in the 20th century. Commercial fisheries are another significant sector in the state's economy. Salmon, halibut, cod, and herring are the principal species landed at ports on Grays Habor, Willapa Bay, and Puget Sound. Developments in aquaculture supplement the harvest with salmon, trout, and shellfish.

Sand, gravel, and clay are the most valuable of the state's limited mineral products. Magnesite, lead, and zinc are produced in the Okanogan Highlands; and coal mining in the Cascades and Puget Sound Lowland has declined during the 20th century. An open-pit coal mine near Centralia provides fuel for a thermoelectric power plant.

Agriculture. Winter wheat is the state's leading crop and a major export from the Columbia Basin, which also grows barley, dry peas, lentils, and hay on dryland farms. Irrigated crops include potatoes, vegetables, fruits, hops, and mint. Washington markets more apples than any other state and is a major producer of pears, cranberries, and wine grapes. Vegetable seeds, berries, vegetables for canning or freezing, and flower bulbs are specialties of the Puget Sound Lowland.

Dairying is a leading rural industry of the northern Puget Sound Lowland, which is also noted for poultry. Beef cattle and sheep graze on the eastern grasslands and the open forestlands of mountain regions.

Farms vary from a few acres to hundreds of acres; since the mid-20th century the tendency has been toward larger and fewer farms. Former agricultural land near large cities has been converted to urban use at an increasing rate.

Manufacturing. For more than a century agriculture, forestry, fisheries, and mining have furnished materials for Washington's processing plants. By the mid-20th century, aircraft and aerospace production in the Seattle area rose to first place among the state's fabricating industries. U.S. Navy facilities on Puget Sound provide for construction and repair of ships; a major installation is the Trident Nuclear Submarine Base near Bremerton.

The state's several aluminum refineries depend on hydroelectricity and imported alumina to produce about one-fourth of the primary aluminum in the United States. Petroleum refineries on northern Puget Sound process Alaskan and foreign crude oil.

Tourism. Tourism has become a major source of income in Washington. The variety of scenic areas, including three national parks, draws increasing numbers of visitors to the state. Boating, hiking, skiing, sports events, and local festivals are other major tourist attractions.

Transportation. Harbours on Puget Sound and the outer coast afford year-round access to world ocean routes, and a state ferry system serves the San Juan Islands and Canada's Vancouver Island. Navigation locks allow boats to pass between Puget Sound and Lake Washington, at Seattle. Barges carry grain and raw materials along the Columbia–Snake route.

Ocean navigation

Airlines link the state's cities with one another and with transcontinental and world air routes. The Seattle–Tacoma Airport ranks among the leading U.S. airports in international passenger travel.

The state has a well-developed system of highways and interstate freeways. Pontoon bridges span Hood Canal on the Olympic Peninsula and Lake Washington at Seattle. Railways crisscross the state but rank behind trucks in freight transport. Pipelines move oil and natural gas from out-of-state and distribute refined products.

Administration and social conditions. *Government.* Washington's constitution of 1889, reflecting the distrust of government that was characteristic of the time, contained many restrictions on state power. One reflection of this was the creation of a divided executive. Unlike the federal executive branch, to which only the president and vice president are elected, the state has nine separately elected officials. The most important is the governor.

The legislature comprises the Senate of 49 members and the House of 98 members, elected, respectively, for four- and two-year terms. Important limitations on legislative powers include the earmarking of certain funds to specific purposes—*e.g.,* the gasoline tax to highways. Because the constitution prohibits a state income tax, Washington depends on more than one-half of its tax revenues from a general sales tax, which accounts for about 30 percent of the state's total income. Initiative and referendum on legislation and recall of elected officials give the voters a check on the legislature. The governor's power of "item veto" has been expanded to include all legislation, except referendums and initiatives, to the extent of eliminating lines in budget acts or sections of other laws.

The courts are divided into four levels. Courts of limited jurisdiction—justice, municipal, and police—are local and hear traffic cases, minor criminal and civil cases, and small-claims actions. Superior courts are general trial courts, having original jurisdiction in felonies and in civil cases not delegated to the limited courts. The Supreme Court and the appellate courts are almost solely courts of review. All judges, except for some classes of appointed municipal and police judges, are elected on nonpartisan ballots. Grand juries, created by a superior court, are used mainly to investigate political corruption, though their legal powers are considerably broader.

Washington's 39 counties are classified according to population by the legislature. The governing body in most counties is the board of county commissioners, whose three members act as both the chief executive officers and the legislative body for the county. The Optional Municipal Code was adopted in 1969, substantially expanding the powers of cities choosing to come under it. Cities with populations of 10,000 or more can adopt a home-rule charter if such a referendum is approved by the electorate, while municipalities of 300 to 10,000 are granted optional, noncharter home rule by statute.

Elections and political parties are regulated by state law. The unique feature of the nomination process in Washington is the "blanket primary," which replaced the closed primary in 1935, permitting citizens to vote for any candidate without disclosing their party membership. This law reflects a characteristic independence among the state's voters. Split voting has been reported by three-fourths of the voters in both primaries and final elections.

The blanket primary

Education. The State Board of Education sets general requirements of public school curricula, which are administered by an elected superintendent of public instruction and more than 300 district school boards. Attendance is required for children age eight through 16. Higher education is predominantly a state function, the largest institution being the University of Washington in Seattle, established in 1861. Washington State University at Pullman was founded in 1890 as a land-grant college for agricultural and mechanical arts. Three state colleges—at Bellingham, Ellensburg, and Cheney—evolved from teacher-training institutions in the 1890s to university status in 1977, and Evergreen State College at Olympia was added in 1971. A system of community colleges was combined under state administration in 1967. Several private, denominational institutions augment postsecondary opportunities.

Health and welfare. In 1936, responding to the Social Security Act of 1935, the state assumed broad responsibilities for welfare programs. The Department of Social and Health Services administers benefits for children, the aged, and families; it oversees both private and public medical services, including Medicaid. Washington ranks among the top 10 states in public aid to families with dependent children. Separate agencies provide aid to the blind and veterans. There are also commissions for Human Rights and Insurance Consumer Protection. An Employment

Security Department assists those who seek jobs and disburses unemployment insurance payments.

Cultural life. A young state, Washington has a Western and pioneering outlook. There is great interest among the people in archaeological explorations and the cultural patterns of the Indian inhabitants. Interdisciplinary field studies by scientists at Washington State University have discovered artifacts of two quite different archaeological sites. Marmes Rock Shelter, in arid eastern Washington, has yielded a 10,000-year sequence of tools left by hunters and gatherers along with some of the oldest well-documented skeletal remains in the Western Hemisphere. The Ozette site, on the Olympic Coast, has a unique collection of well-preserved clothing, basketry, and harpoons of people who fished and hunted seals and whales 500 years ago. Tools of a similar culture dating from 2,000 years ago were also found there. These and other sites in the state reflect the diverse cultural forms that evolved after prehistoric migrations from northeastern Asia.

Marmes Rock Shelter and Ozette

The arrival of European settlers in the 19th century not only transformed the cultural landscape of Washington but also introduced new social patterns. Contemporary outdoor events usually are based on local history or economic pursuits. Rodeos (mainly in eastern Washington) and "old settlers reunions" recall early struggles to occupy the land. Agricultural fairs, ethnic festivals, blossom festivals, and parades exhibit products and skills. The annual Seattle Seafair features parades, boat races, and water carnivals. Water sports are popular on many lakes and rivers and especially on Puget Sound. Skiing is a favourite winter sport in the Cascades and Okanogan Highlands. Public forestlands, three national parks, and more than 125 state parks attract campers during the summer months.

Washington residents pursue a wide range of interests in the fine arts. The Seattle Symphony, Seattle Opera Association, and the Seattle Repertory Theater draw national attention. The School of Drama at the University of Washington pioneered modern arena staging. Several smaller cities have orchestras and drama groups. Among the approximately 25 major art galleries and museums, the Seattle Art Museum has extensive collections of Oriental art. The Thomas Burke Memorial Washington State Museum on the University of Washington's Seattle campus has an important collection relating to Pacific Northwest Indians. Also of interest are the Museum of History and Industry and the Pacific Science Center, both in Seattle. Among important historical museums is the Washington State Historical Society Museum in Tacoma.

The first library in the state, the Washington State Library in Olympia, traces its financial aid by the Congress to part of the Territorial Act of 1853. Rapid development of the public library system occurred in the first decade of the 20th century, when the Carnegie Foundation provided building funds. More than 250 libraries serve the state. Noteworthy are the two state university libraries, the Washington State Archives, and the Seattle Public Library.

HISTORY

The early frontier. When Europeans first explored the Washington area, they encountered a number of Indian tribes, the most prominent being the Chinook, the Coast Salish, the Nez Percé, and the Yakima. The early history of Washington and of the Northwest is intertwined with efforts to find the Northwest Passage, the development of the fur trade with the Orient, and the attempts of Roman Catholic and Protestant missionaries to convert the Indians. Spaniards had sailed along the coast earlier, but the wealth of sea otter skins secured from the Indians on one of the voyages of Captain James Cook in 1778 marked the start of real exploration and of the maritime fur trade. George Vancouver, sent by Britain in 1792, tried to find the Northwest Passage and to map the coast. Robert Gray was the first trader from the United States; his explorations resulted in the discovery of the Columbia River in 1792. By 1812, the United States almost completely dominated the fur trade. The British Hudson's Bay Company, however, maintained areas of dominance into the 1840s.

Explorers, trappers, missionaries, and settlers

Missionaries were generally welcomed by the Indians, though often not so much for Christian salvation as for the knowledge and material advantages the whites could bring. Among the most famous missions were those of the medical missionary Marcus Whitman and the Reverend Henry Spalding, established in 1836 in southeastern Washington, and the Roman Catholic missions established by Pierre-Jean DeSmet in northeastern Washington.

The Protestant missionaries felt that white civilization was necessary for the Indian and thus encouraged white settlement. With the opening of the Oregon Trail the first large group, about 1,000 people, reached the Northwest in 1843. These and others following first went mainly into the Willamette valley of what became the state of Oregon, and later into the area north of the Columbia River (in present-day Washington), then still dominated by the Hudson's Bay Company. The Indians were initially receptive but the settlers' inconsistent dealings with the Indians led to such conflicts as the Cayuse War (1848–50), the Yakima War (1855–58), and the Nez Percé War (1877).

By the end of the 19th century most of the Indians had been settled on reservations, representing three principal tribal groups: Coast Salish, Interior Salish, and the Sahaptin. Anthropologists have identified numerous distinct tribes on the basis of language and other local cultural characteristics. Among the larger tribes of western Washington are the Makah, Quinault, Lummi, Snohomish, and Puyallup; tribes of eastern Washington include the Okanogan, Yakima, Klickitat, Kalispel, and Spokane.

Territory and state. Until the 1840s citizens of both the United States and Britain by agreement could settle and trade in what was still known as the Oregon country. In 1846 the two countries agreed on the present boundary between the United States and Canada, and in 1848 Congress established the Oregon Territory including all of the present-day states of Oregon, Washington, and Idaho and parts of Wyoming and Montana. This enormous area was difficult to govern from the territorial capital in the Willamette valley. As the population around Puget Sound grew, agitation arose to form a separate territory of the area north and west of the Columbia. In 1853 Congress created the Washington Territory—named for the first president of the United States—and extended it east of the Columbia River to the crest of the Rockies, including parts of present-day Idaho and Montana.

Establishment of Washington Territory

Different rates of population growth and difficulties of communication continued to cause problems, and various movements called for the creation of a separate territory in eastern Washington and even the creation of an independent Pacific Republic. In the 1870s and '80s the extension of the telegraph and the railroads to the Northwest strengthened ties with the United States, and attention turned to seeking statehood, granted in 1889.

Gold discoveries in the interior in the 1850s made Walla Walla the centre of eastern Washington for a time, but these were merely a prelude to Washington's role in provisioning the gold seekers who set out for the Alaskan and Yukon strikes of the late 1890s. The gold stimulated the trade of cities on Puget Sound, and the new prosperity was celebrated at the Alaska–Yukon Exposition in 1909.

Possibly the greatest stimulus to the state's progress in the 20th century was initiated by the development of the Columbia Basin and related projects, which greatly increased hydroelectric power and provided the basis for increased irrigation and flood control. A navigation improvement project was authorized as early as 1911, and work began on the Bonneville and Grand Coulee dams in 1933. Construction was completed on Bonneville in 1937 and on the main structure of Grand Coulee in 1941. The first two Grand Coulee power plants were completed in 1951, and a third power plant began operation in 1975.

Washington's prosperity and its growing role in the commerce of the Pacific were among the features celebrated in the Seattle World's Fair of 1962, named the Century 21 Exposition. Developments in the latter half of the 20th century were increased urbanization, consolidation of agricultural landholdings, improved transportation networks, and expanded trade with the Pacific Basin countries. Increasing concern for the environment led to a series of laws to regulate the impact of a growing population.

(E.Cl./H.J.Cr.)

BIBLIOGRAPHY

Physical and human geography. *The land:* (*Landforms and geology*): The standard work on the landform regions of the United States is still WILLIAM D. THORNBURY, *Regional Geomorphology of the United States* (1965), superbly illustrated. WALTER SULLIVAN, *Landprints: On the Magnificent American Landscape* (1984), while journalistic in tone, is a lively, authoritative, and well-illustrated treatment. An elementary, illustrated textbook is E.C. PIRKLE and W.H. YOHO, *Natural Landscapes of the United States,* 4th ed. (1985). NEVIN M. FENNEMAN, *Physiography of Western United States* (1931), and *Physiography of Eastern United States* (1938), are exhaustive and still standard references. WILLIAM L. GRAF (ed.), *Geomorphic Systems of North America* (1987), is a highly technical discussion. Recommended atlases include CHARLES O. PAULLIN, *Atlas of the Historical Geography of the United States* (1932, reprinted 1975); and GEOLOGICAL SURVEY (U.S.), *The National Atlas of the United States of America* (1970).

(*Climate*): STEPHEN S. VISHER, *Climatic Atlas of the United States* (1954, reprinted 1966), contains more than 1,000 maps. UNITED STATES NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, *Climates of the United States,* 2nd ed., 2 vol. (1980), makes available physical and climatic data in narrative, tabular, and map form. JOHN L. BALDWIN, *Climates of the United States* (1973), is a concise treatment. Scholarly discussions are found in REID A. BRYSON and F. KENNETH HARE (eds.), *Climates of North America* (1974).

(*Plant and animal life*): An authoritative regional treatment of plant and animal ecology is VICTOR E. SHELFORD, *The Ecology of North America* (1963, reprinted 1978). MICHAEL G. BARBOUR and WILLIAM DWIGHT BILLINGS (ed.), *North American Terrestrial Vegetation* (1988), covers all major types. The relationship between climate and natural vegetation is obvious but far from simple; the most ambitious cartographic attempt to correlate them in a North American setting is explained in ROBERT G. BAILEY (comp.), *Description of the Ecoregions of the United States* (1978).

(*Human geography*): Of the several general texts covering the human geography of the continent, perhaps the most stimulating and informative is J. WREFORD WATSON, *North America, Its Countries and Regions,* rev. ed. (1967). D.W. MEINIG, *The Shaping of America,* vol. 1, *Atlantic America, 1492–1800* (1986), is indispensable for an understanding of the origins of America's human geography. A general discussion of the traditional regions of the country is found in WILBUR ZELINSKY, *The Cultural Geography of the United States* (1973). JOEL GARREAU, *The Nine Nations of North America* (1981), is a lively, highly readable description of the emerging socioeconomic regions.

(*Landscape and land use*): STEPHEN S. BIRDSALL and JOHN W. FLORIN, *Regional Landscapes of the United States and Canada,* 3rd ed. (1985), is a general introduction. JOHN R. STILGOE, *Common Landscape of America, 1580 to 1845* (1982), offers a valuable account of the early evolution of settlement. JOHN BRINCKERHOFF JACKSON, *Discovering the Vernacular Landscape* (1984), delves into the meaning of everyday man-made environments. The growth and development of America's cities and towns are detailed in ALEXANDER B. CALLOW, JR. (ed.), *American Urban History,* 3rd ed. (1982); and RICHARD LINGEMAN, *Small Town America: A Narrative History, 1620–the Present* (1980). (P.F.L./W.Ze.)

The people: The *Statistical Abstract of the United States,* published annually by the UNITED STATES BUREAU OF THE CENSUS, is the standard summary of statistics on the country's social, political, and economic composition. Interpretations of demographic data include EDWARD G. STOCKWELL, *Population and People* (1968); and RICHARD M. SCAMMON and BEN J. WATTENBERG, *The Real Majority* (1970). For an analysis of national values, a classic account is GUNNAR MYRDAL, *An American Dilemma: The Negro Problem and Modern Democracy,* 2 vol. (1944, reprinted 1975). OSCAR HANDLIN, *The Uprooted,* 2nd ed. enlarged (1973), covers the era of mass immigration, 1860–1920. Examinations of contemporary American minority groups include STEPHAN THERNSTORM (ed.), *Harvard Encyclopedia of American Ethnic Groups* (1980); and FRANK D. BEAN and W. PARKER FRISBIE (eds.), *The Demography of Racial and Ethnic Groups* (1978). Ethnic patterns as of 1980 are treated definitively in JAMES PAUL ALLEN and EUGENE JAMES TURNER, *We the People: An Atlas of America's Ethnic Diversity* (1988). (O.H.)

The economy: JOHN AGNEW, *The United States in the World* (1987), analyzes the growth of regional economic blocs and their effects on the U.S. economy. ANTHONY S. CAMPAGNA, *U.S. National Economic Policy, 1917–1985* (1987), is a history of changes in America's economic policies over nearly 70 years. Aspects of the economy are treated in HOWARD F. GREGOR, *Industrialization of U.S. Agriculture* (1982), an atlas emphasizing aspects of U.S. industrialized farming, mainly from census information; ROBERT J. NEWMAN, *Growth in the American South* (1984), on the shift of U.S. manufacturing to the Southern states in the 1960s and '70s; and RUTHERFORD H. PLATT and GEORGE MACINKO (eds.), *Beyond the Urban Fringe* (1983), essays on the vast land acreage outside U.S. metropolitan areas.

Administration and social conditions: *The United States Government Manual* (annual) offers a broad overview of the federal structure; while the *Congressional Quarterly Weekly Report* and *National Journal* (weekly) provide closer views of the public record of the federal legislature. *Congressional Quarterly's Guide to Congress,* 3rd ed. (1982), details the development and present organization of Congress. See also *The Book of the States,* published biennially by the COUNCIL OF STATE GOVERNMENTS. DONALD R. WHITNAH (ed.), *Government Agencies* (1983), contains essays on the agencies' purposes and histories, with bibliographies. Discussions of election politics include FRED I. GREENSTEIN and FRANK B. FEIGERT, *The American Party System and the American People,* 3rd ed. (1985); the series by THEODORE H. WHITE, begun with *The Making of the President, 1960* (1961), which continued by covering subsequent presidential elections; and JACK P. GREENE (ed.), *Encyclopedia of American Political History,* 3 vol. (1984). ALEXANDER DeCONDE (ed.), *Encyclopedia of American Foreign Policy,* 3 vol. (1978), also contains useful bibliographies. NEAL R. PEIRCE and JERRY HAGSTROM, *The Book of America: Inside 50 States Today,* rev. and updated ed. (1984), provides an insightful look at persistent social differences among various regions of the country. (Ed.)

Cultural life: The classic statement of the American vision in literary criticism is LIONEL TRILLING, *The Liberal Imagination* (1950, reissued 1976). See also LESLIE FIEDLER, *What Was Literature?* (1982), a radical egalitarian polemic against the division of American literature into "high" and "low" forms. EMORY ELLIOTT (ed.), *Columbia Literary History of the United States* (1988), covers the many aspects of American literature. DANIEL HOFFMAN (ed.), *Harvard Guide to Contemporary American Writing* (1979), is the best general introduction. Useful works on art include DORE ASHTON, *American Art Since 1945* (1982); IRVING SANDLER, *The Triumph of American Painting: A History of Abstract Expressionism* (1970, reissued 1982); and MILTON W. BROWN *et al., American Art* (1979). The renaissance of American dance has produced two great dance critics, Arlene Croce and Edwin Denby. Their works include ARLENE CROCE, *Going to the Dance* (1982), and *Sight Lines* (1987); and EDWIN DENBY, *Dance Writings* (1986). For a history of America's unique contribution to the theatre arts, see GERALD BORDMAN, *American Musical Theater,* expanded ed. (1986). JAMES AGEE, *Agee on Film,* vol. 1, *Reviews and Comments* (1958, reprinted 1983), is still the most eloquent writing about American movies. STEPHEN MAMBER, *Cinema Verite in America: Studies in Uncontrollable Documentary* (1974), is a good introduction to alternative theories about alternative film. H. WILEY HITCHCOCK and STANLEY SADIE (eds.), *The New Grove Dictionary of American Music,* 4 vol. (1986), is an excellent starting point for research. GILBERT CHASE, *America's Music, from the Pilgrims to the Present,* rev. 3rd ed. (1987), is invaluable and readable. (A.Go.)

History. Among the many overviews of U.S. history, the following are representative: SAMUEL ELIOT MORISON, HENRY STEELE COMMAGER, and WILLIAM E. LEUCHTENBURG, *The Growth of the American Republic,* 7th ed. (1980); and JOHN A. GARRATY and ROBERT A. MCCAUGHEY, *The American Nation,* 6th ed. (1987). Reference sources include *Dictionary of American History,* rev. ed., 8 vol. (1976–78); and RICHARD B. MORRIS (ed.), *Encyclopaedia of American History,* 6th ed. (1982).

Discovery and exploration: Useful introductions include SAMUEL ELIOT MORISON, *The European Discovery of America,* 2 vol. (1971–74); and DAVID B. QUINN, *North America from Earliest Discovery to First Settlements: The Norse Voyages to 1612* (1977). (Ed.)

Colonial development to 1763: CHARLES M. ANDREWS, *The Colonial Period of American History,* 4 vol. (1934–38, reprinted 1964), is the starting point for an understanding of the structure of the British Empire in America. LAWRENCE HENRY GIPSON, *The British Empire Before the American Revolution,* 15 vol. (1936–70), represents the culmination of the "British Imperial" school of interpretation. GARY B. NASH, *Red, White, and Black: The Peoples of Early America,* 2nd ed. (1982); and JACK P. GREENE and J.R. POLE (eds.), *Colonial British America* (1984), are excellent surveys. (*Settlement*): PERRY MILLER, *The New England Mind: The Seventeenth Century* (1939, reissued 1983), and a sequel, *The New England Mind: From Colony to Province* (1953, reissued 1967), together constitute perhaps the finest work of intellectual history ever written by an American historian. FRANCIS JENNINGS, *The Invasion of America* (1975); and JAMES AXTELL, *The European and the Indian* (1982), are important accounts of white–Indian relations. (*Imperial organization*): Useful surveys include MICHAEL KAMMEN, *Empire and*

Interest: The American Colonies and the Politics of Mercantilism (1970); and STEPHEN SAUNDERS WEBB, *1676, the End of American Independence* (1984). (*The growth of provincial power*): JAMES A. HENRETTA, *The Evolution of American Society, 1700–1815* (1973), is an excellent survey of the American economic and political order. JACK P. GREENE, *Pursuits of Happiness* (1988), seeks to demonstrate the variety of colonial social developments. CARL BRIDENBAUGH, *Myths and Realities: Societies of the Colonial South* (1952, reprinted 1981), argues persuasively that the colonial South consisted of not one but three sections. RHYS ISAAC, *The Transformation of Virginia, 1740–1790* (1982), imaginatively surveys the social order of 18th-century Virginia. GARY B. NASH, *The Urban Crucible: Social Change, Political Consciousness, and the Origins of the American Revolution* (1979), surveys the growth of American cities in the 18th century. JOHN J. MCCUSKER and RUSSELL R. MENARD, *The Economy of British America, 1607–1789* (1985), is a good survey. (*Cultural and religious development*): DANIEL J. BOORSTIN, *The Americans: The Colonial Experience* (1958, reissued 1988), gives a brilliant, if overstated, account of American uniqueness. HENRY F. MAY, *The Enlightenment in America* (1976), provocatively examines American intellectual development. See also BROOKE HINDLE, *The Pursuit of Science in Revolutionary America, 1735–1789* (1956, reprinted 1974). ALAN HEIMERT, *Religion and the American Mind, from the Great Awakening to the Revolution* (1966), makes an important though polemical contribution to the understanding of the Great Awakening. (*America, England, and the wider world*): Overviews are found in FRANCIS PARKMAN, *A Half-Century of Conflict*, 2 vol. (1892, reprinted 1965); HOWARD H. PECKHAM, *The Colonial Wars, 1689–1762* (1964); and ALAN ROGERS, *Empire and Liberty: American Resistance to British Authority, 1755–1763* (1974). (R.R.B.)

The American Revolution: (*Prelude to revolution*): BERNARD BAILYN, *The Ideological Origins of the American Revolution* (1967), examines the transmission of English republican ideology and its American reception, while his *The Origins of American Politics* (1968) emphasizes political processes. EDWARD COUNTRYMAN, *The American Revolution* (1985), considers American social history in the explanation of how American resistance developed. P.G.D. THOMAS, *British Politics and the Stamp Act Crisis* (1975), is the most scholarly account of British objectives and methods, and his *The Townshend Duties Crisis* (1987) is the most comprehensive account of this episode. (*War of Independence*): JOHN R. ALDEN, *The American Revolution, 1775–1783* (1954, reprinted 1962), is distinguished for its political and military analyses. JACK P. GREENE (ed.), *The American Revolution* (1987), contains a valuable collection of essays. JERRILYN GREENE MARSTON, *King and Congress* (1987), studies how Congress acquired formal "legitimacy" in the course of rebellion. ROBERT MIDDLEKAUFF, *The Glorious Cause* (1982), examines the Revolution from a somewhat older point of view than is now fashionable. PIERS MACKESY, *The War for America, 1775–1783* (1964), explains the British side of the war. J.G.A. POCOCK (ed.), *Three British Revolutions: 1641, 1688, 1776* (1980), sets the American Revolution in the historical context of British experience. MORTON WHITE, *The Philosophy of the American Revolution* (1978), analyzes the concepts that took shape in the Declaration of Independence. JACK N. RAKOVE, *The Beginnings of National Politics* (1979), interprets the complex politics of the Continental Congress. Military histories include JOHN SHY, *Toward Lexington* (1965), on the British army in America; DON HIGGINBOTHAM, *The War of American Independence* (1971, reprinted 1985), the best of the military studies, showing the interrelationship of military and political developments; CHARLES ROYSTER, *A Revolutionary People at War* (1979, reissued 1981); and WILLIAM M. FOWLER, JR., *Rebels Under Sail* (1976), on the American navy. (W.M.Wa./J.R.Po.)

The early federal republic: PETER S. ONUF, *The Origins of the Federal Republic* (1983), stresses the jurisdictional problems of relations among states and between states and the Confederation. GORDON S. WOOD, *The Creation of the American Republic, 1776–1787* (1969), provides a comprehensive "ideological" interpretation emphasizing the transformation of political thought into action. DAVID F. EPSTEIN, *The Political Theory of The Federalist* (1984); and the lengthy introduction to CECELIA M. KENYON, *The Antifederalists* (1966, reprinted 1985), are excellent studies. JACKSON TURNER MAIN, *The Antifederalists: Critics of the Constitution, 1781–1788* (1961, reprinted 1974), analyzes the social origins and aspirations of the Anti-Federalists. JOYCE APPLEBY, *Capitalism and a New Social Order* (1984), persuasively argues that capitalism was seen as a liberating force by Jeffersonians as well as by Hamiltonians. GERALD STOURZH, *Alexander Hamilton and the Idea of Republican Government* (1970), explains Hamilton's vision of American opportunity. JAMES M. BANNER, JR., *To the Hartford Convention: The Federalists and the Origins of Party Politics in Massachusetts, 1789–1815* (1970), examines Federalist rumblings of secession. JOHN ZVESPER, *Political Philosophy and Rhetoric* (1977), analyzes the underlying differences of principle and political thought in the 1790s. RICHARD HOFSTADTER, *The Idea of a Party System* (1969), explains how the idea of opposition was transferred into accepted political institutions between 1780 and 1840. NOBLE E. CUNNINGHAM, *The Jeffersonian Republicans* (1957), is the most scholarly account of the formation of an organized opposition leading to Jefferson's victory in 1800. Cunningham's *The Process of Government Under Jefferson* (1978) and his *The Jeffersonian Republicans in Power* (1963) study the translation of Jeffersonian politics into action. (J.R.Po.)

From 1816 to 1850: (*The Era of Mixed Feelings*): A comprehensive and highly readable overview of the politics of this period is GEORGE DANGERFIELD, *The Era of Good Feelings* (1952, reprinted 1973). SHAW LIVERMORE, JR., *The Twilight of Federalism: The Disintegration of the Federalist Party, 1815–1830* (1962, reissued 1972), is an excellent analysis. GLOVER MOORE, *The Missouri Controversy, 1819–1821* (1953, reissued 1967), skillfully untangles that complex problem. (*Economic development*): PETER TEMIN, *The Jacksonian Economy* (1969), is provocative but not always easy to follow. Still valuable and informative are BRAY HAMMOND, *Banks and Politics in America, from the Revolution to the Civil War* (1957, reissued 1967); EDWARD PESSEN, *Most Uncommon Jacksonians: The Radical Leaders of the Early Labor Movement* (1967, reprinted 1970); and WALTER BUCKINGHAM SMITH, *Economic Aspects of the Second Bank of the United States* (1953, reissued 1969). (*Blacks, slave and free*): Particularly noteworthy studies are EUGENE D. GENOVESE, *Roll, Jordan, Roll: The World the Slaves Made* (1974); HERBERT G. GUTMAN, *The Black Family in Slavery and Freedom, 1750–1925* (1976); LEON F. LITWACK, *North of Slavery: The Negro in the Free States, 1790–1860* (1961, reprinted 1970); and IRA BERLIN, *Slaves Without Masters: The Free Negro in the Antebellum South* (1974, reissued 1981). (*Social and intellectual developments*): Lightly documented but brilliantly insightful is ALEXIS DE TOCQUEVILLE, *Democracy in America*, 2 vol. (1835; originally published in French, 1835), available in many later editions. EDWARD PESSEN, *Riches, Class, and Power Before the Civil War* (1973), challenges Tocqueville's version of equality in Jacksonian America. WILLIAM H. PEASE and JANE H. PEASE, *The Web of Progress: Private Values and Public Styles in Boston and Charleston, 1828–1843* (1985); and BARBARA WELTER, *Dimity Convictions: The American Woman in the Nineteenth Century* (1976), are useful studies. RUSH WELTER, *The Mind of America, 1820–1860* (1975), is valuable if not always persuasive. Surveys of the era's reform movements include MARTIN DUBERMAN (ed.), *The Antislavery Vanguard* (1965); and DAVID BRION DAVIS (comp.), *Ante-Bellum Reform* (1967). (*Jacksonian politics*): ARTHUR M. SCHLESINGER, JR., *The Age of Jackson* (1945, reissued 1953), is an influential study that stimulated a great array of refutations of its pro-Jackson interpretation, including EDWARD PESSEN, *Jacksonian America*, new ed. (1978, reprinted 1985). A stimulating if not always convincing comparison of Jacksonian and earlier America is ROBERT H. WIEBE, *The Opening of American Society: From the Adoption of the Constitution to the Eve of Disunion* (1984). RICHARD P. MCCORMICK, *The Second American Party System* (1966, reissued 1973), is an influential study. MICHAEL PAUL ROGIN, *Fathers and Children: Andrew Jackson and the Subjugation of the American Indian* (1975), is brilliant, original, and controversial. JOHN M. BELOHLAVEK, *Let the Eagle Soar!: The Foreign Policy of Andrew Jackson* (1985), fills a void in the Jacksonian literature. (*Expansionism*): BERNARD DE VOTO, *The Year of Decision, 1846* (1942, reissued 1989); and K. JACK BAUER, *The Mexican War, 1846–1848* (1974), are scholarly treatments. (E.Pe.)

The Civil War: The best general syntheses of modern scholarship are JAMES M. MCPHERSON, *Ordeal by Fire* (1982); and J.G. RANDALL and DAVID DONALD, *The Civil War and Reconstruction*, 2nd ed. rev. (1969). ALLAN NEVINS, *Ordeal of the Union*, 8 vol. (1947–71), provides a comprehensive history. CLEMENT EATON, *A History of the Old South*, 3rd ed. (1975, reissued 1988), is the best general history of the region. Full, critical assessments of slavery are provided by KENNETH M. STAMPP, *The Peculiar Institution* (1956, reprinted 1978); and EUGENE D. GENOVESE (*op. cit.*). A most perceptive account of the political conflicts of the late 1850s is ROY F. NICHOLS, *The Disruption of American Democracy* (1948, reissued 1967); while DON E. FEHRENBACHER, *The Dred Scott Case* (1978), offers a masterful analysis of the constitutional issues. JEAN H. BAKER, *Affairs of Party* (1983), discusses the strong partisan attachments of ordinary citizens. JAMES M. MCPHERSON, *Battle Cry of Freedom* (1988), is an engrossing narrative history of the Civil War. The most comprehensive coverage of the Confederate military effort in the East is DOUGLAS SOUTHALL FREEMAN, *Lee's Lieutenants, a Study in Command*, 3 vol. (1942–44, reissued 1970–72); while WARREN W. HASSLER, JR., *Commanders of the Army of the Potomac* (1962, reprinted 1979), does the same for the Federals. The ablest studies of the war in the Mississippi valley are the works by THOMAS L. CONNELLY, *Army of the*

Heartland: The Army of Tennessee, 1861–1862 (1967); and *Autumn of Glory: The Army of Tennessee, 1862–1865* (1971). The best book on the war's greatest battle is EDWIN B. CODDINGTON, *The Gettysburg Campaign: A Study in Command* (1968, reissued 1984). VIRGIL CARRINGTON JONES, *The Civil War at Sea,* 3 vol. (1960–62), comprehensively describes the naval war. (D.H.D./W.W.H.)

Reconstruction: Excellent syntheses of recent scholarship on the Reconstruction period are REMBERT W. PATRICK, *The Reconstruction of the Nation* (1967); JOHN HOPE FRANKLIN, *Reconstruction* (1961); and KENNETH M. STAMPP, *The Era of Reconstruction, 1865–1877* (1965, reprinted 1975). The fullest account of blacks' experience in the postwar years are LEON F. LITWACK, *Been in the Storm So Long: The Aftermath of Slavery* (1979); and ERIC FONER, *Reconstruction: America's Unfinished Revolution, 1863–1877* (1988). C. VANN WOODWARD, *Reunion and Reaction* (1951, reissued 1966), covers behind-the-scenes political and economic negotiations in the disputed 1876–77 election. The definitive account of the South in the post-Reconstruction era is C. VANN WOODWARD, *Origins of the New South, 1877–1913* (1951, reissued 1971). Important studies of postwar race relations include C. VANN WOODWARD, *The Strange Career of Jim Crow,* 3rd rev. ed. (1974, reissued 1982); and JOEL WILLIAMSON, *The Crucible of Race* (1984). (D.H.D.)

The transformation of American society, 1865–1900: (*National expansion*): A comprehensive study of the American "frontiers" of the period is HAROLD E. BRIGGS, *Frontiers of the Northwest: A History of the Upper Missouri Valley* (1940, reissued 1950). WALTER PRESCOTT WEBB, *The Great Plains* (1931, reprinted 1981), is a scholarly classic; see also RAY ALLEN BILLINGTON and MARTIN RIDGE, *Westward Expansion,* 5th ed. (1982); and RODMAN W. PAUL, *The Far West and the Great Plains in Transition, 1859–1900* (1988). HENRY E. FRITZ, *The Movement for Indian Assimilation, 1860–1890* (1963, reprinted 1981), traces the development of this policy after the Civil War. Two excellent studies of the occupation of the Plains by the farmers are FRED A. SHANNON, *The Farmer's Last Frontier: Agriculture, 1860–1897* (1945, reprinted 1977); and GILBERT C. FITE, *The Farmers' Frontier, 1865–1900* (1966, reissued 1987). (*Industrial development*): EDWARD C. KIRKLAND, *Industry Comes of Age* (1961), recounts development from the Civil War to 1897. SAMUEL P. HAYS, *The Response to Industrialism, 1885–1914* (1957), offers a perceptive appraisal of the impact of industry on American life. The best brief summary of the role of the trade unions during the second half of the 19th century is NORMAN J. WARE, *The Labor Movement in the United States, 1860–1895* (1929, reprinted 1964). (*Politics*): SEAN DENNIS CASHMAN, *America in the Gilded Age: From the Death of Lincoln to the Rise of Theodore Roosevelt,* 2nd ed. (1988), provides an overview of the era. LEONARD D. WHITE, *The Republican Era, 1869–1901* (1958, reissued 1965), presents a careful and useful analysis. H. WAYNE MORGAN, *From Hayes to McKinley: National Party Politics, 1877–1896* (1969); and HAROLD U. FAULKNER, *Politics, Reform, and Expansion, 1890–1900* (1959, reissued 1963), are also valuable. Studies of populism include JOHN D. HICKS, *The Populist Revolt* (1931, reprinted 1981); and LAWRENCE GOODWYN, *Democratic Promise: The Populist Moment in America* (1976). (H.W.Br./Ed.)

Imperialism, progressivism, and America's rise to power in the world, 1896–1920: (*American imperialism*): Varying interpretations of imperialism are presented by ERNEST R. MAY, *Imperial Democracy* (1961, reissued 1973); WALTER LAFEBER, *The New Empire: An Interpretation of American Expansion, 1860–1898* (1963); and RICHARD E. WELCH, JR., *Response to Imperialism: The United States and the Philippine-American War, 1899–1902* (1979). DAVID F. TRASK, *The War with Spain* (1981), is an excellent account of the Spanish-American War. JULIUS W. PRATT, *America's Colonial Experiment* (1950, reissued 1964), discusses the administration of the American overseas empire. A. WHITNEY GRISWOLD, *The Far Eastern Policy of the United States* (1938, reissued 1966), remains the standard work; but, for the Open Door policy and relations with China, see also TYLER DENNETT, *John Hay: From Poetry to Politics* (1933, reissued 1963). The U.S. penetration and domination of the Caribbean is most authoritatively recounted in DANA G. MUNRO, *Intervention and Dollar Diplomacy in the Caribbean, 1900–1921* (1964, reprinted 1980). (*The Progressive era*): The best introduction to the United States during the Progressive era is JOHN WHITECLAY CHAMBERS II, *The Tyranny of Change* (1980); but ARTHUR S. LINK and RICHARD L. MCCORMICK, *Progressivism* (1983), is also valuable. Still the standard treatment of populism is the work by JOHN D. HICKS cited in the previous paragraph. (*The rise to world power*): An overview of the period is JOHN M. DOBSON, *America's Ascent: The United States Becomes a Great Power, 1880–1914* (1978). The best surveys of American national politics from Roosevelt through Wilson are GEORGE E. MOWRY, *The Era of Theodore Roosevelt, 1900–1912* (1958, reprinted 1962); ARTHUR S. LINK, *Woodrow Wilson and the Progressive Era, 1910–1917* (1954, reprinted 1963); and ROBERT H. FERRELL, *Woodrow Wilson and World War I, 1917–1921* (1985). On the neutrality issue, see ERNEST R. MAY, *The World War and American Isolation, 1914–1917* (1959); and ARTHUR S. LINK, *Wilson,* 5 vol. (1947–65), especially the last three volumes. American mobilization is well covered by DANIEL R. BEAVER, *Newton D. Baker and the American War Effort, 1917–1919* (1966); and NEIL A. WYNN, *From Progressivism to Prosperity: World War I and American Society* (1986). ARNO J. MAYER, *Political Origins of the New Diplomacy, 1917–1918* (1959, reissued 1970), and a sequel, *Politics and Diplomacy of Peacemaking: Containment and Counterrevolution at Versailles, 1918–1919* (1967), include a brilliant account of the development of Wilson's peace program in its worldwide context. The best study on Wilson and American diplomacy at the Paris peace conference is ARTHUR WALWORTH, *Wilson and His Peacemakers* (1986). For the best account of the fight over the treaty in the United States, see WILLIAM C. WIDENOR, *Henry Cabot Lodge and the Search for an American Foreign Policy* (1980). WESLEY M. BAGBY, *The Road to Normalcy: The Presidential Campaign and Election of 1920* (1962), is an excellent study. (A.S.L.)

From 1920 to 1945: GEOFFREY PERRETT, *America in the Twenties* (1982), gives extensive overviews of political, social, and cultural aspects of this period. The best scholarly history is WILLIAM E. LEUCHTENBURG, *The Perils of Prosperity, 1914–32* (1958). NORMAN H. CLARK, *Deliver Us from Evil* (1976), provides a challenging revisionist history of Prohibition. FREDERICK LEWIS ALLEN, *Only Yesterday* (1931, reprinted 1986), is a contemporaneous account, covering all aspects of the years 1919–31; its companion volume is *Since Yesterday* (1940, reprinted 1986), on the 1930s. The standard account of politics in the 1930s is WILLIAM E. LEUCHTENBURG, *Franklin D. Roosevelt and the New Deal, 1932–1940* (1963). J.C. FURNAS, *Stormy Weather: Crosslights on the Nineteen Thirties* (1977), is a complete survey. IRVING BERNSTEIN, *Turbulent Years: A History of the American Worker, 1933–1941* (1969), is authoritative. GEOFFREY PERRETT, *Days of Sadness, Years of Triumph* (1973, reprinted 1985), comprehensively covers the war years 1939–45. JOHN MORTON BLUM, *V Was for Victory: Politics and American Culture During World War II* (1976), offers a critique of the war period. Military history is provided by KENNETH S. DAVIS, *Experience of War: The United States in World War II* (1965; also published as *The American Experience of War, 1939–1945,* 1967).

From 1945 to the present: JAMES GILBERT, *Another Chance: Postwar America, 1945–1985,* 2nd ed. edited by R. JACKSON WILSON (1986), is a useful survey. Coverage of the Cold War is provided by RALPH B. LEVERING, *The Cold War, 1945–1987,* 2nd ed. (1988); and JOHN LEWIS GADDIS, *Strategies of Containment* (1982), a brilliant analysis of U.S. Cold War policies. BURTON I. KAUFMAN, *The Korean War* (1986), is a reliable overview. One of the most useful histories of the Civil Rights movement is TAYLOR BRANCH, *Parting the Waters: America in the King Years, 1954–1963* (1988). GEORGE C. HERRING, *America's Longest War: The United States and Vietnam, 1950–1975,* 2nd ed. (1986), is solid. WILLIAM L. O'NEILL, *Coming Apart: An Informal History of America in the 1960's* (1971), is a study of the quality of American life under the impact of changing social values. FREDERICK F. SIEGEL, *Troubled Journey: From Pearl Harbor to Ronald Reagan* (1984), analyzes the relationship between American social and cultural life and government policy. (W.L.O'N.)

General state information: The series of guides written under the Works Progress Administration is an excellent source for information about many aspects of each state; many of these guides have been reprinted or revised. State atlases and historical atlases present graphically the settlement, development, and current condition of each state. Statistical abstracts and state yearbooks or directories ("blue books") are the best sources for data and for information on the political organization of each state. The triennial fact books for each state, published by Clements Research, give comprehensive county-by-county coverage—*e.g., Illinois Facts.* Books about place-names in each state provide information on local history. Two historical series are good introductions: *The States and the Nation Series* and *A History of the American Colonies.* State historical journals publish current historical research.

New England. *Connecticut.* THOMAS R. LEWIS and JOHN E. HARMON, *Connecticut: A Geography* (1986), is a concise contemporary treatment. DAVID M. ROTH (ed.), *Series in Connecticut History,* includes five separately titled volumes, all published in 1975, covering the early years to the present. Bibliographies of historical sources include CHRISTOPHER COLLIER and BONNIE B. COLLIER, *The Literature of Connecticut History* (1983), definitive and thoroughly annotated; and ROGER PARKS (ed.), *Connecticut* (1986). (I.J.S.)

Maine: STANLEY BEARCE ATTWOOD, *The Length and Breadth of Maine* (1946, reprinted 1974), describes the state's geography and natural resources. An authoritative general history is LOUIS CLINTON HATCH (ed.), *Maine: A History,* 5 vol. (1919; vol. 1–3 reprinted 1974). RONALD F. BANKS, *Maine Becomes a State: The Movement to Separate Maine from Massachusetts, 1785–1820* (1970), is a scholarly treatment of this important period. JOHN D. HASKELL, JR. (ed.), *Maine* (1977, reprinted 1983), is a useful bibliographic research guide. (J.N.C.)

Massachusetts: GEORGIA ORCUTT, *Massachusetts,* 2 vol. (1988), is a historical geography. Historical works include ALBERT BUSHNELL HART (ed.), *Commonwealth History of Massachusetts,* 5 vol. (1927–30, reissued 1966), on the years 1605–1930; JACK TAGER and JOHN W. IFKOVIC (eds.), *Massachusetts in the Gilded Age* (1985); and MARTIN KAUFMAN, JOHN W. IFKOVIC, and JOSEPH CARVALHO III (eds.), *A Guide to the History of Massachusetts* (1988). JOHN D. HASKELL, JR. (ed.), *Massachusetts* (1976, reprinted 1983), is an extensive bibliographic work. (M.L.Cl.)

New Hampshire: J. DUANE SQUIRES, *The Granite State of the United States,* 4 vol. (1956), covering the period from 1623 to the present; and RONALD JAGER and GRACE JAGER, *New Hampshire* (1983), are general histories. Particular periods are examined by JERE DANIELL, *Experiment in Republicanism: New Hampshire Politics and the American Revolution, 1741–1794* (1970); LYNN WARREN TURNER, *The Ninth State: New Hampshire's Formative Years* (1983); and JAMES WRIGHT, *The Progressive Yankees: Republican Reformers in New Hampshire, 1906–1916* (1987). JOHN D. HASKELL, JR., and T.D. SEYMOUR BASSETT (eds.), *New Hampshire* (1979), lists more than 6,000 items on history and culture. (J.R.D.)

Rhode Island: GEORGE H. KELLNER and J. STANLEY LEMONS, *Rhode Island* (1982), provides a lively treatment of various historical topics. PATRICK T. CONLEY, *An Album of Rhode Island History, 1636–1986* (1986), is also useful. PETER J. COLEMAN, *The Transformation of Rhode Island, 1790–1860* (1963, reprinted 1985), presents the shift from commerce to manufacturing across the state. ROGER PARKS (ed.), *Rhode Island* (1983), is an invaluable bibliographic reference. (M.I.W.)

Vermont: HAROLD A. MEEKS, *Time and Change in Vermont: A Human Geography* (1986), discusses both continuity and change. CHARLES W. JOHNSON, *The Nature of Vermont* (1980), combines history and natural history. RANDOLPH A. ROTH, *The Democratic Dilemma* (1987), is an in-depth look at social change in Vermont from 1791 to 1850. FRANK M. BRYAN, *Yankee Politics in Rural Vermont* (1974), explores the changing nature of 20th-century Vermont politics. T.D. SEYMOUR BASSETT (ed.), *Vermont* (1981), is an excellent bibliographic guide to history sources. (C.T.Mo./D.G.Sa.)

The Middle Atlantic region. *Delaware:* Overviews of state history are found in CAROL E. HOFFECKER (ed.), *Readings in Delaware History* (1973); and JOHN A. MUNROE, *History of Delaware,* 2nd ed. (1984). For specific periods, see C.A. WESLAGER, *The English on the Delaware: 1610–1682* (1967); HAROLD B. HANCOCK, *The Loyalists of Revolutionary Delaware* (1977), and *Delaware Two Hundred Years Ago: 1780–1800* (1987); and JOHN A. MUNROE, *Federalist Delaware, 1775–1815* (1954). H. CLAY REED and MARION BJÖRNSON REED (comps.), *A Bibliography of Delaware Through 1960* (1966), is supplemented by UNIVERSITY OF DELAWARE LIBRARY REFERENCE DEPARTMENT, *Bibliography of Delaware, 1960–1974* (1976). (J.A.Mu.)

Maryland: Maryland's physical and human geography is discussed in EUGENE L. MEYER, *Maryland Lost and Found* (1986); and JAMES E. DILISIO, *Maryland* (1983). JOSEPH L. ARNOLD, *Maryland* (1985), focuses on the state's economic conditions and industries. Politics and government are addressed by GEORGE H. CALLCOTT, *Maryland & America, 1940 to 1980* (1985). Historical works include ROBERT J. BRUGGER, *Maryland: A Middle Temperament, 1634–1980* (1988); and HAROLD R. MANAKEE, *Maryland in the Civil War* (1961). (Ja.H.B.)

New Jersey: Geography and geology are covered in detail by CHARLES H. STANSFIELD, JR., *New Jersey: A Geography* (1983); PETER O. WACKER, *Land and People* (1975), on the preindustrial state; and PETER E. WOLFE, *The Geology and Landscapes of New Jersey* (1977). Historical treatments include RICHARD P. MCCORMICK, *New Jersey from Colony to State, 1609–1789,* rev. ed. (1981); and LARRY R. GERLACH, *Prologue to Independence* (1976), on the pre-Revolutionary years. (P.O.W.)

New York: Physical features are described in JOHN H. THOMPSON (ed.), *Geography of New York State* (1966, reprinted 1977). Historical works include DAVID M. ELLIS *et al., A History of New York State,* rev. ed. (1967); DAVID M. ELLIS, *New York: State & City* (1979); and DAVID M. ELLIS, JAMES A. FROST, and WILLIAM B. FINK, *New York,* 5th ed. (1980). (P.J.Sc.)

Pennsylvania: A detailed work on Pennsylvania's geography is RAYMOND E. MURPHY and MARION MURPHY, *Pennsylvania* (1937). On the people of Pennsylvania, see JOHN A. HOSTETLER, *Amish Society,* 3rd ed. (1980); and WILLIAM T. PARSONS, *The Pennsylvania Dutch: A Persistent Minority* (1976). Historical works include PHILIP S. KLEIN and ARI HOOGENBOOM, *A History of Pennsylvania,* 2nd enlarged ed. (1980); JOSEPH J. KELLEY, *Pennsylvania: The Colonial Years, 1681–1776* (1980); and SYLVESTER K. STEVENS, *Pennsylvania: Birthplace of a Nation* (1964), and *Pennsylvania: The Keystone State,* 2 vol. (1956). (C.L.T.)

The South. *Alabama:* JESSE M. RICHARDSON (ed.), *Alabama Encyclopedia,* vol. 1, *Book of Facts* (1965), provides comprehensive data on a variety of topics and includes an excellent chronology. Studies of the state's history include THOMAS PERKINS ABERNATHY, *The Formative Period in Alabama, 1815–1828,* 2nd ed. (1965); J. MILLS THORNTON III, *Politics and Power in a Slave Society: Alabama, 1800–1860* (1978, reprinted 1981); WALTER L. FLEMING, *Civil War and Reconstruction in Alabama* (1905, reprinted 1978); and ALLEN J. GOING, *Bourbon Democracy in Alabama, 1874–1890* (1951, reprinted 1972). Historical discussions of race relations are provided in PETER KOLCHIN, *First Freedom: The Response of Alabama's Blacks to Emancipation and Reconstruction* (1972); ROBERT J. NORRELL, *Reaping the Whirlwind: The Civil Rights Movement in Tuskegee* (1985); and DAN T. CARTER, *Scottsboro: A Tragedy of the American South,* rev. ed. (1979). (R.J.No.)

Arkansas: JOHN GOULD FLETCHER, *Arkansas* (1947, reprinted 1989), is one of the best single volumes about the state. DIANN SUTHERLIN SMITH, *The Arkansas Handbook* (1984), provides general information. Arkansas's population is analyzed in DAVID M. TUCKER, *Arkansas: A People and Their Reputation* (1985). Historical topics are examined in JAMES M. WOODS, *Rebellion and Realignment: Arkansas's Road to Secession* (1987); GEORGE H. THOMPSON, *Arkansas and Reconstruction: The Influence of Geography, Economics, and Personality* (1976); and C. CALVIN SMITH, *War and Wartime Changes: The Transformation of Arkansas, 1940–1945* (1986). (B.A.D.)

Florida: General information sources include *The Florida Handbook* (biennial); and DEL MARTH and MARTHA J. MARTH (eds.), *The Florida Almanac, 1988–89,* 7th ed. (1988). ROBERT B. MARCUS and EDWARD A. FERNALD, *Florida: A Geographical Approach* (1975), is comprehensive. Studies of the state's environment include MARJORY STONEMAN DOUGLAS, *The Everglades: River of Grass,* rev. ed. (1988); and NELSON MANFRED BLAKE, *Land Into Water—Water Into Land: A History of Water Management in Florida* (1980). The best general history is CHARLTON W. TEBEAU, *A History of Florida,* rev. ed. (1980). (R.H.Fu.)

Georgia: LAWRENCE R. HEPBURN (ed.), *Contemporary Georgia* (1987), is a good overview. For Georgia's history, see KENNETH COLEMAN (ed.), *A History of Georgia* (1977); E. MERTON COULTER, *Georgia,* rev. and enlarged ed. (1960), dated but still highly readable, with the flavour of the pre-integration South; HARVEY H. JACKSON and PHINIZY SPALDING (eds.), *Forty Years of Diversity: Essays on Colonial Georgia* (1984); MICHAEL P. JOHNSON, *Toward a Patriarchal Republic: The Secession of Georgia* (1977); and NUMAN V. BARTLEY, *From Thurmond to Wallace: Political Tendencies in Georgia, 1948–1968* (1970), a review of post-World War II politics in the state. (L.DeV.)

Kentucky: General geographic information may be found in P.P. KARAN (ed.), *Kentucky* (1973); KARL B. RAITZ, *The Kentucky Bluegrass* (1980); and WILFORD A. BLADEN, *A Geography of Kentucky* (1984). HARRY M. CAUDILL, *A Darkness at Dawn: Appalachian Kentucky and the Future* (1976), focuses on political and social conditions in eastern Kentucky. The state's history is examined in LEWIS COLLINS, *Collins' Historical Sketches of Kentucky,* rev. and enlarged by RICHARD H. COLLINS, 2 vol. (1874, reissued 1979), still valuable; THOMAS D. CLARK, *A History of Kentucky,* rev. ed. (1960), and *Kentucky* (1968); GEORGE MORGAN CHINN, *Kentucky Settlement and Statehood, 1750–1800* (1975); and HAMBLETON TAPP and JAMES C. KLOTTER, *Kentucky: Decades of Discord, 1865–1900* (1977). (W.A.Bl.)

Louisiana: FRED B. KNIFFEN and SAM BOWERS HILLIARD, *Louisiana,* rev. ed. (1988), provides a concise cultural geography. JAMES BOLNER (ed.), *Louisiana Politics* (1982), highlights the structure and practice of Louisiana's government. MARK T. CARLETON, PERRY H. HOWARD, and JOSEPH B. PARKER (comps.), *Readings in Louisiana Politics,* 2nd ed. (1988), provides classic statements of social, economic, and political history. NOEL GRAY, *A Short History of Louisiana* (1965), gives attention to the black community's role. GLENN R. CONRAD (ed.), *Readings in Louisiana History* (1978), is comprehensive and includes maps and bibliographies. A helpful bibliography is LIGHT TOWNSEND CUMMINS and GLEN JEANSONNE (eds.), *A Guide to the History of Louisiana* (1982). (P.H.H.)

Mississippi: The most complete history is RICHARD AUBREY MCLEMORE (ed.), *A History of Mississippi,* 2 vol. (1973). On

slavery, see CHARLES SACKETT SYDNOR, *Slavery in Mississippi* (1933, reissued 1965). The pre- and post-Civil War periods are covered in EDWIN A. MILES, *Jacksonian Democracy in Mississippi* (1961, reprinted 1985); PERCY L. RAINWATER, *Mississippi: Storm Center of Secession, 1856–1861* (1938, reissued 1969); JOHN K. BETTERSWORTH, *Confederate Mississippi* (1943, reprinted 1978); and WILLIAM C. HARRIS, *Presidential Reconstruction in Mississippi* (1967), and *The Day of the Carpetbagger* (1979). JAMES W. SILVER, *Mississippi: The Closed Society*, new enlarged ed. (1966), explains with remarkable clarity both the causes and consequences of Mississippi's official policy of racial segregation. (D.G.S.)

North Carolina: JAMES A. CRUTCHFIELD (ed.), *The North Carolina Almanac and Book of Facts* (1986), provides an overview. Introductions to North Carolina's history are found in LINDLEY S. BUTLER and ALAN D. WATSON, (eds.), *The North Carolina Experience* (1984); and H.G. JONES, *North Carolina Illustrated, 1524–1984* (1983). Plantations and slavery are examined in JEFFREY J. CROW, *The Black Experience in Revolutionary North Carolina* (1977). (D.G.M.)

South Carolina: CHARLES F. KOVACIK and JOHN J. WINBERRY, *South Carolina* (1987), discusses historical and economic geography. JACK BASS, *Porgy Comes Home: South Carolina . . . After 300 Years* (1972), looks at the social and economic revolutions of the 1960s. DAVID DUNCAN WALLACE, *The History of South Carolina*, 4 vol. (1934), remains the classic history of the state. For specific eras, see ALFRED GLAZE SMITH, *Economic Readjustment of an Old Cotton State: South Carolina, 1820–1860* (1958); and ERNEST MCPHERSON LANDER, *A History of South Carolina, 1865–1960*, 2nd ed. (1970). (J.J.W.)

Tennessee: For basic physical geography, see RALPH O. FULLERTON and JOHN B. RAY (eds.), *Tennessee* (1977); and EDWARD T. LUTHER, *Our Restless Earth: The Geologic Regions of Tennessee* (1977). MARGUERITE OWEN, *The Tennessee Valley Authority* (1973), gives the history and details of the programs of this important agency. A good history of the state is ROBERT E. CORLEW, *Tennessee*, 2nd ed. (1981), which contains excellent chapter bibliographies. See also SAM B. SMITH (ed. and comp.), *Tennessee History: A Bibliography* (1974). (S.McC.H.)

Virginia: An introduction to the state is provided by JEAN GOTTMAN, *Virginia in Our Century* (1969), a comprehensive analysis through 1968. EDMUND S. MORGAN, *American Slavery, American Freedom: The Ordeal of Colonial Virginia* (1975), analyzes the social basis of early Virginian political thought. Various periods in the state's past are described in RICHARD L. MORTON, *Colonial Virginia*, 2 vol. (1960); VIRGINIUS DABNEY, *Virginia, the New Dominion* (1971, reprinted 1983); PARKE ROUSE, *Virginia* (1966); and A.W. MOGER, *Virginia: Bourbonism to Byrd, 1870–1925* (1968). (C.L.Q.)

West Virginia: For the state's physical features, see RAYMOND E. JANSSEN, *Earth Science: A Handbook on the Geology of West Virginia* (1964). PHIL CONLEY, *History of the West Virginia Coal Industry* (1960), examines this important industry. OTIS K. RICE, *West Virginia: A History* (1985), provides an overview. PHIL CONLEY (ed.), *The West Virginia Encyclopedia* (1929), is a classic, but dated, general reference, still good for history; see also SAM CLAGG, *West Virginia Historical Almanac* (1975). (S.E.C.)

The Midwest. *Illinois:* RONALD E. NELSON (ed.), *Illinois: Land and Life in the Prairie State* (1978); and A. DOYNE HORSLEY, *Illinois: A Geography* (1986), are good general works. R. DAN NEELY and CARLA G. HEISTER (comps.), *The Natural Resources of Illinois* (1987), contains information on climate and biological and geologic features. JOHN CLAYTON, *The Illinois Fact Book and Historical Almanac, 1673–1968* (1970), is comprehensive. Illinois literary figures are profiled in ILLINOIS STATE LIBRARY, *Illinois Authors* (1983). Illinois history is detailed in ROBERT P. HOWARD, *Illinois* (1972). (R.T.L./V.D.S.)

Indiana: Overviews of Indiana history include DWIGHT W. HOOVER and JANE RODMAN, *A Pictorial History of Indiana* (1980); and JAMES H. MADISON, *The Indiana Way: A State History* (1986). The series *The History of Indiana* offers detailed coverage of various periods. (Ed.)

Iowa: H.L. NELSON, *A Geography of Iowa* (1967), surveys many aspects; while WAYNE I. ANDERSON, *Geology of Iowa* (1983), analyzes the important natural resources, as well as the geologic processes that have formed the state. TOM C. COOPER and NYLA SHERBURNE HUNT (eds.), *Iowa's Natural Heritage* (1982), examines the state's natural history. STEPHEN WILBERS, *The Iowa Writers' Workshop: Origins, Emergence, & Growth* (1980), traces the development of this creative writing program. Overviews of state history include CYRENUS COLE, *Iowa Through the Years* (1940), a political treatment; and LELAND L. SAGE, *A History of Iowa* (1974, reissued 1987). (N.E.S.)

Kansas: A guide to Kansas geology is REX C. BUCHANAN (ed.), *Kansas Geology* (1984). JOSEPH T. COLLINS (ed.), *Natural Kansas* (1985), recounts the state's natural history. Introductions to Kansas history include BLISS ISELY and W. MARVIN RICHARDS, *The Kansas Story* (1961); ROBERT W. RICHMOND, *Kansas*, 3rd ed. (1989); LOUISE BARRY, *The Beginning of the West: Annals of the Kansas Gateway to the American West, 1540–1854* (1972); CRAIG MINER, *West of Wichita: Settling the High Plains of Kansas, 1865–1890* (1986); and FRANCIS W. SCHRUBEN, *Kansas in Turmoil, 1930–1936* (1969). (C.G.P.)

Michigan: The state's physical features are described in LAWRENCE M. SOMMERS, *Michigan* (1984); and RICHARD A. SANTER, *Michigan, Heart of the Great Lakes* (1977). GEORGE S. MAY, *A Most Unique Machine: The Michigan Origins of the American Automobile Industry* (1975); and ANGUS MURDOCH, *Boom Copper* (1943, reissued 1964), focus on Michigan's economy. Useful general histories include RICHARD J. HATHAWAY (ed.), *Michigan: Visions of Our Past* (1989); GEORGE S. MAY, *Michigan* (1987); BRUCE A. RUBENSTEIN and LAWRENCE E. ZIEWACZ, *Michigan* (1981); and WILLIS F. DUNBAR, *Michigan*, rev. ed. by GEORGE S. MAY (1980). (Ri.J.H.)

Minnesota: PATRICIA CONDON JOHNSTON, *Minnesota: Portrait of the Land and Its People* (1987), is an illustrated overview. Physical features are described by GEORGE M. SCHWARTZ and GEORGE A. THIEL, *Minnesota's Rocks and Waters: A Geological Story*, rev. ed. (1963); and RICHARD W. OJAKANGAS and CHARLES L. MATSCH, *Minnesota's Geology* (1982). JUNE DRENNING HOLMQUIST (ed.), *They Chose Minnesota: A Survey of the State's Ethnic Groups* (1981), covers the years from 1850 to 1980. MARY ANN GROSSMANN and TOM THOMSEN (eds.), *The Minnesota Almanac, 1988*, 3rd ed. (1987), compiles statistics on a variety of subjects. Studies of Minnesota's past may be found in WILLIAM WATTS FOLWELL, *A History of Minnesota*, rev. ed., 4 vol. (1956–69); and THEODORE C. BLEGEN, *Minnesota*, expanded ed. (1975). (N.C.G.)

Missouri: MILTON D. RAFFERTY, *Missouri* (1983), includes discussion of physical, economic, political, and cultural topics. RUSSEL L. GERLACH, *Settlement Patterns in Missouri* (1986), is a concise work. Historical treatments include LOUIS HOUCK, *A History of Missouri from the Earliest Explorations and Settlements Until the Admission of the State into the Union*, 3 vol. (1908; reissued 3 vol. in 1, 1971), the definitive history up to statehood; WILLIAM E. PARRISH (ed.), *A History of Missouri* (1971–), a comprehensive historical survey; and DUANE MEYER, *The Heritage of Missouri*, 3rd ed. (1982), perhaps the best one-volume history of the state. (M.D.R.)

Nebraska: BRADLEY H. BALTENSPERGER, *Nebraska* (1985), offers an excellent treatment of the state's recent development. FREDERICK C. LUEBKE, "Nebraska: Time, Place, and Culture," in JAMES H. MADISON (ed.), *Heart Land: Comparative Histories of the Midwestern States* (1988), pp. 226–247, is the best interpretation. Political processes are the concern of ROBERT W. CHERNY, *Populism, Progressivism, and the Transformation of Nebraska Politics, 1885–1915* (1981); and STANLEY B. PARSONS, *The Populist Context: Rural Versus Urban Power on a Great Plains Frontier* (1973). JAMES AUCOIN, *Water in Nebraska: Use, Politics, Policies* (1984), discusses development of water resources. JAMES C. OLSON, *History of Nebraska*, 2nd ed. (1966, reissued 1974), is the best general historical work. (H.A.D.)

North Dakota: FRANCIE M. BERG, *North Dakota* (1977), is a valuable descriptive work. MELVIN E. KAZECK, *North Dakota: A Human and Economic Geography* (1956), although dated, is also useful. JOHN P. BLUEMLE, *The Face of North Dakota: The Geologic Story* (1977), is a sound introduction. WILLIAM C. SHERMAN and PLAYFORD V. THORSON (eds.), *Plains Folk: North Dakota's Ethnic History* (1988), provides detailed information. A good collection of well-documented essays on major political personages and movements is THOMAS M. HOWARD (ed.), *The North Dakota Political Tradition* (1981). ELWYN B. ROBINSON, *History of North Dakota* (1966, reissued 1982), is the authoritative and exemplary history up to about 1960. (B.O.'K.)

Ohio: ALBERT J. KORSOK, *Ohio* (1975), is an introduction. MICHAEL B. LAFFERTY (ed.), *Ohio's Natural Heritage* (1979), describes the state's natural history. Sources of information on the people, economy, and government of Ohio include RAYMOND BORYCZKA and LORIN LEE CARY, *No Strength Without Union: An Illustrated History of Ohio Workers, 1803–1980* (1982); and ANDREW R.L. CAYTON, *The Frontier Republic: Ideology and Politics in the Ohio Country, 1780–1825* (1986). Ohio's history is detailed in EUGENE H. ROSEBOOM and FRANCIS P. WEISENBURGER, *A History of Ohio*, 2nd ed. (1967, reissued 1986); GEORGE W. KNEPPER, *An Ohio Portrait* (1976); and DAVID A. GERBER, *Black Ohio and the Color Line, 1860–1915* (1976). (G.W.Kn.)

South Dakota: Overviews of the land and people include J. LEONARD JANNEWEIN and JANE BOORMAN (eds.), *Dakota Panorama*, 3rd ed. (1973); and FRANCIE M. BERG, *South Dakota* (1982). HERBERT T. HOOVER and LARRY ZIMMERMAN (eds.), *South Dakota Leaders* (1989), is the best interpretative histor-

ical volume. Other history sources include HERBERT S. SCHELL, *History of South Dakota,* 3rd ed. rev. (1975); and HOWARD ROBERT LAMAR, *Dakota Territory, 1861–1889* (1956, reissued 1966). JACK W. MARKEN and HERBERT T. HOOVER, *Bibliography of the Sioux* (1980); and SUE LAUBERSHEIMER (ed.), *South Dakota: Changing, Changeless, 1889–1989* (1985), contain annotated lists of sources on state history. (H.T.H.)

Wisconsin: The state's features are covered by ROBERT W. FINLEY, *Geography of Wisconsin,* rev. ed. (1976), a broad picture of the state's physical and cultural geography; GWEN M. SCHULTZ, *Wisconsin's Foundations: A Review of the State's Geology and Its Influence on Geography and Human Activity* (1986), a well-illustrated work for the nonspecialist; and INGOLF VOGELER, *Wisconsin* (1986). WILLIAM FLETCHER THOMPSON (ed.), *The History of Wisconsin* (1973–), is a well-researched comprehensive work. ROBERT C. NESBIT, *Wisconsin: A History* (1973), is an introduction. (R.W.F.)

The Southwest. *Arizona:* Geographic studies include MALCOLM L. COMEAUX, *Arizona* (1981); UNIVERSITY OF ARIZONA, *Arizona: Its People and Resources,* rev. 2nd ed. (1972); and TOM MILLER (ed.), *Arizona: The Land and the People* (1986). For history, see THOMAS EDWIN FARISH, *History of Arizona,* 8 vol. (1915–18), still a good source for early history; MARSHALL TRIMBLE, *Arizona* (1977); and JAY J. WAGONER, *Early Arizona: Prehistory to Civil War* (1975), and *Arizona Territory, 1863–1912* (1970). (J.W.B.)

New Mexico: IRA G. CLARK, *Water in New Mexico* (1987), provides a voluminous study of the state's water problem. ARRELL MORGAN GIBSON, *The Santa Fe and Taos Colonies: Age of the Muses, 1900–1942* (1983), is the best study on the art colonies; while MARTA WEIGLE and KYLE FIORE, *Santa Fe and Taos: The Writer's Era, 1916–1941* (1982), gives an overview of the leading authors of the period. Historical studies of New Mexico include WARREN A. BECK, *New Mexico: A History of Four Centuries* (1962); and ROBERT W. LARSON, *New Mexico's Quest for Statehood, 1846–1912* (1968), and *New Mexico Populism* (1974). (W.A.B.)

Oklahoma: KENNY A. FRANKS, *The Oklahoma Petroleum Industry* (1980); and EDWARD EVERETT DALE, *The Range Cattle Industry: Ranching on the Great Plains from 1865 to 1925,* new ed. (1960), profile two important industries. Modern political subjects are analyzed in JAMES R. SCALES and DANNEY GOBLE, *Oklahoma Politics* (1982). The best comprehensive historical work is ARRELL MORGAN GIBSON, *Oklahoma,* 2nd ed. (1981). ANNE HODGES MORGAN and H. WAYNE MORGAN (eds.), *Oklahoma* (1982), collects essays on a variety of topics. (J.S.E.)

Texas: A good overview is TERRY G. JORDAN, JOHN L. BEAN, JR., and WILLIAM M. HOLMES, *Texas* (1984). WALTER PRESCOTT WEBB and ELDON STEPHEN BRANDA (eds.), *The Handbook of Texas,* 3 vol. (1952–76), offers encyclopaedic historical information. Excellent studies of the two largest minority groups in the state are ALWYN BARR, *Black Texans: A History of Negroes in Texas, 1528–1971* (1973, reprinted 1982); and ARNOLDO DE LEÓN, *The Tejano Community, 1836–1900* (1982). Histories include RUPERT N. RICHARDSON, ERNEST WALLACE, and ADRIAN ANDERSON, *Texas,* 5th ed. (1988), the best one-volume state history; and T.R. FEHRENBACH, *Lone Star* (1968, reissued 1985). LIGHT TOWNSEND CUMMINS and ALVIN R. BAILEY, JR. (eds.), *A Guide to the History of Texas* (1988), includes historiographic and topical subject essays. (R.A.Wo.)

The Mountain region. *Colorado:* MEL GRIFFITHS and LYNNELL RUBRIGHT, *Colorado* (1983), is a systematic look at the state's geography. CORNELIA FLEISCHER MUTEL and JOHN C. EMERICK, *From Grassland to Glacier: The Natural History of Colorado* (1984), describes various Colorado ecosystems. PHILIP L. FRADKIN, *A River No More: The Colorado River and the West* (1981), examines the role of water in the Colorado basin. GLEAVES WHITNEY, *Colorado Front Range: A Landscape Divided* (1983), traces the rapid growth along the Front Range corridor. A historical treatment may be found in CARL UBBELOHDE, *A Colorado History,* 6th ed. (1988). (J.L.Di.)

Idaho: MERLE WELLS and ARTHUR A. HART, *Idaho* (1985), provides information about the state, its industries, and its history. RANDY STAPILUS, *Paradox Politics: People and Power in Idaho* (1988), analyzes popular government. MERRILL D. BEAL and MERLE W. WELLS, *History of Idaho,* 3 vol. (1959), is the most complete history. (B.A.M.)

Montana: Two popular geographies cover the state: JOHN A. ALWIN, *Western Montana* (1983), and *Eastern Montana* (1982). A popular history is HARRY W. FRITZ, *Montana: Land of Contrast* (1984). Among the several scholarly histories of Montana are MERRILL G. BURLINGAME and K. ROSS TOOLE, *A History of Montana,* 3 vol. (1957); and the comprehensive MICHAEL P. MALONE and RICHARD B. ROEDER, *Montana* (1976). A highly insightful and interpretive history is K. ROSS TOOLE, *Montana: An Uncommon Land* (1959), with a companion volume, *Twentieth-century Montana: A State of Extremes* (1972). (J.M.C.)

Nevada: A study of the state's settlement is found in WILBUR S. SHEPPERSON, *Restless Strangers: Nevada's Immigrants and Their Interpreters* (1970). RUSSELL R. ELLIOTT, *History of Nevada,* 2nd ed. rev. (1987), is the most detailed and comprehensive reference. JAMES W. HULSE, *The Nevada Adventure: A History,* 5th ed. (1981), provides a general survey from prehistory to modern times. Older but still useful is JAMES G. SCRUGHAM (ed.), *Nevada: A Narrative of the Conquest of a Frontier Land,* 3 vol. (1935). STANLEY W. PAHER, *Nevada: An Annotated Bibliography* (1980), is comprehensive. (R.J.Z.)

Utah: Good references on the land and people include HELEN Z. PAPANIKOLAS (ed.), *The Peoples of Utah* (1976); and THOMAS K. MARTIN, TIM B. HEATON, and STEPHEN J. BAHR (eds.), *Utah in Demographic Perspective: Regional and National Contrasts* (1986). The geologic history is told in WILLIAM LEE STOKES, *Geology of Utah* (1986). Political and economic developments are described in NELS ANDERSON, *Desert Saints: The Mormon Frontier in Utah* (1942, reissued 1966); and GUSTIVE O. LARSON, *The "Americanization" of Utah for Statehood* (1971). Excellent histories include DEAN L. MAY, *Utah* (1987); WAYNE K. HINTON, *Utah: Unusual Beginning to Unique Present* (1988); and RICHARD D. POLL (ed.), *Utah's History* (1978). (L.J.A.)

Wyoming: ROBERT HAROLD BROWN, *Wyoming* (1980), describes the land and its resources. DAVID LAGESON and DARWIN SPEARING, *Roadside Geology of Wyoming* (1988); and D.L. BLACKSTONE, JR., *Traveler's Guide to the Geology of Wyoming,* 2nd ed. (1988), trace the state's geologic history. T.A. LARSON, *History of Wyoming,* 2nd ed. rev. (1978), is recommended. (G.R.We.)

The Pacific Coast. *Alaska:* Reference works include *The Alaska Almanac* (annual); and R.K. WOERNER (ed.), *The Alaska Handbook* (1986), an encyclopaedia of information and statistics. Two books from the NATIONAL GEOGRAPHIC SOCIETY (U.S.), *Alaska,* by BERN KEATING, 2nd ed. (1971), and *Alaska: High Roads to Adventure* (1976), offer illustrated essays on geographic regions, people, and industries. ERNEST GRUENING, *The State of Alaska,* expanded ed. (1968), is an authoritative text on politics and economics in Alaska in the decade before statehood. CLARENCE C. HULLEY, *Alaska: Past and Present,* 3rd ed. (1970, reprinted 1981), provides a general history from the Russian days to the 1960s. CLAUDE-M. NASKE and HERMAN E. SLOTNICK, *Alaska,* 2nd ed. (1987), includes chapters on native land claims, conservation, and the oil boom. (M.M.M.)

California: JAMES D. HART, *A Companion to California,* new ed., rev. and expanded (1987), includes entries on all aspects of the state. For geography and landscape, see DAVID W. LANTIS, RODNEY STEINER, and ARTHUR E. KARINEN, *California,* rev. 3rd ed. (1981); CRANE S. MILLER and RICHARD S. HYSLOP, *California: The Geography of Diversity* (1983); and MARY HILL, *California Landscape: Origin and Evolution* (1984). Cultural history is addressed by KEVIN STARR, *Americans and the California Dream, 1850–1915* (1973, reprinted 1986), and *Inventing the Dream: California Through the Progressive Era* (1985); CAREY MCWILLIAMS, *Southern California Country* (1946, reissued as *Southern California,* 1979); and NEIL MORGAN, *The California Syndrome* (1969). Histories include JOHN W. CAUGHEY and NORRIS HUNDLEY, JR., *California,* 4th ed. (1982); WALTON BEAN and JAMES J. RAWLS, *California,* 5th ed. (1988); and ANDREW ROLLE, *California,* 4th ed. (1987). (N.Mo.)

Hawaii: Physical characteristics are examined by JOSEPH R. MORGAN, *Hawaii* (1983); and GORDON A. MACDONALD, AGATIN T. ABBOTT, and FRANK L. PETERSON, *Volcanoes in the Sea: The Geology of Hawaii,* 2nd ed. (1983). ANDREW W. LIND, *An Island Community* (1938, reissued 1968), is an excellent study of the migrations to Hawaii and subsequent race relations; it is complemented by ELVI WHITTAKER, *The Mainland Haole: The White Experience in Hawaii* (1986). Good histories include GAVAN DAWS, *Shoal of Time: A History of the Hawaiian Islands* (1968, reissued 1974), one of the best single-volume histories of Hawaii; and RALPH S. KUYKENDALL and A. GROVE DAY, *Hawaii,* rev. ed. (1961, reissued 1976). (L.S.M.)

Oregon: SAMUEL N. DICKEN and EMILY F. DICKEN, *Two Centuries of Oregon Geography,* 2 vol. (1979–82), covers both historical and descriptive geography. EWART M. BALDWIN, *Geology of Oregon,* 3rd ed. (1981), is a handy summary reference. Historical treatment is provided by PHIL F. BROGAN, *East of the Cascades,* 4th ed. (1977), a popularization of central Oregon's history; and CHARLES H. CAREY, *General History of Oregon Through Early Statehood,* 3rd ed. (1971). (R.M.Hi.)

Washington: D.W. MEINIG, *The Great Columbia Plain: A Historical Geography, 1805–1910* (1968), covers the settlement and organization of an area that includes eastern Washington; see also JOHN A. ALWIN, *Between the Mountains: A Portrait of Eastern Washington* (1984). DANIEL M. OGDEN, JR., and HUGH A. BONE, *Washington Politics* (1960, reprinted 1981), is an informative account. An excellent general history is MARY W. AVERY, *Washington,* rev. ed. (1965). (H.J.Cr.)

Uruguay

Uruguay (officially the República Oriental del Uruguay, "Oriental" designating its position on the eastern shore of the Uruguay River) is the smallest republic in South America after Suriname. About 3,000,000 people live within the country's area of some 68,000 square miles (176,120 square kilometres), almost half of them concentrated in the metropolitan area of Montevideo, the capital. The resulting demographic disequilibrium finds the remaining half of the population distributed throughout the rest of the national territory. The populations of the largest provincial cities are just fractions of that of the capital. The wedge-shaped country is bounded by Brazil on the north and east, by the Atlantic Ocean on the southeast, and by the Río de la Plata, an estuary of the Atlantic Ocean, on the south, while the Uruguay River serves as its western boundary, separating the small republic from Argentina.

With an economy based largely on agricultural exports, chiefly livestock products, Uruguay developed throughout much of the 20th century as one of Latin America's more progressive societies. During that time the nation was recognized for its political stability and advanced social legislation. However, social unrest grew out of economic problems that had developed in the 1960s, giving rise to the Tupamaros, an urban guerrilla movement named after Tupac Amaru II, an 18th-century Inca who had rebelled against Spanish rule. The violent activities of the guerrillas so disrupted public life that the democratic governmental structure was, under military leadership, abridged for more than a decade. It was not until 1985 that the military government, under increasing social and economic pressure, reluctantly relinquished political control to a popularly elected civilian government.

This article is divided into the following sections:

Physical and human geography 458
- The land 458
  - Relief, soils, and drainage
  - Climate
  - Plant and animal life
  - Settlement patterns
- The people 459
- The economy 460
  - Resources
  - Agriculture
  - Industry
  - Finance and trade
  - Transportation
- Administration and social conditions 461
  - Government
  - Education
  - Health and welfare
- Cultural life 462
  - The arts
  - Cultural institutions
  - Recreation
  - Press and broadcasting

History 462
- Before the Europeans 462
- Colonial period 462
- The struggle for national identity (1810–80) 462
- The modernization of Uruguay (1880–1930) 463
- Economic decline and military takeover (1930–73) 463
- The military regime (1973–85) 464
- Civilian government 464

Bibliography 464

Physical and human geography

THE LAND

Relief, soils, and drainage. Most of the Uruguayan landscape is characterized by gently rolling land and low hills. In the southern and eastern half of the country the land exhibits a subsoil of weathered ancient schist (crystalline rock) and granite through which less-weathered rocks protrude. In north central and central Uruguay the ancient schist is overlaid by horizontal layers of rock dating from the Permian period (from 286,000,000 to 245,000,000 years ago) that form a plateau. The northwestern portion is occupied by a plateau composed of horizontal beds of Triassic red sandstone (208,000,000 to 245,000,000 years old), in places faulted and capped by sheets of Triassic basalt. The plains are covered with deposits of sand and clay of the Pleistocene epoch (1,600,000 to 10,000 years ago) and with alluvial beds.

In northeastern Uruguay the low, rolling hills are a southward extension of the Brazilian Highlands. Tidal lakes and sand dunes fringe the coastline. The northwestern section of the republic presents a greater variety of relief, with occasional ridges, broad valleys, and low plateaus. No hill in Uruguay rises to more than 1,700 feet (520 metres).

Drainage

Although it is a well-watered land, no large rivers flow entirely within Uruguay. The largest stream, the Negro River, is navigable only in its lower part. The Uruguay River, along the western border of the nation, is navigable for ships of 14-foot (4.3-metre) draft from its mouth to Paysandú, and for smaller vessels above that point to the falls at Salto. The Santa Lucía, Cebollatí, and Queguay Grande rivers are the other principal streams.

Climate. Uruguay enjoys a generally pleasant, temperate climate. The average temperature for the midwinter month of July varies from 54° F (12° C) at Salto in the northern interior to 50° F (10° C) at Montevideo in the south. The midsummer month of January varies from a warm average of 79° F (26° C) at Salto to 72° F (22° C) at Montevideo. Frost is almost unknown along the coast. Both summer and winter weather may vary from day to day with the passing of storm centres; a hot northerly wind on occasion may be followed by a cold wind from the Argentine pampas, the *pampero*.

Precipitation

Uruguay has neither a decidedly dry nor a rainy season. The heaviest rainfall occurs during March and April, autumn months in the Southern Hemisphere, although more frequent rains occur in winter. The mean annual precipitation is generally greater than 40 inches (1,000 millimetres), decreasing with distance from the seacoast, and is relatively evenly distributed throughout the year. Frequent thunderstorms occur in the summer.

Plant and animal life. Tall prairie grasses cover most of Uruguay's land surface. Only a small percentage of the land is forested, most of the trees growing in narrow stretches along the watercourses. The principal species are the ombu, a scrubby, treelike plant, and the alder. Others include the willow, eucalyptus, poplar, acacia, and aloe. The algaroba (carob tree) and the quebracho (whose wood and bark are utilized in tanning and dyeing) are prevalent, and indigenous palms grow in the valleys of the Sierra de San José Ignacio and along the southeastern coast. Mimosa, myrtle, rosemary, and the scarlet-flowered ceibo are common.

Wild animals native to Uruguay have largely disappeared. The puma and the jaguar, however, are still sometimes seen in remote areas. Other native mammals include the fox, deer, wildcat, and several types of rodent, including the huge capybara. A small armadillo, the mulita, is re-

lated to the extinct mylodon and megatherium, the fossils of which are found throughout the pampa. The scorpion is rare, but venomous spiders are common.

Birdlife

Among the birdlife are a few specimens of the vulture, a native crow, and partridges and quails. Parakeets are plentiful in the hills, and the lagoons swarm with waterfowl. Characteristic of the prairies is the sight of the tiny burrowing owl perched atop an eminence. Large flocks of the lapwing terutero are common, and bright-plumaged birds include the hummingbird and cardinal. The rhea is now found mostly in semidomesticated settings, while white herons, cranes, and flamingos are seen in swampy lands along the coast.

The principal reptiles include a lizard, a tortoise, a dangerous viper, and the rattlesnake. The alligator-like caiman inhabits the upper waters of the Uruguay River, and seals are found on small islands off the southeast coast, particularly on Lobos Island.

Settlement patterns. When Uruguay became independent in 1828, its national territory was used almost exclusively for the grazing of herds of cattle and sheep on the unfenced range; there were few permanent settlements outside of Montevideo and Colonia del Sacramento and the villages along the Uruguay River. The grazing lands along the eastern shore of the Uruguay River constituted a kind of no-man's-land between the Portuguese Brazilians and the Spanish Argentines.

After independence, Uruguay received a small influx of immigrants, chiefly from Italy and Spain. At that time Argentina was torn by civil strife, and settlement on the pampas was not yet attractive. The newcomers entered Uruguay through Montevideo and settled southern Uruguay in a zone along the Río de la Plata and Uruguay River. But from the early 1850s the European immigrants to the Plata region went largely to Argentina, causing the agricultural zone in southern Uruguay to remain static. Livestock grazing and crop cultivation thrived in the north, but only farming was practiced in the south. It was not until after the mid-1940s that there was a great increase in the area devoted to farming and the sharp contrast between the two regions disappeared.

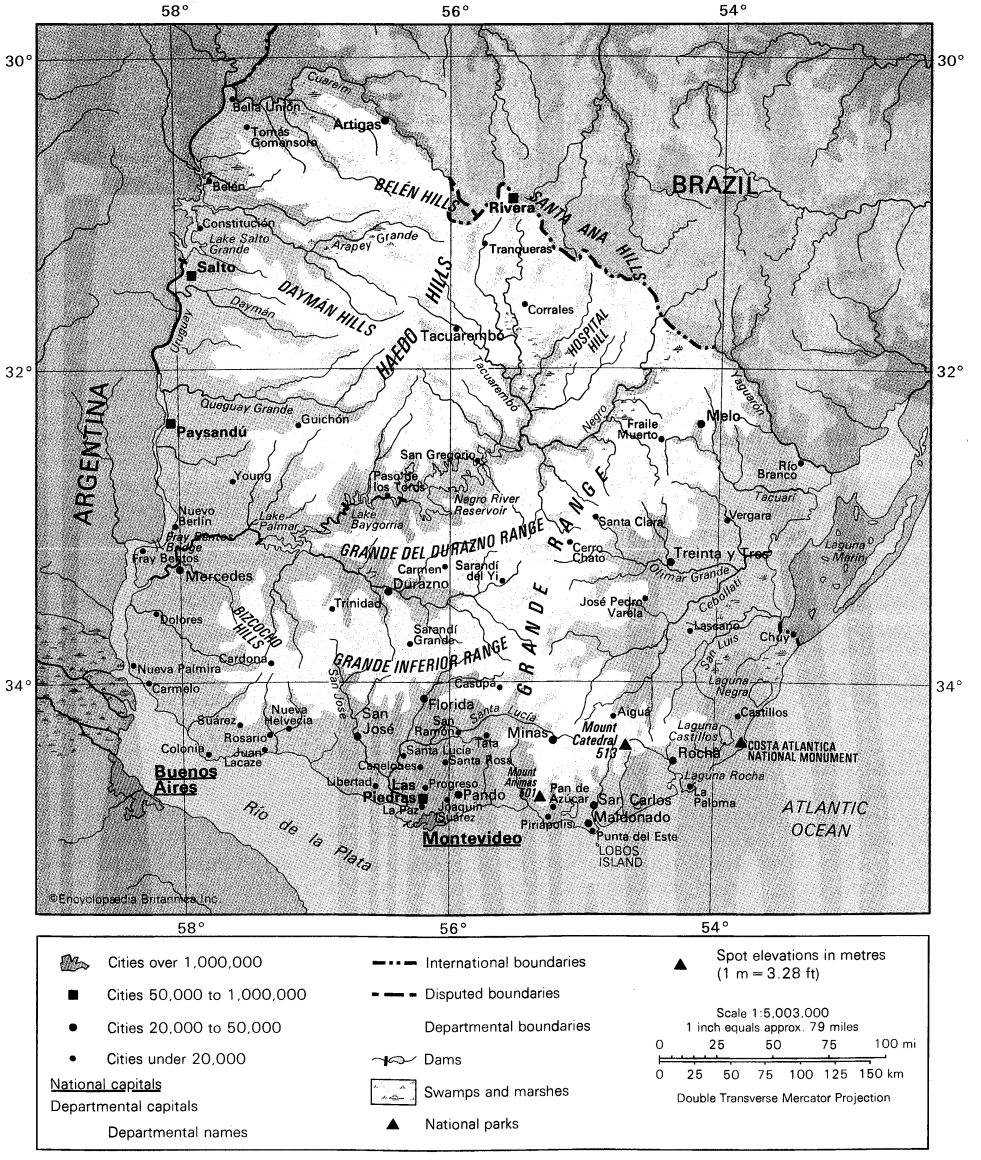
Cattle raising

The grazing of animals in the pastoral region is no longer on unfenced range. Barbed wire is used to separate one pasture from another and to border the highways. Planted pasture grasses have replaced the native grasses. Large ranches, or estancias, some larger than 25,000 acres (10,000 hectares), are common in the pastoral region. Sheep far outnumber cattle in the northwest, but cattle are of major importance south of the Negro River.

Montevideo, on the coast, developed as the country's one large urban centre, with a city population exceeding 1,000,000. Other, much smaller cities include Salto, Paysandú, Las Piedras, and Rivera. (P.E.J./M.H.A.)

THE PEOPLE

The Uruguayans of today are predominantly of European origin, mostly descendants of 19th- and 20th-century immigrants from Spain and Italy and, to a lesser degree,



MAP INDEX

Political subdivisions

Artigas ..... 30 35 s 57 00 w
Canelones ..... 34 35 s 56 00 w
Cerro Largo .... 32 20 s 54 20 w
Colonia ..... 34 10 s 57 30 w
Durazno ..... 33 05 s 56 05 w
Flores ..... 33 35 s 56 50 w
Florida ..... 33 50 s 55 55 w
Lavalleja ..... 34 00 s 55 00 w
Maldonado ..... 34 40 s 54 55 w
Montevideo ..... 34 50 s 56 10 w
Paysandú ..... 32 00 s 57 15 w
Río Negro ..... 32 45 s 57 20 w
Rivera ..... 31 30 s 55 15 w
Rocha ..... 34 00 s 54 00 w
Salto ..... 31 25 s 57 00 w
San José ..... 34 15 s 56 45 w
Soriano ..... 33 30 s 57 45 w
Tacuarembó .... 32 10 s 55 30 w
Treinta y Tres ... 33 00 s 54 15 w

Cities and towns

Aiguá ..... 34 12 s 54 45 w
Artigas ..... 30 24 s 56 28 w
Belén ..... 30 47 s 57 47 w
Bella Unión ..... 30 15 s 57 35 w
Canelones ..... 34 32 s 56 17 w
Cardona ..... 33 53 s 57 23 w
Carmelo ..... 34 00 s 58 17 w
Carmen ..... 33 15 s 56 01 w
Castillos ..... 34 12 s 53 50 w
Casupá ..... 34 07 s 55 39 w
Cerro Chato .... 33 06 s 55 08 w
Chuy ..... 33 41 s 53 27 w
Colonia ..... 34 28 s 57 51 w
Constitución .... 31 05 s 57 50 w
Corrales ..... 31 35 s 55 28 w
Dolores ..... 33 33 s 58 13 w
Durazno ..... 33 22 s 56 31 w
Florida ..... 34 06 s 56 13 w
Fraile Muerto .... 32 31 s 54 32 w
Fray Bentos .... 33 08 s 58 18 w
Guichón ..... 32 21 s 57 12 w
Joaquín Suárez .. 34 44 s 56 02 w
José Pedro Varela ..... 33 27 s 54 32 w
Juan Lacaze .... 34 26 s 57 27 w
La Paloma ..... 34 40 s 54 10 w
La Paz ..... 34 46 s 56 15 w
Las Piedras ..... 34 44 s 56 13 w
Lascano ..... 33 40 s 54 12 w
Libertad ..... 34 38 s 56 39 w
Maldonado ..... 34 54 s 54 57 w
Melo ..... 32 22 s 54 11 w
Mercedes ..... 33 16 s 58 01 w
Minas ..... 34 23 s 55 14 w
Montevideo ..... 34 53 s 56 11 w
Nueva Helvecia .. 34 18 s 57 14 w
Nueva Palmira ... 33 53 s 58 25 w
Nuevo Berlín .... 32 59 s 58 03 w
Pan de Azúcar .. 34 48 s 55 14 w
Pando ..... 34 43 s 55 57 w
Paso de los Toros ..... 32 49 s 56 31 w
Paysandú ..... 32 19 s 58 05 w
Piriápolis ..... 34 54 s 55 17 w
Progreso ..... 34 40 s 56 13 w
Punta del Este .. 34 58 s 54 57 w
Río Branco ..... 32 34 s 53 25 w
Rivera ..... 30 54 s 55 31 w
Rocha ..... 34 29 s 54 20 w
Rosario ..... 34 19 s 57 21 w
Salto ..... 31 23 s 57 58 w
San Carlos ..... 34 48 s 54 55 w
San Fructuoso, see Tacuarembó
San Gregorio .... 32 37 s 55 40 w
San José ..... 34 20 s 56 42 w
San Ramón ..... 34 18 s 55 58 w
Santa Clara ..... 32 55 s 54 58 w
Santa Lucía ..... 34 27 s 56 24 w
Santa Rosa ..... 34 30 s 56 03 w
Sarandí del Yi ... 33 21 s 55 38 w
Sarandí Grande .. 33 44 s 56 20 w
Suárez (Tarariras) ..... 34 17 s 57 37 w
Tacuarembó (San Fructuoso) . 31 44 s 55 59 w
Tala ..... 34 21 s 55 46 w
Tarariras, see Suárez
Tomás Gomensoro .... 30 26 s 57 26 w
Tranqueras ..... 31 12 s 55 45 w
Treinta y Tres ... 33 14 s 54 23 w
Trinidad ..... 33 32 s 56 54 w
Vergara ..... 32 56 s 53 57 w
Young ..... 32 41 s 57 38 w

Physical features and points of interest

Ánimas, Mount .. 34 48 s 55 23 w
Arapey Grande, *river* ..... 30 55 s 57 49 w
Atlantic Ocean .. 35 00 s 53 00 w
Baygorria, Lake, *reservoir* ..... 32 54 s 56 49 w
Belén Hills ..... 30 55 s 56 30 w
Bizcocho Hills ... 33 30 s 57 50 w
Castillos, Laguna . 34 20 s 53 54 w
Catedral, Mount . 34 23 s 54 43 w
Cebollatí, *river* ... 33 09 s 53 38 w
Costa Atlantica National Monument .... 34 20 s 53 30 w
Cuareim, *river* ... 30 12 s 57 36 w
Daymán, *river* ... 31 30 s 58 02 w
Daymán Hills .... 31 38 s 57 10 w
Grande del Durazno Range ..... 33 05 s 56 00 w
Grande Inferior Range ..... 33 50 s 56 10 w
Grande Range .. 33 15 s 55 07 w
Haedo Hills ..... 31 40 s 56 18 w
Hospital Hill .... 31 53 s 54 33 w
Lobos Island .... 35 02 s 54 53 w
Merín, Laguna ... 32 45 s 52 50 w
Negra, Laguna .. 34 03 s 53 40 w
Negro, *river* ..... 33 24 s 58 22 w
Negro River Reservoir ..... 32 45 s 56 00 w
Olimar Grande, *river* ..... 33 16 s 53 52 w
Palmar, Lake, *reservoir* ..... 33 06 s 57 21 w
Plata, Río de la, *estuary* ..... 35 00 s 57 00 w
Queguay Grande, *river* ... 32 09 s 58 09 w
Rocha, Laguna .. 34 38 s 54 18 w
Salto Grande, Lake ..... 31 14 s 57 55 w
San José, *river* .. 34 38 s 56 29 w
San Luis, *river* ... 33 32 s 53 33 w
Santa Ana Hills .. 31 15 s 55 15 w
Santa Lucía, *river* ..... 34 48 s 56 22 w
Tacuarembó, *river* ..... 32 25 s 55 29 w
Tacuarí, *river* .... 32 46 s 53 18 w
Uruguay, *river* ... 34 12 s 58 18 w
Yaguarón, *river* .. 32 39 s 53 12 w
Yi, *river* ..... 33 07 s 57 08 w

France and Britain. Earlier settlers had migrated from Argentina and Paraguay. Few descendants of Uruguay's original population remain. Of the small number of blacks in the country, most came southward from Brazil.

Uruguay is one of the few Latin-American republics not overwhelmingly Roman Catholic. According to Uruguayan census data, about 60 percent of the people identify themselves as Roman Catholic, but, in terms of practicing communicants, there are indications of a still lower percentage. The establishment of Methodist churches in provincial cities attests to the Protestant missionary efforts of the 1920s. There are also sizable congregations of Anglicans and other Protestant groups. Jews, mostly in Montevideo, make up a very small minority group in Uruguay, but they are nevertheless one of the larger Jewish communities in South America. Mormon missionaries have become increasingly active.

Language

Without an Indian minority, Uruguayans universally speak Spanish, although in the borderland towns close to Brazil an admixture of Portuguese and Spanish words and phrases can be heard. In and around the border city of Rivera, the bilingualism has yielded a slang called Portunol, from the words Portuguese and *español.*

Of the 19 *departamentos* into which the republic is administratively divided, the *departamento* of Montevideo, almost coterminous with the national capital, is the most populous, followed by the *departamento* of Canelones; none of the other *departamentos* is close to those two in population. Nevertheless, Uruguay is a highly urban country, with more than 80 percent of the people classified as urban. The population, which was estimated at more than 3,000,000 in the late 1980s, has grown much more slowly than that of most other Latin-American countries. Uruguay's birth rate and proportion of young people in the population are lower than those of most other South American countries. (M.H.A.)

Demographics

THE ECONOMY

The foundation of Uruguay's economy is said to have been laid in 1603, when a governor of Paraguay, Hernando Arias de Saavedra, shipped a number of cattle and horses downstream from Asunción. The animals were landed on the Uruguayan riverbank, where they were left to run wild. Later in the century the herds were so abundant that they attracted gauchos, who crossed the Río de la Plata from Buenos Aires and began a trade in hides. As more cattlemen arrived, boundaries had to be fixed, setting the stage for the development of the great estancias of the country.

Uruguay's gross national product (GNP) per capita is among the highest in Latin America. The nation's relatively high standard of living continues to be closely related to the earnings from pastoral and agricultural exports; economic well-being is somewhat precarious because these primary products are subject to sudden fluctuations in world demand and prices. To reduce the nation's dependence on external trade, successive governments have encouraged the development of domestic industry by means of protective tariffs, import controls, exemptions of machinery from import duties, and preferential exchange rates. Because of a lack of local sources, Uruguay is obliged to import most of its fuel, industrial raw materials, vehicles, and industrial machinery.

Structure of the economy

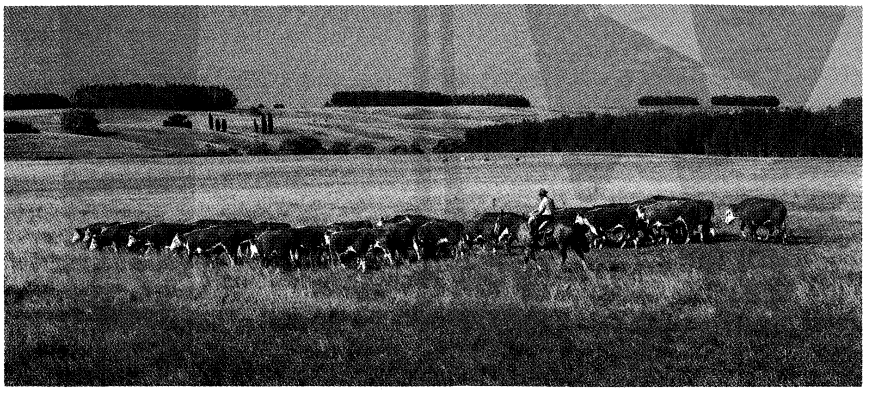
Although Uruguay is a heavily agricultural country, a large urban middle class dominates its society. The country is noted for its highly developed social services, but these have placed an excessive burden on national resources.

Resources. Uruguay has no domestic commercial sources of petroleum, natural gas, coal, or iron, a leading factor in the limitation of industrial growth in the country. The low, rolling countryside of Uruguay is not generally suited to hydroelectric development, and most of the electric installations were formerly dependent on imported coal or oil. This changed, however, after the hydroelectric plants built along the Negro and Uruguay rivers began producing at full power by the early 1980s.

Agriculture. Pastoral farming, chiefly sheep and cattle raising, is Uruguay's most important economic activity. Livestock, livestock products, and skins and hides account for about 40 percent of Uruguay's income from exports. Beef and wool are particularly important. All available land is in use for agriculture, and production can be increased only by improved techniques. With the major emphasis on livestock, little of the arable land has been available for field crops. Consequently, increases in farm products have been rare, rice production being an exception. Major crops in addition to rice include sugarcane, wheat, sugar beets, sorghum, potatoes, grapes, and corn. The grape harvest sustains a local wine industry.

Industry. The state operates a large number of public corporations. It controls electricity and the refining of

The public corporations



Gaucho herding cattle on a ranch in central Uruguay.
Carl Frank—Photo Researchers

imported petroleum, manufactures alcohol and cement, directs the largest meat-packing plants and the processing of fish, and controls the railways, several banking institutions, and most insurance sales.

The processing of food and other products of the land is the basic form of industry. The subsidized wool-combing industry has added to the country's exports. Consumer goods manufactured locally include textiles, tires and other rubber goods, shoes, and household appliances.

Finance and trade. Both state and private banks operate in Uruguay, the private banks being regulated by the Central Bank of Uruguay (founded in 1967), which is also the note-issuing bank. Other state banks include the Bank of the Eastern Republic of Uruguay, which is the country's largest commercial bank, with responsibilities in regulating foreign trade, and the Mortgage Bank of Uruguay. Uruguay's balance of payments has been in deficit for much of the post-World War II period, although there have been intermittent periods of surplus. As noted above, the main exports are related to livestock raising, particularly cattle, and include wool, meat, and hides and skins. The chief imports include crude petroleum, machinery and appliances, chemical products, plastics, resins, rubber, and transport equipment. Traditionally, the main trading partner of Uruguay has been Brazil; others of importance include the United States, Argentina, Nigeria, and West Germany.

Transportation. The Uruguayan landscape offers little in the way of impediments to transportation, and a comparatively extensive road and rail network has been developed, Montevideo serving as its hub. Paved roads connect the city to other urban centres in the country, the main highways leading to the border and neighbouring cities. Numerous unpaved roads connect farms and small towns. Most of the country's domestic freight and passenger service is by road, the railroads having declined somewhat as the road system grew and vehicular traffic increased. The basic railroad network, purchased from the British after World War II, also radiates from Montevideo and connects with the Argentine and Brazilian systems.

Railroads

Most of Uruguay's external trade is handled by foreign shipping, its own fleet being inadequate to carry much of the volume. Vessels of various sizes navigate the inland waters, and a hydrofoil service connects Buenos Aires and Montevideo across the Río de la Plata. The international airport lies near the Carrasco beach resort some 13 miles from downtown Montevideo. The government-owned airline (Primeras Líneas Uruguayas de Navegación Aérea; PLUNA) links Montevideo to the provincial capitals and provides flights to neighbouring countries.

(I.C.Cn./G.Pe./C.P.D./M.H.A.)

ADMINISTRATION AND SOCIAL CONDITIONS

Government. Uruguayans lived under successive constitutional governments from 1903 until 1973, when, under deteriorating economic conditions, the military took control to quell civil disorder fomented chiefly by an urban guerrilla group, the Tupamaros, and other civil dissenters. Constitutional rights guaranteed by the 1966 constitution were suspended during the military regime, which lasted until 1985. The elected government that took office in that year once again followed the structure of the 1966 constitution. A president and the Council of Ministers hold executive power, and a vice president serves as president of the legislature. The president and vice president are elected for five-year terms by a simple majority vote. The legislative branch, the General Assembly, consists of the Senate and the Chamber of Representatives, whose members are elected for five years by direct popular vote. All Uruguayans 18 years of age or older may vote.

The legislature, or Departmental Board, for each of the 19 *departamentos* formulates local policies. Each departmental capital has a chief executive who combines some of the duties of a departmental governor and a mayor under the title *intendente municipal.*

The two principal traditional political parties are the Colorado Party (which traditionally has had a liberal urban base), and the Blanco, or National, Party (supported by the more conservative landowners). A third party, the Broad Front (Frente Amplio), representing a coalition of Christian Democrats, Socialists, Communists, and dissident members of the two other parties, was outlawed by the military regime, but the member parties of the coalition were reinstated in 1984 and 1985. The revolutionary Tupamaros were also allowed to form a political party after the civilian government took office.

Political parties

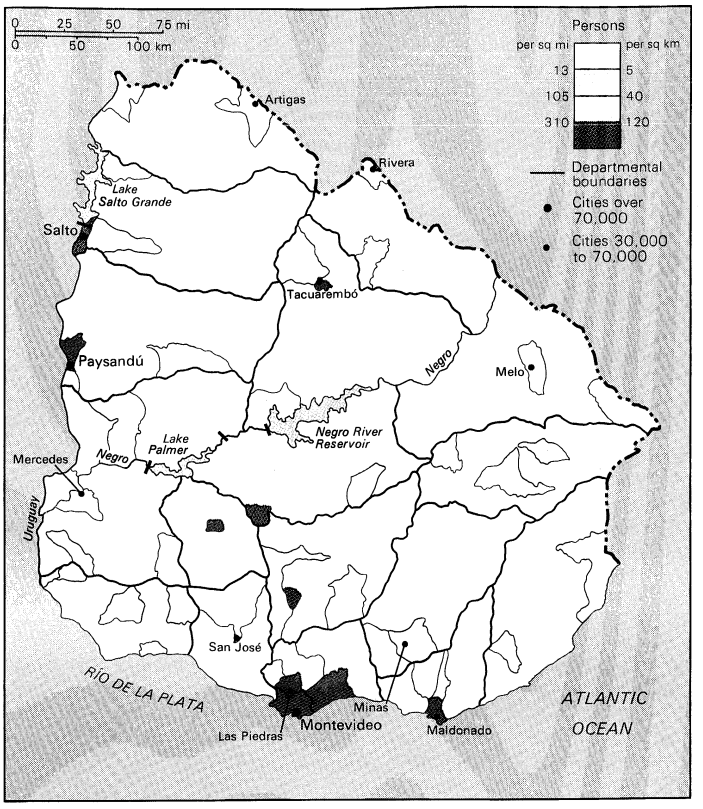
At the head of the judiciary is the Supreme Court, composed of five justices. They are elected by the General Assembly for a term of 10 years, and they are eligible for reelection five years after their previous term ends. A screening commission presents a list of outstanding candidates to the assembly for debate and selection for each vacancy. The Appellate Tribunals, each composed of three judges, form the next-highest judicial level, followed by the Courts of Record. The Supreme Court justices select Appellate Tribunal judges for confirmation by the Senate. The Electoral Court of nine judges monitors elections, both national and local.

Education. Uruguay has long been renowned for its progress in education, after reforms instituted in the late 19th century that established a system of free, compulsory, coeducational, and secular education. This progress is reflected in the high literacy rate, which exceeds that of most Latin-American countries.

The University of the Republic (founded in 1849) has 10 faculties, including a distinguished medical school that draws students from many South American countries. It also includes the School of Agronomy and the Federico E. Capurro School of Librarianship and Related Sciences. There is a privately supported Institute of Higher Studies devoted to scientific research. Vocational training is given by the Labour University of Uruguay.

Higher education

Health and welfare. Since the enactment of elaborate legislation in 1912 and in 1929, Uruguay has long been recognized for its advanced welfare programs offering extensive subsidized health care and numerous benefits to the unemployed, low-wage workers, and the aged. Uruguayan employees with low annual incomes may receive maternity benefits, and mothers who earn low wages can receive child-care benefits. The huge Hospital de Clínicas has



Population density of Uruguay.

been for decades a low-cost medical service centre for the needy as well as a research centre. (G.I.B./M.H.A.)

CULTURAL LIFE

Uruguayan culture reflects some of the same characteristics as are found in neighbouring Argentina. Both countries are strongly European in their orientation and, unlike many South American countries, Uruguay is minimally influenced by the original Indian inhabitants. The tradition of the gaucho has been an important element in the art and folklore of both countries. Uruguay's theatre and music have not been the province of a small elite, as in Bolivia and Peru and some other Latin-American nations, but rather are broadly based in terms of support and participation. The Italian influence in language and customs is also shared by Uruguay and Argentina.

The arts. José Enrique Rodó, a modernist, is considered Uruguay's most significant literary figure. His book *Ariel* (1900), which stresses the importance of upholding spiritual (Latin-American) over materialistic (U.S.) values, continues to influence young writers. Outstanding among Latin-American playwrights is Florencio Sánchez; his plays, written around the beginning of the 20th century and dealing with contemporary social problems, are still performed. From about the same period and somewhat later came the romantic poetry of Juan Zorrilla de San Martín, Juana de Ibarbourou, and Delmira Agustini and the short stories of Horacio Quiroga. In the 1960s and '70s the works of the radical writer Carlos Quijano and the economic historian Luis Carlos Benvenuto remained highly popular.

Leading writers

The folk and popular music of Uruguay shares with Argentina not only its gaucho roots but also the tango, a dance form that originated in Argentina. One of the most famous tangos, "La Cumparsita" (1917), was written by the Uruguayan composer Gerardo Matos Rodríguez. Early classical forms showed heavy Spanish and Italian influence, but in the 20th century a number of composers of classical music, including Eduardo Fabini, Vicente Ascone, and Héctor Tosar, have made use of Latin-American musical idioms. The 19th-century painter Juan Manuel Blanes, whose works depict historical events, was the first Uruguayan artist to gain widespread recognition. The Postimpressionist painter Pedro Figari achieved international recognition for his pastel studies of subjects in Montevideo and the countryside.

Cultural institutions. Uruguay's principal cultural institutions are in Montevideo. The National Library is located there, as are the national museums of history, natural history, and art. The government supports two symphony orchestras and a theatre company, the National Theatre, and it sponsors schools of dramatic arts, fine arts, and ballet. Private dramatic and musical groups also perform in Montevideo and other cities.

Recreation. Football (soccer) remains the major sport in number of both participants and spectators, but in the 1980s basketball became a runner-up. Water sports are popular, and resort areas, particularly on the coast, attract visitors throughout most of the year. Among these is Punta del Este, renowned as a meeting place for high-level international conferences. The most important festival, Carnival Week, is held just before Lent in Montevideo.

Press and broadcasting. Most of the Uruguayan daily newspapers are published in Montevideo, and several have national circulation. The major dailies are owned by or affiliated with the principal political parties. *El Día,* the nation's most prestigious paper, was founded in 1886 by José Batlle y Ordóñez and has been associated with his Colorado Party. *El País,* the paper of the rival conservative Blanco, or National, Party, has the largest circulation. Both government and private radio and television stations operate in Uruguay. Radio broadcasting began as a daily service in 1922, and the first television station began broadcasting in 1956. (M.H.A.)

Major newspapers

For statistical data on the land and people of Uruguay, see the *Britannica Word Data* section in the BRITANNICA WORLD DATA ANNUAL.

History

BEFORE THE EUROPEANS

Before the arrival of Europeans, the territory which is now Uruguay supported a small population estimated at no more than 5,000 to 10,000. The principal groups were the Charrúa and Chaná Indians. These seminomadic tribes did not develop the tools necessary to farm Uruguay's grasslands. They moved to the shore in summer to fish and gather clams, fruits, and roots, and inland in the winter to hunt deer, rheas, and smaller game with bolas and bows and arrows. Bands of eight to 12 families under a chief lived in villages of five to six houses made of matted windscreens. The fierce Charrúa attacked enemies to expand hunting grounds and capture women and children.

COLONIAL PERIOD

Though the area was explored by the Spanish explorer Juan Díaz de Solís in 1516, the absence of mineral wealth or sedentary Indians who could be compelled to work made the Banda Oriental del Uruguay (the east bank of the Uruguay River) unattractive for Spanish settlement. Cattle from neighbouring regions, allowed to roam freely in Uruguayan territory, multiplied over the years until their numbers reached the millions. They were hunted for their hides by gauchos, cowboys of mixed Spanish-Indian ancestry, who nevertheless did not settle the land. In 1680 Portuguese Brazil established Colônia do Sacramento in the Banda Oriental on the Río de la Plata opposite Buenos Aires in order to trade with the Spanish settlers in contraband and siphon off silver from Peru. Spanish authorities countered this move by founding San Felipe de Montevideo as a fortified city in 1726 and attacking Colônia do Sacramento, which was finally ceded to Spain in 1777. Montevideo became the major Spanish port of the South Atlantic, and the process of dividing the Banda Oriental into huge unfenced ranches began. In 1776 the Banda Oriental became part of the Viceroyalty of la Plata, the capital of which was Buenos Aires.

The founding of Montevideo

THE STRUGGLE FOR NATIONAL IDENTITY (1810–80)

When the independence movement in the Viceroyalty of la Plata broke out in Buenos Aires in 1810, Montevideo with its Spanish military and naval contingents was a roy-

alist stronghold. In the interior of the Banda Oriental, José Gervasio Artigas, commander of the Blandengues, a corps of irregulars, led the fight against Spain. He was supported by rural inhabitants, antiroyalist Montevideo leaders, and an army from Buenos Aires. Artigas won first the interior and then Montevideo. The victorious Artigas wanted a loose confederation of provinces of la Plata, and he also considered a rival confederation to Buenos Aires centring on Montevideo. These plans, coupled with Artigas' growing power and egalitarian policies, made him a threat to Buenos Aires, which acquiesced when the Portuguese, seizing the opportunity, took over the Banda Oriental in 1820, driving Artigas into exile.

"Brazilianization" was resisted within the Banda Oriental and by Uruguayan exiles as well. The Argentines were also disturbed, and their government felt compelled to support Juan Antonio Lavalleja, one of Artigas' exiled officers, and his "33 *orientales*" when they crossed the river to free their homeland in 1825. The ensuing war was a stalemate, but British diplomats mediated a settlement in 1827, and in 1828 a treaty was ratified making the province La República Oriental del Uruguay, a separate nation to serve as a buffer between Brazil and Argentina.

On July 18, 1830, when the constitution for the new nation was approved, Uruguay had scarcely 74,000 inhabitants. Twenty years of war and depredation had greatly reduced cattle numbers, and the lands and fortunes of many colonial families had been destroyed. Both Argentina and Brazil still coveted Uruguay. Not surprisingly, the new nation's first years of independence were disastrous. The factions of the first and second presidents, José Fructuoso Rivera and Manuel Oribe, battled each other. Oribe's adherents, who wore white, became the Blanco Party and controlled the interior. From 1843 to 1851, supported by armies of the Argentine dictator Juan Manuel de Rosas, the Blancos besieged Rivera and his followers, who wore red and became the Colorado Party, based in Montevideo. The Colorados were aided first by France and England, then by Brazil. When in 1851 the Great War ended without a clear victory for either side, the Uruguayan interior was devastated, the government was bankrupt, and the disappearance of an independent Uruguay had become a real possibility. Intellectuals wanted to abolish the political parties that had brought the country so low, but the war had made too deep an impact on ordinary Uruguayans, who now felt themselves polarized into Colorados or Blancos. Political disputes, civil wars, and Brazilian and Argentine involvement (the Colorados repaid the help given by these countries in an 1865 victory over the Blancos by joining Brazil and Argentina in their war against Paraguay, 1865–70) kept Uruguay from taking advantage of the export opportunities opened by Europe's industrialization. In the 1870s the military took over the government.

Origin of the Blancos and Colorados

THE MODERNIZATION OF URUGUAY (1880–1930)

The military's brutality, corruption, and internal dissension reduced their public support so severely that civilian Colorado government returned in 1890. It was, however, the military's temporary pacification of the interior during the previous decade that permitted the acceleration of the economic development begun earlier: the fencing in of ranches, the introduction of sheep to graze together with cattle, and the improvement of livestock by the importation of pedigreed bulls and rams. Wool, hide, and dried-beef exports fueled railroad building and other British investments as well as the modernization of Montevideo, which in turn encouraged immigration. By 1900, Uruguay's population was 13 times greater than that at independence. Even so, Uruguay had only a million inhabitants, and immigration lagged far behind that of Argentina. Ongoing conflict between Colorados and Blancos continued to impede Uruguay's modernization.

Blanco demands for a larger role in government had escalated into the Revolution of 1897, led by Aparicio Saravia, which ended when the Colorado president, Juan Idiarte Borda, was killed by an assassin not associated with the Blancos. With tensions mounting, José Batlle y Ordóñez, a strong Colorado leader, was elected president in 1903. In the next year the Blancos revolted, and eight bloody months of fighting ensued before Saravia was killed in battle and government forces emerged victorious. This was followed in 1905 by a Colorado victory in the first legislative elections in 30 years the results of which had not been arranged beforehand.

Accession of José Batlle y Ordóñez

Batlle, now the Colorado hero, used the economic prosperity that came with domestic stability to change Uruguay. He instituted policies of state intervention to expand ranching production, reduce import dependency and the role of foreign capital, and improve workers' conditions through far-reaching social reforms. State economic intervention was combined with expanded education and a reduction of the role of the Roman Catholic church; legislation was passed covering divorce, the end of the death penalty, and rights for illegitimate children.

Batlle y Ordóñez was president for two terms, 1903–07 and 1911–15, but he realized that his program required many years to be fully implemented and was in danger of being reversed by a future president or dictator. His solution was constitutional reform to end the presidency and replace it with a plural executive, the *colegiado.* Such audacity split the Colorados and reinvigorated the Blanco opposition. The *colegiado* was defeated in 1916 in an election for a constitutional convention in which the secret ballot was used for the first time. Batlle's prestige, however, was sufficiently great to partially rescue the *colegiado,* and a compromise system splitting executive responsibility between the president and a National Council of Administration went into effect in 1919.

The plural executive

The *colegiado*'s 1916 defeat showed that Uruguayans could not fully accept Batlle's reform impetus. What emerged was a consensus government whose policies, except in social legislation, were more cautious than innovative. Higher living standards were supported by a ranching economy that had stopped growing, a dilemma hidden by the high export prices of the late 1920s.

ECONOMIC DECLINE AND MILITARY TAKEOVER (1930–73)

In 1930 the Colorado Gabriel Terra successfully maneuvered his presidential candidacy through the political vacuum created by the death in 1929 of Batlle y Ordóñez, who had held an increasingly complex political and governmental structure together. When the effects of the Great Depression hit Uruguay, President Terra first blamed the plural executive's economic policies and then, supported by Blanco leader Luis Alberto de Herrera, carried out a coup on March 31, 1933, that abolished the National Council and returned Uruguay to full presidential government. Terra's dictatorship, followed by the presidency of his brother-in-law General Alfredo Baldomir from 1938 to 1942, formulated a conservative response to the Great Depression. The state interfered with unions, postponed social legislation, and preserved as much as it could of the British market for Uruguayan meat at the price of giving up on nationalizing foreign, mainly British, enterprises in Uruguay. Although ideologically advocating free-market principles, the government had to involve itself still further in the economy. It apportioned scarce foreign exchange, built a hydroelectric dam, and tried to ease unemployment and maintain political support by hiring public employees under a system of political quotas. Hard times sped migration from the interior to Montevideo. The shortage of foreign exchange due to the shrunken export markets curbed manufactured exports and encouraged the industrialization of Montevideo. From the Great Depression had come an increased urbanization and bureaucratization of Uruguay which was accompanied by a decline in the birth rate.

With World War II, Europe began eagerly to buy Uruguay's meat, wool, and hides, bringing a period of genuine prosperity. The war also strengthened Uruguay's democratic parties, which sought to share the benefits of property more widely. This was confirmed by the 1946 election of Tomás Berreta, a Batllista (member of the Colorado Batllista Party founded by Batlle in 1919) as president. After his sudden death, Vice President Luis Batlle Berres, Batlle's nephew, became president. An economic boom, based on high U.S. wool prices stimulated by the Korean War, enabled Batlle Berres to buy the British-

Batlle Berres' economic program

owned railroads and public utilities, inaugurate new state enterprises, encourage industrialization, subsidize agriculture, and lower food prices. There was full employment. In 1951, constitutional reform replaced the presidency with a nine-member plural executive, the traditional cornerstone of the Batllista program. During this period Uruguay combined a strong democracy with the highest income per capita in Latin America.

By the mid-1950s, when the end of the Korean War lowered wool prices, it was evident that Uruguay's ranching economy could not continue to support the standard of living which the country enjoyed. Politicians, responding to voters' demands, tried to keep consumption up, first by spending Uruguay's foreign exchange, then by borrowing abroad and devaluing the peso. Economic conditions deteriorated, combining annual inflation rates of more than 60 percent, the breakdown of public services, industry closings, and a high emigration rate.

Voter dissatisfaction brought the Blancos to power in 1958 for the first time since 1865. Although reelected for a second term, the Blanco administration failed to improve conditions, and in 1966 a new constitution was ratified, returning the country to the presidential system. Elections in that year brought new leadership under Colorado conservatives, but inflation and a production slump continued to grip the country. The economic crisis precipitated increasingly greater protests and demonstrations. Subsequent government crackdowns on students and unions created an ugly climate intensified by the activities of an urban guerrilla group called the Tupamaros, who fought to bring a socialist revolution through terrorism. When the police could not stop the Tupamaros, the government unleashed the military. This brought defeat to the Tupamaros, but the military then took power in 1973 from the elected politicians, who were by then discredited.

The Tupamaros

THE MILITARY REGIME (1973–85)

The military acted with a ferocity and thoroughness previously unknown to Uruguay. During this period Uruguay was reputed to have the highest ratio of political prisoners to population in the world. Human-rights groups protested numerous cases of abuse, including those of torture, killings, and disappearances. Imprisonment, censorship, suspension of civil political activity, and dissolution of unions were designed not only to control the people but also to force a new economic outlook on them. Foreign banks and lenders were attracted by high interest rates, industrialists and ranchers were encouraged to borrow and modernize, wages were held down, and strikes were forbidden. The government built roads and other public works. Though real wages fell and many businesses failed because they could not compete with cheap imports, the policy had some successes, including the growth of manufactured exports, a building boom, and the reemergence of Montevideo as a banking and financial centre. In 1980 the military decided to institutionalize their control of Uruguay through constitutional plebiscite. To their surprise, since they controlled the media and severely restricted the opposition, voters rejected the military's constitution.

The plebiscite destroyed the military regime's legitimacy, and its economic policy was undermined by the 1982 tightening of foreign loans to Latin America and by the downturn (caused partly by the Falkland Islands war) of the Argentine economy, to which Uruguay was closely tied. Because borrowing from abroad could no longer support the foreign exchange value of the Uruguayan peso, the military, despite previous assurances, had to let the peso fall. Businesses, ranchers, and the government saw their debts dramatically increase. With Uruguay's economic crisis worsening, the military reluctantly negotiated with politicians the return to elections in 1984.

CIVILIAN GOVERNMENT

The Sanguinetti administration

Julio María Sanguinetti, a Colorado Batllista, was elected president in 1984. His great achievement, which he bequeathed to his successors, was the reestablishment of full democracy and human and civil rights. To assure this he sponsored a general amnesty to the military despite popular preference for criminal trials. This move weakened confidence in the president, whose popularity was further eroded by striking workers trying to get back some of the real wages they had lost under the military. Uruguay's enormous foreign debt, $5,000,000,000 for a population of not quite 3,000,000, inhibited economic recovery, but Sanguinetti refused to embark on dramatic economic programs that would carry high risks. In this he generally reflected the cautious mood of Uruguayans of the period. A referendum in April 1989 upheld the amnesty law, but the Colorado Party lost the presidency the following November when the Blancos returned to power for the first time in 23 years with the election of Luis Alberto Lacalle. (M.I.V.)

For later developments in the history of Uruguay, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

For coverage of related topics in the *Macropædia* and *Micropædia*, see the *Propædia*, sections 964, 966, and 974, and the *Index*.

BIBLIOGRAPHY

Physical and human geography: General introductions to the geography, economy, culture, and history of Uruguay include GEORGE PENDLE, *Uruguay*, 3rd ed. (1963, reprinted 1985); MARVIN ALISKY, *Uruguay: A Contemporary Survey* (1969); THOMAS E. WEIL *et al.*, *Area Handbook for Uruguay* (1971); JORGE CHEBATAROFF, *Geografía de la República Oriental del Uruguay* (1979); and PRESTON E. JAMES, C.W. MINKEL, AND EILEEN W. JAMES, "Uruguay," in *Latin America*, 5th ed. (1986), pp. 450–462. Studies of the people include CARLOS M. RAMA, *Las clases sociales en el Uruguay: estructura, morfología* (1960); and CÉSAR A. AGUIAR, *Uruguay: país de emigración* (1982). Works concerning the economy and the government include LUIS MACADAR, *Uruguay, 1974–1980: un nuevo ensayo de reajuste económico?* (1982); JORGE NOTARO, *La política económica en el Uruguay, 1968–1984* (1984); LUIS COSTA BONINO, *Crisis de los partidos tradicionales y movimiento revolucionario en el Uruguay*, trans. from French (1985); GERMAN W. RAMA, *La democracia en Uruguay* (1987); and MARTIN WEINSTEIN, *Uruguay, Democracy at the Crossroads* (1988). See also ALEJANDRO ROVIRA, *Subversion, Terrorism, Revolutionary War: The Uruguayan Experience* (1981; originally published in Spanish, 1981).

History: *Historia Uruguaya*, 7 vol. (1975–87), is a multiauthored work covering Uruguayan history from European discovery to 1958, published by Ediciones de la Banda Oriental. The politics, economy, and social system during the period 1830–1930 are described in detail in EDUARDO ACEVEDO, *Anales históricos del Uruguay*, 6 vol. (1933–36); JUAN E. PIVEL DEVOTO and ALCIRA RANIERI DE PIVEL DEVOTO, *Historia de la República Oriental del Uruguay (1830–1930)*, 3rd ed. (1966); JUAN A. ODDONE, "The Formation of Modern Uruguay, *c.* 1870–1930," chap. 13 in LESLIE BETHELL (ed.), *The Cambridge History of Latin America*, vol. 5 (1986), pp. 453–474; and ALBERTO ZUM FELDE, *Proceso histórico del Uruguay*, 10th ed. (1987). PETER WINN, "British Informal Empire in Uruguay in the Nineteenth Century," *Past & Present*, no. 73, pp. 100–126 (November 1976), explains British political and economic imperialism in Uruguay. JOSÉ P. BARRÁN and BENJAMÍN NAHUM, *Historia rural del Uruguay moderno*, 7 vol. in 8 (1967–78), is the major interpretive history of rural Uruguay. JOHN STREET, *Artigas and the Emancipation of Uruguay* (1959), is a concise, documented study. SIMON G. HANSON, *Utopia in Uruguay: Chapters in the Economic History of Uruguay* (1938, reprinted 1982), studies the economic effects of Batlle's reforms from 1911 to 1930. GÖRAN G. LINDAHL, *Uruguay's New Path: A Study in Politics During the First Colegiado, 1919–33*, trans. from Swedish (1962), reviews politics of the first plural executive. JOSÉ P. BARRÁN and BENJAMÍN NAHUM, *Batlle, los estancieros y el Imperio Británico*, 8 vol. (1979–87), analyzes the sources and resistance to Batlle y Ordóñez' reforms to 1916; the first volume is also a pathbreaking history of turn-of-the-century Uruguay. MILTON I. VANGER, *José Batlle y Ordoñez of Uruguay: The Creator of His Times, 1902–1907* (1963, reprinted 1980), and *The Model Country: José Batlle y Ordoñez of Uruguay, 1907–1915* (1980), are the most important biographies of Batlle, with historical analyses of the era of his presidencies. M.H.J. FINCH, *A Political Economy of Uruguay since 1870* (1981), presents the best historical explanation of Uruguay's economic stagnation. More recent events are described in EDY KAUFMAN, *Uruguay in Transition: From Civilian to Military Rule* (1979), on the 1973 military coup and its aftermath; CHARLES G. GILLESPIE, "Uruguay's Transition from Collegial Military-Technocratic Rule," in GUILLERMO O'DONNELL, PHILIPPE C. SCHMITTER, and LAURENCE WHITEHEAD (eds.), *Transitions from Authoritarian Rule: Latin America* (1986), pp. 173–195; and JUAN RIAL ROADE, *Uruguay, elecciones de 1984: sistema electoral y resultados* (1986). (M.H.A./M.I.V.)

Velázquez

The most important Spanish painter of the 17th century, Diego Velázquez is universally acknowledged as one of the world's greatest artists. The naturalistic style in which he was trained provided a language for the expression of his remarkable power of observation in portraying both the living model and still life. Stimulated by the study of 16th-century Venetian painting, he developed from a master of faithful likeness and characterization into the creator of masterpieces of visual impression unique in his time. With brilliant diversity of brushstrokes and subtle harmonies of colour, he achieved effects of form and texture, space, light, and atmosphere, that make him the chief forerunner of 19th-century French Impressionism.

The principal source of information about Velázquez' early career is the treatise *Arte de la pintura* ("The Art of Painting"), published in 1649 by his master and father-in-law Francisco Pacheco, who is more important as a biographer and theoretician than as a painter. The first complete biography of Velázquez appeared in the third volume (*El Parnaso español;* "The Spanish Parnassus") of *El Museo pictórico y escala óptica* ("The Pictorial Museum and Optical Scale"), published in 1724 by the court painter and art scholar Antonio Palomino. This was based on biographical notes made by Velázquez' pupil Juan de Alfaro, who was Palomino's patron. The number of personal documents is very small, and official documentation relating to his paintings is relatively rare. Since he seldom signed or dated his works, their identification and chronology has often to be based on stylistic evidence alone. Though many copies of his portaits were evidently made in his studio by assistants, his own production was not large and his surviving autograph works number fewer than 150. He is known to have worked slowly, and during his later years much of his time was occupied by his duties as a court official in Madrid.

Seville. Born in Seville, Diego Rodríguez de Silva Velázquez was baptized on June 6, 1599. According to Palomino, his first master was the Sevillian painter Francisco Herrera the Elder (*c.* 1576–1656). In 1611 he was formally apprenticed to Francisco Pacheco, whose daughter he married in 1618. "After five years of education and training," Pacheco writes, "I married him to my daughter, moved by his virtue, integrity, and good parts and by the expectations of his disposition and great talent." Although Pacheco was himself a mediocre Mannerist painter, it was through his teaching that Velázquez developed his early naturalistic style. "He worked from life," writes Pacheco, "making numerous studies of his model in various poses and thereby he gained certainty in his portraiture." He was not more than 20 when he painted the "Water Seller of Seville" (*c.* 1619; Wellington Museum, London), in which the control of the composition, colour, and light, the naturalness of the figures and their poses, and realistic still life already reveal his keen eye and prodigious facility with the brush. The strong modelling and sharp contrasts of light and shade of Velázquez' early illusionistic style closely resemble the technique of dramatic lighting called tenebrism, which was one of the innovations of the Italian painter Caravaggio (1573–1610). Velázquez' early subjects were mostly religious or genre (scenes of daily life). He popularized a new type of composition in Spanish painting, the *bodegón,* a kitchen scene with prominent still life, such as the "Old Woman Frying Eggs." Sometimes the *bodegones* had religious scenes in the background, as in "Christ in the House of Martha and Mary." The "Adoration of the Magi" is one of the few Sevillian paintings of Velázquez that have remained in Spain.

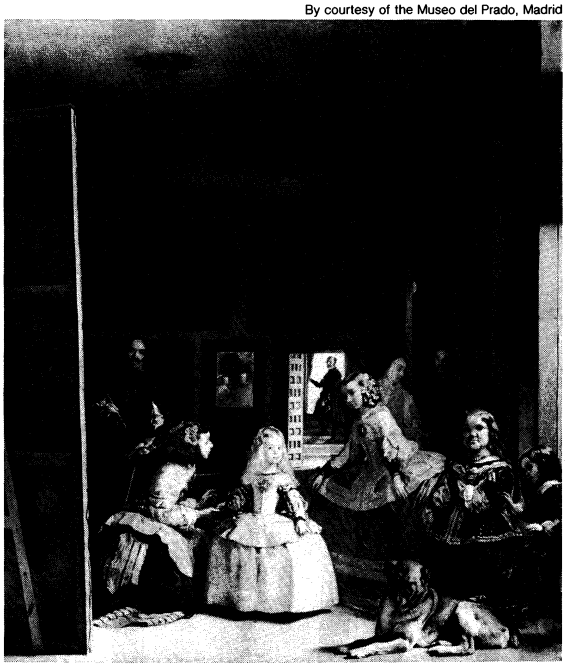
First paintings

Court painter in Madrid. In 1622, a year after Philip IV came to the throne, Velázquez visited Madrid for the first time, in the hope of obtaining royal patronage. He painted a portait of the poet Luis de Góngora (1622; Museum of Fine Arts, Boston), but there was no opportunity of portraying the King or Queen. In the following year he was recalled to Madrid by the prime minister, Count Olivares, a fellow Sevillian and a future patron. Soon after his arrival he painted a portrait of Philip IV that won him immediate success. He was appointed court painter with a promise that no one else should portray the King. Pacheco describes an equestrian portrait of Philip (lost) painted soon afterward, "all taken from life, even the landscape"; the portrait was exhibited publicly "to the admiration of all the Court and the envy of members of the profession." The envy of fellow artists, who accused Velázquez of only being able to paint heads, is said to have been the occasion of the King ordering him to paint a historical subject, the "Expulsion of the Moriscos" (lost), in competition with other court painters. Velázquez was awarded the prize and the appointment in 1627 of gentleman usher to the King. Though he continued to paint other subjects, as court painter he was chiefly occupied in portraying members of the royal family and their entourage, and he painted numerous portraits of Philip IV during the course of his life. "The liberality and affability with which he is treated by such a great monarch is unbelievable," writes Pacheco. "He has a workshop in his gallery and His Majesty has a key to it and a chair in order to watch him painting at leisure, nearly every day."

Velázquez' position at the court gave him access to the royal collections, rich in paintings by the Venetian Renaissance master Titian (*c.* 1490–1576), who was to have more influence than any other artist on the development of his style. The full-length portraits of Philip IV (*c.* 1626; Prado) and his brother the Infante Don Carlos (*c.* 1626; Prado) are in the tradition of Spanish royal portraits established by Titian and are to some extent influenced by his style. In these portraits the detailed description and tenebrism of Velázquez' Sevillian paintings have been modified; only the faces and hands are accentuated and the dark figures stand out against a light background. In his later

Influence of Titian and Rubens

By courtesy of the Museo del Prado, Madrid



"Las Meninas" (with a self-portrait of the artist at the left and reflections of Philip IV and Queen Mariana in the mirror at the back of the room and the Infanta Margarita with her *meninas,* or maids of honour, in the foreground), oil on canvas by Diego Velázquez, *c.* 1656. In the Prado, Madrid, 3.18 × 2.76 m.

court portraits, Velázquez was to adopt something of the more elaborate decor and richer colouring of the Flemish Baroque master Peter Paul Rubens (1577–1640), whom he met during the latter's second visit to the Spanish court in 1628. Pacheco tells how Rubens praised Velázquez' works very highly because of their simplicity. Velázquez' painting of Bacchus, known as "Los Borrachos" ("The Topers" or "The Triumph of Bacchus") seems to have been inspired by Titian and Rubens, but his realistic approach to the subject is characteristically Spanish and one that Velázquez was to preserve throughout his life.

First Italian journey. Velázquez' visit, with Rubens, to see the famous paintings in the royal monastery of the Escorial near Madrid is said by Palomino to have aroused his desire to go to Italy. Having obtained leave and two years' salary from the King and money and letters of recommendation from Olivares, he sailed from Barcelona to Genoa in August 1629. In letters from Italian ambassadors in Madrid he is referred to as a young portrait painter, favourite of the King and Olivares, who was going to Italy to study and to improve his painting. The visit did in fact have an important effect on his artistic evolution. He stopped in Venice, where Palomino says he made drawings after Tintoretto (1518–94), the master of late 16th-century Venetian painting, and then hurried on to Rome. Pacheco relates that he was given rooms in the Vatican palace, which he found very isolated. Having obtained permission to return to the Vatican to make drawings after Michelangelo's "Last Judgment" and the paintings of Raphael, he moved to the Villa Medici, which was "high and airy" and had "antique sculptures to copy." An attack of fever obliged him later to move nearer to the Spanish ambassador. After a year in Rome he returned to Spain, stopping on the way in Naples; he arrived back in Madrid early in 1631.

None of Velázquez' Italian drawings appear to have survived. Of the few paintings that he made in Italy, a "famous portrait of himself" painted in Rome, mentioned by Pacheco, is possibly the "self-portrait" known only in replicas (Uffizi, Florence; Museo Provincial de Bellas Artes de San Carlos, Valencia). The chief works of his Italian visit are the two "celebrated pictures" painted in Rome, which Palomino records he took back to Spain and offered to the King: "Joseph's Bloody Coat Brought to Jacob" and "The Forge of Vulcan." These two monumental figure compositions are far removed from the limited realism in which he had been trained. As a result of his Italian studies, particularly of Venetian painting, his development in the treatment of space, perspective, light, and colour and his broader technique mark the beginning of a new phase in his lifelong pursuit of the truthful rendering of visual appearance.

Middle years. After his return from Italy, Velázquez entered upon the most productive period of his career. He took up again his chief office of portrait painter and was occasionally called on to represent mythological subjects for the decoration of the royal apartments. From now on his religious works are rare and individual. The devotional quality of his early Sevillian paintings finds moving expression in the "Christ on the Cross," a composition of monumental simplicity and naturalness. In "The Coronation of the Virgin" the solemnity and dignity of the holy persons are set off by their voluminous, colourful robes in a composition of exceptional splendour specially fitting for a painting of the Queen of Heaven made to adorn the oratory of the Queen of Spain.

Equestrian portraits

For the decoration of the throne room of the new Buen Retiro palace, completed in 1635, Velázquez painted a series of royal equestrian portraits (Prado), following a tradition that goes back in Spain to Titian's portrait of "Charles V at Mühlberg" (1548; Prado) and was continued by Rubens. Velázquez' equestrian groups have a balance and poise closer to Titian's than to Rubens' Baroque compositions, and after his return from Italy, he achieved a three-dimensional effect without detailed drawing or strong contrasts of light and shade but with a broad technique of brushwork and natural outdoor lighting. "The Surrender of Breda," Velázquez' famous contribution to the series of military triumphs painted for the same throne room, is his only surviving historical subject. Though the elaborate composition was based on a pictorial formula of Rubens, he creates a vivid impression of actuality and of human drama by means of accurate topographical details and the lifelike portraiture of the principal figures.

Though Velázquez frequently followed traditional compositions, particularly for his royal portraits, it was from no lack of ability to compose or invent. With his portraits of Philip IV (*c.* 1635; Prado), the Infante Fernando (*c.* 1632–35; Prado), and Prince Baltasar Carlos as huntsman, painted for the King's new hunting lodge, the Torre de la Parada, he created a new type of informal royal portrait. For the same place he painted hunting scenes of which "Philip IV Hunting Wild Boar" (National Gallery, London) is possibly an example, and some classical subjects, including probably the portraitlike figures of "Aesop" and "Menippus" (1639–40; Prado). The portraits of court dwarfs (Prado), painted during the next few years, display the same impartial and discerning eye as those of royal and noble sitters, while the character of the dwarfs' deformities is revealed through their awkward, unconventional poses, their individual expressions, and by the exceptionally free and bold brushwork. The "Lady with a Fan," one of the few informal portraits of women, is, on the other hand, remarkable for the subtle and delicate painting and for the sensitive portrayal of personal charm.

Second Italian journey. At the beginning of 1649 Velázquez left Spain on a second visit to Italy. This time he was on official business as gentleman of the bedchamber. He was given a carriage for pictures, possibly gifts from Philip IV to Pope Innocent X. The chief purpose of the journey was to buy paintings and antiques for the King for the decoration of new apartments in the royal palace and also to engage fresco painters to decorate the ceilings of the apartments and to reintroduce fresco painting into Spain. Again Velázquez found fresh inspiration in Italy, particularly from Titian. First he went to Venice, where he bought paintings by Titian, Tintoretto, and Veronese. He then went on to Modena, where he saw the famous ducal collection, which included his own portrait of the Duke of Modena, painted in Madrid in 1638. According to Palomino, he stopped in many other cities, including Bologna, where he contracted fresco painters to work in Madrid. Palomino recounts that Velázquez was befriended in Rome by eminent prelates and artists, including the French painter Nicolas Poussin (1594–1665) and Gian Lorenzo Bernini (1598–1680), the leading Italian sculptor of the Baroque style. He gives a list of the antiques selected by Velázquez, from which it seems that he followed the tradition of great collectors since the 16th century: rather than inferior originals he chose casts of the most famous statues in Rome. "Without neglecting his other business he also did many paintings," in addition to the portrait of Innocent X. Palomino relates that before portraying the Pope, as an exercise in painting a head from life, Velázquez made the portrait of his mulatto slave, Juan de Pareja (freed by Velázquez in 1650). This is an exceptional unofficial portrait, unusually boldly painted, which creates a powerful effect of familiar and living likeness. In 1970 the sum of $5,544,000 was paid for this picture—at the time the highest price paid for a work of art at auction.

The portrait of Innocent X

For the portrait of Innocent X, one of his most important official works, Velázquez followed a tradition for papal portraits created by Raphael in the likeness of "Julius II" (*c.* 1511–12; National Gallery, London) and later used by Titian in portraying "Paul III and His Grandsons Ottavio and Cardinal Alessandro Farnese" (1546; Museo e Gallerie Nazionali di Capodimonte, Naples). The powerful head, the brilliant combinations of crimson of the curtain, the chair, and the cope are painted with fluent technique and almost imperceptible brush strokes that go far beyond the late manner of Titian and announce the last stage in Velázquez' development in the direction of Impressionism. This portrait, which has long been Velázquez' most famous painting outside Spain, was copied innumerable times and won him immediate and lasting renown in Italy. In 1650 he was made a member of the Accademia di San Luca (Academy of St. Luke) and of the Congregazione dei Virtuosi al Pantheon, Rome's two most prestigious orga-

nizations of artists. The portrait earned for him the Pope's support for his application for membership of the most exclusive Spanish military order, though the difficulties arising from the fact that he was not of noble birth were so great that he did not receive the habit of the Order of Santiago until 1659.

The two small views of the Villa Medici, where Velázquez stayed during his first visit to Rome, must, for stylistic reasons, have been painted during his second visit. They are unique examples of pure landscape in his surviving work and among those of his achievements that foreshadow 19th-century Impressionism. The so-called "Rokeby Venus" was also probably painted in Italy and is one of the few representations of the female nude in Spanish painting before the 19th century. The theme of the toilet of Venus, the rich colouring and warm flesh tones, are inspired mainly by Titian and other Venetian painters. But Velázquez has characteristically made no attempt to disguise or idealize his model, and his superbly painted Venus is exceptional for his time as a lifelike portrayal of a living nude woman.

Last years. Velázquez returned to Madrid in the summer of 1651 with some of his purchases and was warmly welcomed by the King who, in the following year, appointed him chamberlain of the palace, an office that entailed the arrangement of the royal apartments and of the King's journeys. During his absence Philip had remarried, and the young queen Mariana of Austria with her children provided new subjects for him to portray. For his portraits of the Queen (1652–53; Prado) and of the King's oldest daughter, the Infanta María Teresa (1652/53; Kunsthistorisches Museum, Vienna), he used similar compositional formulas, and numerous studio replicas of them were made. The royal ladies appear as doll-like figures with their enormous coiffures and farthingale hoops. The effect of form, texture, and ornament is achieved in Velázquez' late manner without any definition of detail, in a free, "sketchy" technique. The portraits of the young infanta Margarita (1659; Kunsthistorisches Museum, Vienna) and Prince Felipe Próspero, similar in composition and manner, are among the most colourful of his works, and he most sensitively reveals the childlike character of his sitters behind the facade of royal dignity. Velázquez' late bust portraits of Philip IV (*c.* 1654, Prado; and *c.* 1656, National Gallery, London), of which many studio versions exist, are very different in character and are exceptional as royal portraits for their informal appearance. These last close-up views of the sad and aging monarch are among the most intimate of all Velázquez' royal characterizations.

Late portraits

In addition to his many official portraits, Velázquez painted during his last years two of his most original figure compositions and greatest masterpieces. "Las Hilanderas," a genre scene in a tapestry factory, is at the same time an illustration of the ancient Greek fable of the spinning contest between Pallas Athena and Arachne. Here, the mythological subject—like the religious scene in some of the early *bodegones*—is in the background. But in this late work there is no barrier between the world of myth and reality; they are united in an ingenious composition by formal and aerial perspective. In "Las Meninas" ("The Maids of Honour"), also known as "The Royal Family," he has created the effect of a momentary glance at a casual scene in the artist's studio while he is painting the King and Queen—whose reflection only is seen in the mirror in the background—in the presence of the Infanta Margarita with her *meninas* and other attendants. In this complex composition, the nearly life-size figures are painted in more or less detail according to their relation to the central figure of the Infanta and to the source of light, creating a remarkable illusion of reality never surpassed by Velázquez or any other artist of his age.

Velázquez' last activity was to accompany the King and court to the French border, in the spring of 1660, to arrange the decoration of the Spanish pavilion for the marriage of the Infanta María Teresa with Louis XIV. Shortly after his return to Madrid, he fell ill and died on August 6, 1660. Velázquez left few pupils or immediate followers. His European fame dates from the beginning of the 19th century. Many of his early Sevillian paintings were acquired then by foreign (chiefly English) collectors. Most of his later official works were incorporated in the Prado Museum, in Madrid.

MAJOR WORKS

"Old Woman Frying Eggs" (1618; National Gallery of Scotland, Edinburgh); "Christ in the House of Martha and Mary" (*c.* 1618; National Gallery, London); "Water Seller of Seville" (*c.* 1619; Wellington Museum, London); "Adoration of the Magi" (1619; Prado); "Los Borrachos" ("The Topers") or "The Triumph of Bacchus" (*c.* 1628–29; Prado); "The Forge of Vulcan" (1630; Prado); "Joseph's Bloody Coat Brought to Jacob" (1630; El Escorial); "Christ on the Cross" (*c.* 1631; Prado); "Equestrian Portrait of Philip IV" (*c.* 1634–35; Prado); "Equestrian Portrait of Prince Baltasar Carlos" (*c.* 1634–35; Prado); "Equestrian Portrait of Olivares" (*c.* 1634–35; Prado); "The Surrender of Breda" (1634–35; Prado); "Prince Baltasar Carlos as Huntsman" (1635–36; Prado); "Duke of Modena" (1638; Galleria e Museo Estense, Modena); "Lady with a Fan" (*c.* 1635–40; Wallace Collection, London); "The Coronation of the Virgin" (*c.* 1641–42; Prado); "Philip IV at Fraga" (1644; Frick Collection, New York); "Mars" (*c.* 1645–48; Prado); "Portrait of Juan de Pareja" (1649–50; Metropolitan Museum of Art, New York); "Portrait of Pope Innocent X" (1650; Galleria Doria-Pamphili, Rome); "Views of the Villa Medici, Rome" (*c.* 1650; Prado); "The Toilet of Venus," known as "The Rokeby Venus" (*c.* 1651; National Gallery, London); "Queen Mariana of Austria" (1652–53; Prado); "The Infanta Margarita in a Pink Dress" (1653–54; Kunsthistorisches Museum, Vienna); "Las Hilanderas" ("The Spinners") or "The Fable of Arachne" (*c.* 1655; Prado); "Las Meninas" ("The Maids of Honour"; 1656; Prado); "Mercury and Argus" (*c.* 1659; Prado); "Prince Felipe Próspero" (1659; Kunsthistorisches Museum, Vienna).

BIBLIOGRAPHY. CARL JUSTI, *Diego Velazquez und sein Jahrhundert,* 2 vol. (1888; 2nd rev. ed, 1903; Eng. trans., *Diego Velazquez and His Times,* 1889; Spanish ed., 1953), a classic monograph, though out-of-date, still the most important and comprehensive study of the subject; MADLYN M. KAHR, *Velázquez: The Art of Painting* (1976), for the general reader; MAURICE SÉRULLAZ, *Velázquez* (1981), an overview; AUGUST L. MAYER, *Velázquez: A Catalogue Raisonné of the Pictures and Drawings* (1936), a well-illustrated catalog of works by and related to Velázquez; ENRIQUE LAFUENTE, *Velazquez: Complete Edition* (1943), an introductory essay followed by a critical catalog, fully illustrated, of all the paintings and drawings considered by the author to be by Velázquez; ELIZABETH DU G. TRAPIER, *Velázquez* (1948), a well-documented study of the artist's life and work that includes most of the information brought to light since the publication of Justi's book, with a comprehensive bibliography, *Varia velazqueña: homenaje a Velázquez en el III centenario de su muerte, 1660–1960,* 2 vol. (1960), and *Velázquez: homenaje en el tercer centenario de su muerte* (1960), two volumes containing transcriptions of all known source material, biographical and documentary, relating to Velázquez; JOSÉ LÓPEZ-REY, *Velázquez: A Catalogue Raisonné of His Oeuvre, with an Introductory Study* (1963), a revised and enlarged edition of Mayer's *Catalogue,* amended in the light of new discoveries and of the author's own views on attribution and dating—a comprehensive and fully illustrated corpus of the artist's work; JOSÉ ORTEGA Y GASSET, *Velazquez, Goya and the Dehumanization of Art* (1972), essays on art, including discussion of Velázquez; JUAN A. GAYA NUÑO, *Bibliografía crítica y antológica de Velázquez* (1963, reissued 1973), a comprehensive critical bibliography of Velázquez, with analytical indexes, a work of major importance for the study of the artist.

(E.Ha./Ed.)

Venezuela

Venezuela (Republic of Venezuela; Spanish: República de Venezuela), located at the northern extremity of the South American continent, occupies a roughly triangular area of 352,144 square miles (912,050 square kilometres), which is about one and a half times the size of France. It is bounded by the Caribbean Sea and the Atlantic Ocean to the north, Guyana to the east, Brazil to the south, and Colombia to the southwest and west. The national capital, Caracas, is the primary centre of industry, commerce, education, and tourism. Reflecting the nation's rapid urbanization, the Caracas metropolitan area population increased sixfold between 1950 and 1980. Venezuela has jurisdiction over a number of Caribbean islands, among which are Margarita Island, the largest, Los Roques, Tortuga, and La Blanquilla. There is a long-standing dispute with Guyana, dating to 1840, over territorial jurisdiction of that country's area west of the Essequibo River, of which Venezuela claims some 53,000 square miles. A dispute with another neighbour, Colombia, over the precise demarcation of maritime boundaries has also endured at length.

A physically diverse country, Venezuela has landscapes that are a complex product of the interaction of culture and nature. The republic's development pattern has been unique among Latin-American countries in terms of the speed, sequence, and timing of economic and demographic growth. The 20th century witnessed the transformation of a relatively poor agrarian society to a rapidly urbanizing one, a condition brought about by the wealth and prosperity derived from the exploitation of the country's huge petroleum reserves. These changes, however, have been accompanied by problems of economic imbalance and regional inequalities, with often acute political debate over the meaning and direction of local and national development. The government must deal with a massive and largely uncontrolled rural-to-urban migration, as well as mass immigration, much of it illegal, from Colombia and other neighbours. Once primarily a food exporter, Venezuela has become heavily dependent upon imported food and commodities to satisfy the domestic material demands and life-styles of its urban population. Years of petrodollar abundance were accompanied by undisciplined borrowing from overseas financial institutions, leaving the country with a growing debt-repayment problem, a condition that was intensified by the oil depression of the late 1970s and '80s. Venezuela has also not been exempt from the type of governmental patronage and corruption commonly regarded as traditional in Latin-American countries. On the other hand the republic can point to a notable degree of political stability and basic democracy since the late 1950s.

This article is divided into the following sections:

Physical and human geography 468
- The land 468
  - Relief
  - Drainage and soils
  - Climate
  - Plant and animal life
  - Settlement patterns
- The people 473
  - Ethnic composition
  - Demographic trends
- The economy 473
  - Resources
  - Agriculture, fishing, and forestry
  - Industry
  - Finance and trade
  - Transportation
- Government and social conditions 475
  - Government
  - Education, health, and welfare
- Cultural life 476
  - Cultural traditions
  - The arts
  - Cultural institutions
  - Recreation
  - Press and broadcasting

History 476
- The colonial era 476
- The independence movement 477
- The caudillos (1830–1935) 477
  - Páez and the Conservatives
  - The Monagas and the civil wars
  - The reigns of Guzmán Blanco and Crespo
  - The Andinos
- Venezuela since 1935 478

Bibliography 479

Physical and human geography

Venezuelan geography presents a number of discernible contrasts of landscape, people, and culture. Towering mountains, tropical jungles, broad river plains, and arid coastal plains provide a diversity of natural habitats that present challenging opportunities for regional development and territorial integration. Sprawling metropolitan centres, densely populated mountain valleys, aboriginal log houses along riverbanks, and open, sparsely settled Llanos (plains) present a diversity of settlement patterns. Pre-Hispanic, colonial Hispanic, and modern social elements coexist in a rich variety of cultural identities. A stable network of towns and cities, whose interconnections define the directions of resource distribution, reflect social continuity within the nation.

THE LAND

Relief. Altitude and relief are the important discriminators of Venezuela's physical landscape, with seasonal climatic variation providing environmental diversity within broader physiographic domains. Three broad regional divisions can be delimited: the Cordillera, a mountain region (reaching heights of about 16,400 feet [5,000 metres] above sea level), the lowland plains (reaching to 1,600 feet), and the interior forested uplands (reaching to 8,200 feet). Within these three broad divisions, seven physiographic regions can be distinguished: the Andes Mountains; the Lake Maracaibo Lowlands; the islands and coastal plains (including the Orinoco delta); the coastal mountain system; the valleys and hill ranges of northwestern Venezuela, also called the Segovia Highlands; the Llanos; and the Guiana Highlands.

Physiographic regions

The Andes Mountains, including the highest peaks in Venezuela, extend northward into Venezuela in two branches, between which lie the Lake Maracaibo Lowlands. The western branch, the Sierra de Perijá, extending along the border with Colombia, is an extension of the Colombian Andes' Cordillera Oriental. The eastern branch, the Cordillera de Mérida, extends from the Tachira depression, which separates it from the Cordillera Oriental, to the Barquisimeto depression. Physiographically, the Segovia Highlands, northwest of Barquisimeto, and the coastal ranges, which in this treatment are regarded separately, may also be considered a part of the Andes chain. The highest point of the Venezuelan Andes is Pico Bolívar (16,427 feet) in the Mérida Mountains; a considerable part of the mountain region has temperate

(*templado*) conditions, but the cold (*frio*) zone is much smaller than in other Andean countries.

The high Andean ranges overlook the lowlands of Lake Maracaibo, much of which are filled by the lake itself—actually a large shallow inlet, or embayment, of the sea. The main oil-producing region of the country, it is covered by dense tropical rain forest in the south, but closer to the Caribbean there is xerophytic thorn scrubland.

Islands and coastal plains

The islands and coastal plains are located to the north and northeast of Venezuela. They include the Caribbean "Islands to the Leeward," such as Margarita and Tortuga, and several peninsulas, including the head-shaped Paraguaná to the west and, to the east, Araya and Paria, the latter a finger of land pointing at Trinidad. The coastal plains extend from the Colombian border and the Gulf of Venezuela eastward to the foothills of the Coastal Range; the Unare Basin, the country's largest inland embayment, spreads out east of the range. Beyond Unare, in the east, the Orinoco delta opens onto the Atlantic Ocean through a number of distributories (*canos*); an early gateway to the interior, it is a low, dank, and swampy area heavily dissected by streams.

The coastal mountain system, in effect two parallel ranges—the Coastal Range and the Interior Range—contains Venezuela's greatest concentration of population, although covering only 3 percent of the national territory. In the intermontane valleys are the major cities of Caracas, Valencia, and Maracay, and all but the steepest slopes are populated. The highest point is Naiguatá peak (9,069 feet).

The valleys and hill ranges of the northwest form a transitional upland zone between the Coastal and Andean mountains. Elevations there range from 1,600 to 5,500 feet. The only desert in Venezuela—the city of Coro's sand dunes—is found in this region.

Between the Mérida Range and the Orinoco River lie the Llanos, a relatively level savanna region, where the only undulations of relief are between low mesalike interfluves and shallow, meandering, braided river courses. Cattle grazing predominates in this sparsely populated region, which experiences river flooding in summer and drought in winter as seasonal extremes. From the Andean foothills to the Orinoco delta, these Llanos extend for some 800 miles, varying in width from 100 miles in the east to 300 miles in the west.

The Guiana Highlands

From the Orinoco through the southernmost (Amazonas) territory bordering Colombia, Brazil, and Guyana is the Guiana Highlands, or Guayana, largely an upland surface of rounded hills and narrow valleys formed from ancient crystalline rocks. Occupying almost 50 percent of the country, it is the most remote and least explored area of Venezuela. Along the southern border with Brazil are groups of massive plateaus and mesas, known as *tepuis*, capped with resistant sandstone, which are remnants of the former erosional surfaces of the original continental landmass. These *tepuis* are covered with intermingled savanna and semideciduous forest, and like the lowland savannas of the Llanos, they experience extreme rainy and dry seasons. In the southeastern Guiana Highlands, in the region called La Gran Sabana (a high plain), are the Angel Falls, the highest waterfall in the world (3,212 feet). The highlands are sparsely settled. An estimated 30,000 native Indians live in the tropical forests of the Amazonas territory bordering Brazil and Colombia. Other settlements are clustered around mission stations along the rivers. The highlands, however, have tremendous resource potential; they abound in deposits of iron ore, gold, and diamonds, and they possess considerable hydroelectric potential, as well as hardwood forest resources. The long-standing conflict with Guyana over territorial jurisdiction of bordering areas, as well as the tensions over Brazilian and Colombian settlements and illegal crossings of people, cattle, and contraband on the Amazonas borders, have made the highlands a region of increasing strategic importance.

Drainage and soils. The Venezuelan drainage network consists almost entirely of two watersheds, the largest emptying into the Atlantic Ocean via the Orinoco River system and the other, into the Caribbean Sea.

The Orinoco River basin

The Orinoco River and its main tributary, the Caroní, drain a 366,000-square-mile basin; some 80 percent of the country's drainage flows into this interior basin. The Orinoco's source is in the southern Guiana Highlands; it first flows northwestward, then north, and finally eastward to its delta, spreading to the Atlantic Ocean across a 265-mile-wide seafront. In the Orinoco's middle course, where it flows eastward through the wide Llanos, it is joined by tributaries from the Llanos interior, such as the Apure and Meta, and by other tributaries originating in the Guiana Highlands to the south, such as the Caroní. One unusual configuration occurs near the river's source, where, because of the almost level gradient, the Orinoco channel divides, one branch discharging southwest into the Casiquiare River, which joins the Negro, a tributary of the Amazon, the other branch continuing its northern flow through Venezuela.

The intermontane basins and valleys of the Andes and Coastal Range are drained mainly by other tributaries of the Orinoco. The Caracas valley is an exception, however; there the Tuy River runs eastward to the Caribbean. Landlocked Lake Valencia is Venezuela's only example of interior drainage, but it is steadily shrinking because of sedimentation and evaporation.

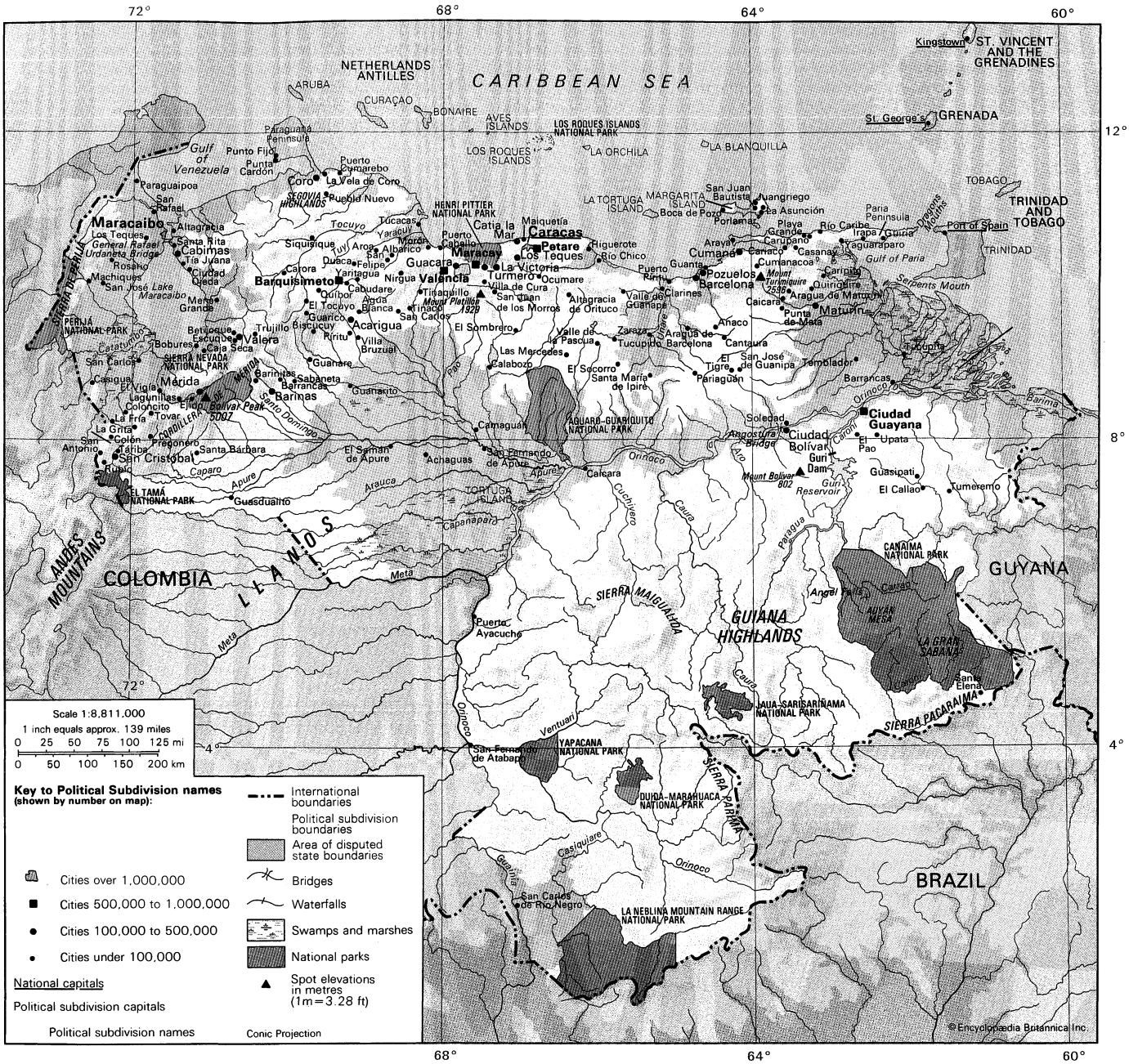
Lake Maracaibo is about 100 miles long from north to south and 75 miles in width at its widest. Short, fast-moving rivers run off the surrounding mountain slopes, depositing silt into the lake's shallow basin. A narrow strait connects Lake Maracaibo to the Caribbean, so that at its surface it contains fresh water in the south and brackish water in the north.

Relatively infertile reddish latosols are Venezuela's most common type of soil. Tropical heat and abundant moisture characteristically leach latosols of all but the most insoluble minerals, iron, and aluminum sesquioxides. These minerals are collectively known as laterite, and they give the soil its name and its distinctive red colour. Latosols are the characteristic soil of the Llanos and the Guiana Highlands except where alluvium is abundant. The country's most fertile soils are formed by such well-drained, transported material as river alluvium or recent volcanic deposits. Alluvial soils are found in the southern Maracaibo Lowlands, along the fringes of the Llanos where rivers draining the surrounding highlands have laid thick deposits of alluvium, and in the broad valley bottoms in the northern mountains. The Orinoco delta and adjacent plains are also rich in alluvium, though poor drainage is a problem with soils developed in these low areas. Volcanic soils cover the slopes of many of the northern mountains, but, though inherently fertile, these soils often suffer severe erosion because of deforestation and shifting agricultural practices.

Climate. Venezuela lies well within the tropics, and, despite regional variation in relief, precipitation, and vegetation, temperatures are relatively uniform across the country and vary little throughout the year. More than 90 percent of Venezuela has a mean annual temperature above 75° F (24° C), and altitude is responsible for the appreciable differences that do occur. The average mean temperature for Caracas, lying in a high valley, is, for example, about 72° F (22° C), while its nearby port of La Guaira averages about 81° F (27° C). Mérida, at more than 4,900 feet, averages 66° F (19° C), while low-lying Maracaibo, at sea level, averages 82° F (28° C). Day-to-night variations of temperature are more pronounced than month-to-month variations, a characteristic trait of the tropics.

Mean temperatures

Venezuela's climatic year is divided into two seasons; the wet season lasts from May to October and even continues sporadically through November, and the dry season begins in December and continues until the end of March. Regional variations in precipitation, however, are marked. Although lying in the northeast trade winds zone, only the northeastern coastal areas receive appreciable precipitation in the summer. The northwestern coastal regions experience descending air, and the resultant aridity can be severe, some places receiving less than 20 inches (500 millimetres) of rainfall annually. La Guaira, for example, receives an average of only 11 inches. Rain shadow areas behind coastal and upland ranges also receive little rain, while their corresponding windward slopes are generally



MAP INDEX

Political subdivisions

Amazonas ..... 3 30 N 66 00 W
Anzoátegui ..... 9 00 N 64 30 W
Apure ..... 7 10 N 68 50 W
Aragua ..... 10 00 N 67 10 W
Barinas ..... 8 10 N 69 50 W
Bolívar ..... 6 20 N 63 30 W
Carabobo ..... 10 10 N 68 05 W
Cojedes ..... 9 20 N 68 20 W
Delta Amacuro ... 8 30 N 61 30 W
Distrito Federal .. 10 30 N 66 55 W
Falcón ..... 11 00 N 69 50 W
Guárico ..... 8 40 N 66 35 W
Lara ..... 10 10 N 69 50 W
Mérida ..... 8 30 N 71 10 W
Miranda ..... 10 15 N 66 25 W
Monagas ..... 9 20 N 63 00 W
Nueva Esparta ..... 11 00 N 64 00 W
Portuguesa ..... 9 10 N 69 15 W
Sucre ..... 10 25 N 63 30 W
Táchira ..... 7 50 N 72 05 W
Trujillo ..... 9 25 N 70 30 W
Yaracuy ..... 10 20 N 68 45 W
Zulia ..... 10 00 N 72 10 W

Cities and towns

Acarigua ..... 9 33 N 69 12 W
Achaguas ..... 7 46 N 68 14 W
Agua Blanca ..... 9 40 N 69 06 W
Albarico ..... 10 24 N 68 42 W
Altagracia ..... 10 07 N 71 14 W
Altagracia de Orituco ..... 9 52 N 66 23 W
Anaco ..... 9 27 N 64 28 W
Aragua de Barcelona ..... 9 28 N 64 49 W
Aragua de Maturín ..... 9 58 N 63 29 W
Araya ..... 10 34 N 64 15 W
Aroa ..... 10 26 N 68 54 W
Barcelona ..... 10 08 N 64 42 W
Barinas ..... 8 38 N 70 12 W
Barinitas ..... 8 45 N 70 25 W
Barquisimeto .... 10 04 N 69 19 W
Barrancas ..... 8 46 N 70 06 W
Barrancas ..... 8 42 N 62 11 W
Betijoque ..... 9 23 N 70 44 W
Biscucuy ..... 9 22 N 69 59 W
Bobures ..... 9 15 N 71 11 W
Boca de Pozo ... 11 00 N 64 23 W
Cabimas ..... 10 23 N 71 28 W
Cabudare ..... 10 02 N 69 16 W
Caicara (Caicara de Orinoco) ..... 7 37 N 66 10 W
Caicara ..... 9 49 N 63 36 W
Caja Seca ..... 9 09 N 71 05 W
Calabozo ..... 8 56 N 67 26 W
Camaguán ..... 8 06 N 67 36 W
Cantaura ..... 9 19 N 64 21 W
Caracas ..... 10 30 N 66 55 W
Cariaco ..... 10 29 N 63 33 W
Caripito ..... 10 08 N 63 06 W
Carora ..... 10 11 N 70 05 W
Carúpano ..... 10 40 N 63 14 W
Casanay ..... 10 30 N 63 25 W
Casigua ..... 8 46 N 72 30 W
Catia la Mar .... 10 36 N 67 02 W
Ciudad Bolívar ... 8 08 N 63 33 W
Ciudad Guayana (San Félix) ..... 8 23 N 62 40 W
Ciudad Ojeda ... 10 12 N 71 19 W
Clarines ..... 9 56 N 65 10 W
Colón (San Juan de Colón) ..... 8 02 N 72 16 W
Coloncito ..... 8 20 N 72 05 W
Coro ..... 11 25 N 69 41 W
Cumaná ..... 10 28 N 64 10 W
Cumanacoa .... 10 15 N 63 55 W
Duaca ..... 10 18 N 69 10 W
Ejido ..... 8 33 N 71 14 W
El Callao ..... 7 21 N 61 49 W
El Pao ..... 8 01 N 62 38 W
El Samán de Apure ..... 7 55 N 68 44 W
El Socorro ..... 8 59 N 65 44 W
El Sombrero ..... 9 23 N 67 03 W
El Tigre ..... 8 55 N 64 15 W
El Tigrito, see San José de Guanipa
El Tocuyo ..... 9 47 N 69 48 W
El Vigía ..... 8 38 N 71 39 W
Escuque ..... 9 18 N 70 40 W
Guacara ..... 10 14 N 67 53 W
Guanare ..... 9 03 N 69 45 W
Guanarito ..... 8 42 N 69 12 W
Guanta ..... 10 14 N 64 36 W
Guarico ..... 9 32 N 69 48 W
Guasdualito ..... 7 15 N 70 44 W
Guasipati ..... 7 28 N 61 54 W
Güiria ..... 10 34 N 62 18 W
Higuerote ..... 10 29 N 66 06 W
Irapa ..... 10 34 N 62 35 W
Juangriego ..... 11 05 N 63 57 W
La Asunción .... 11 02 N 63 53 W
La Fría ..... 8 13 N 72 15 W

La Grita ..... 8 08 N 71 59 W
La Vela de Coro . 11 27 N 69 34 W
La Victoria ..... 10 14 N 67 20 W
Lagunillas ..... 8 31 N 71 24 W
Las Mercedes .... 9 07 N 66 24 W
Los Teques ..... 10 21 N 67 02 W
Los Teques ..... 10 37 N 71 50 W
Machiques ..... 10 04 N 72 34 W
Maiquetía ..... 10 36 N 66 57 W
Maracaibo ..... 10 40 N 71 37 W
Maracay ..... 10 15 N 67 36 W
Maturín ..... 9 45 N 63 11 W
Mene Grande .... 9 49 N 70 56 W
Mérida ..... 8 36 N 71 08 W
Morón ..... 10 29 N 68 11 W
Nirgua ..... 10 09 N 68 34 W
Ocumare ..... 10 07 N 66 46 W
Paraguaipoa .... 11 21 N 71 57 W
Pariaguán ..... 8 51 N 64 43 W
Petare ..... 10 29 N 66 49 W
Píritu ..... 9 23 N 69 12 W
Playa Grande ... 10 40 N 63 18 W
Porlamar ..... 10 57 N 63 51 W
Pozuelos ..... 10 11 N 64 39 W
Pregonero ..... 8 01 N 71 46 W
Pueblo Nuevo ... 11 07 N 69 28 W
Puerto Ayacucho ..... 5 40 N 67 35 W
Puerto Cabello .. 10 28 N 68 01 W
Puerto Cumarebo .... 11 29 N 69 21 W
Puerto Píritu .... 10 04 N 65 03 W
Punta Cardón ... 11 38 N 70 14 W
Punta de Mata ... 9 43 N 63 38 W
Punto Fijo ..... 11 42 N 70 13 W
Quíbor ..... 9 56 N 69 37 W
Quiriquire ..... 9 59 N 63 13 W
Río Caribe ..... 10 42 N 63 07 W
Río Chico ..... 10 19 N 65 59 W
Rosario (Villa del Rosario) ..... 10 19 N 72 19 W
Rubio ..... 7 43 N 72 22 W
Sabaneta ..... 8 46 N 69 56 W
San Antonio (San Antonio del Táchira) .... 7 50 N 72 27 W
San Carlos ..... 9 40 N 68 36 W
San Carlos ..... 9 01 N 71 55 W
San Carlos de Río Negro ..... 1 55 N 67 04 W
San Cristóbal .... 7 46 N 72 14 W
San Felipe ..... 10 20 N 68 44 W
San Félix, see Ciudad Guayana
San Fernando de Apure ..... 7 54 N 67 28 W
San Fernando de Atabapo ..... 4 03 N 67 42 W
San José ..... 10 01 N 72 23 W
San José de Guanipa (El Tigrito) ..... 8 54 N 64 09 W
San Juan Bautista ..... 11 01 N 63 57 W
San Juan de Colón, see Colón
San Juan de los Morros ..... 9 55 N 67 21 W
San Rafael ..... 10 58 N 71 44 W
Santa Bárbara .... 7 47 N 71 10 W
Santa Elena ..... 4 37 N 61 08 W
Santa María de Ipire ..... 8 49 N 65 19 W
Santa Rita ..... 10 32 N 71 32 W
Siquisique ..... 10 34 N 69 42 W
Soledad ..... 8 10 N 63 34 W
Táriba ..... 7 49 N 72 13 W
Temblador ..... 8 59 N 62 44 W
Tía Juana ..... 10 16 N 71 22 W
Tinaco ..... 9 42 N 68 26 W
Tinaquillo ..... 9 55 N 68 18 W
Tovar ..... 8 20 N 71 46 W
Trujillo ..... 9 22 N 70 26 W
Tucacas ..... 10 48 N 68 19 W
Tucupido ..... 9 17 N 65 47 W
Tucupita ..... 9 04 N 62 03 W
Tumeremo ..... 7 18 N 61 30 W
Turmero ..... 10 14 N 67 29 W
Upata ..... 8 01 N 62 24 W
Valencia ..... 10 11 N 68 00 W
Valera ..... 9 19 N 70 37 W
Valle de Guanape ..... 9 54 N 65 41 W
Valle de la Pascua ..... 9 13 N 66 00 W
Villa Bruzual ..... 9 20 N 69 06 W
Villa de Cura .... 10 02 N 67 29 W
Villa del Rosario, see Rosario
Yaguaraparo .... 10 34 N 62 49 W
Yaritagua ..... 10 05 N 69 08 W
Zaraza ..... 9 21 N 65 19 W

Physical features and points of interest

Aguaro-Guariquito National Park ... 8 25 N 66 40 W
Andes Mountains . 6 30 N 72 00 W
Angel (Churún-Merú) Falls ..... 5 57 N 62 30 W
Angostura Bridge ..... 8 05 N 63 34 W
Apure, *river* ..... 7 37 N 66 25 W
Arauca, *river* ..... 7 24 N 66 35 W
Aro, *river* ..... 8 01 N 64 11 W
Auyán Mesa ..... 5 55 N 62 32 W
Aves Islands .... 11 58 N 67 33 W
Barima, *river* ..... 8 35 N 60 25 W
Bolívar, Mount .... 7 28 N 63 25 W
Bolívar Peak ..... 8 33 N 71 03 W
Canaima National Park ... 5 30 N 62 00 W
Capanaparo, *river* ..... 7 01 N 67 07 W
Caparo, *river* ..... 7 46 N 70 23 W
Caribbean Sea .. 12 30 N 66 00 W
Caroní, *river* ..... 8 21 N 62 43 W
Carrao ..... 6 17 N 62 51 W
Casiquiare, *river* .. 2 01 N 67 07 W
Catatumbo, *river* .. 9 21 N 71 45 W
Caura, *river* ..... 7 38 N 64 53 W
Churún, *river* ..... 6 04 N 62 36 W
Cuchivero, *river* ... 7 40 N 65 57 W
Churún-Merú, see Angel Falls
Dragons Mouths, *marine channel* . 10 45 N 61 46 W
Duida-Marahuaca National Park ... 3 25 N 65 40 W
El Tamá National Park ..... 7 27 N 72 14 W
General Rafael Urdaneta Bridge ..... 10 35 N 71 35 W
Guainía, *river* ..... 2 01 N 67 07 W
Guiana Highlands ..... 4 00 N 64 00 W
Guri Dam ..... 7 42 N 62 52 W
Guri Reservoir .... 7 31 N 62 50 W
Henri Pittier National Park .. 10 25 N 67 40 W
Jaua-Sarisariñama National Park ... 4 40 N 64 15 W
La Blanquilla, *island* ..... 11 51 N 64 37 W
La Gran Sabana, *plateau* ..... 5 30 N 61 30 W
La Neblina Mountain Range National Park ... 1 15 N 66 05 W
La Orchila, *island* ..... 11 48 N 66 10 W
La Tortuga Island ..... 10 56 N 65 20 W
Llanos, *plains* .... 5 00 N 70 00 W
Los Roques Islands ..... 11 50 N 66 45 W
Los Roques Islands National Park ..... 11 50 N 66 45 W
Maigualida, Sierra ..... 5 30 N 65 10 W
Maracaibo, Lake .. 9 40 N 71 30 W
Margarita Island ..... 11 00 N 64 00 W
Mérida, Cordillera de ... 9 00 N 71 00 W
Meta, *river* ..... 6 12 N 67 28 W
Orinoco, *river* .... 8 37 N 62 15 W
Pacaraima, Sierra ..... 4 05 N 61 30 W
Pao, *river* ..... 8 33 N 68 01 W
Paragua, *river* .... 6 55 N 62 55 W
Paraguaná Peninsula ..... 11 55 N 70 00 W
Paria, Gulf of .... 10 20 N 62 00 W
Paria Peninsula .. 10 40 N 62 30 W
Parima, Sierra .... 2 30 N 64 00 W
Perijá, Sierra de .. 10 00 N 73 00 W
Perijá National Park ..... 9 30 N 72 20 W
Platillón, Mount ... 9 54 N 67 31 W
Santo Domingo, *river* ..... 8 01 N 69 33 W
Segovia Highlands ..... 11 00 N 70 00 W
Serpents Mouth, *marine channel* ..... 10 10 N 61 58 W
Sierra Nevada National Park ... 8 35 N 70 45 W
Tocuyo, *river* .... 11 03 N 68 20 W
Tortuga Island .... 7 08 N 67 02 W
Turimiquire, Mount ..... 10 07 N 63 52 W
Tuy, *river* ..... 10 41 N 69 08 W
Unare, *river* ..... 10 06 N 65 12 W
Venezuela, Gulf of ..... 11 30 N 71 00 W
Ventuari, *river* .... 3 58 N 67 02 W
Yapacana National Park .... 3 45 N 66 55 W
Yaracuy, *river* ... 10 35 N 68 14 W

well watered. Inland, the Llanos and the southern interior of the country generally receive sufficient rainfall to support tropical savanna, lush tropical rain forest (selva), and cultivated cropland and grassland pastures.

Plant and animal life. Most of Venezuela's vegetation is tropical and nondeciduous or semideciduous, retaining foliage throughout the year or shedding it minimally. Altitude and rainfall variations determine differences in vegetation. Some two-fifths of the country is covered with forests and about half of it with savanna grasses; less than 5 percent of the land, mostly in the valleys of the Andes and of the coastal ranges, is under permanent cultivation.

Altitudinal variations in vegetation

Altitudinal contrasts abound. True tropical growths do not generally flourish above 1,500 feet, although the selva intermingles with tall savanna grasslands in transitional zones in the interior; mangrove swamp is found in the Orinoco delta. The selva gives way to semitropical vegetation, reaching up to about 5,000 feet and characteristically supporting plant life that includes tree ferns and epiphytes, such as orchids. Higher up the Andean slopes, this fern forest gives way to a transitional zone of mountain vegetation, and above 10,000 feet is the *páramo* vegetation, which has few trees but a variety of small alpine shrubs and lichens. The drier northwestern coastal area, which was formerly covered with dry scrub woodland and grasses, has suffered from clearing, soil denudation, and overcultivation. Except in the remoter interior, indigenous and introduced species coexist on the forested slopes and around the settled lowland plains and valleys.

The fauna of Venezuela is essentially an assemblage of South American species found throughout the northern tropical forest and savanna areas of the continent. The indigenous variety is considerable. Seven species of the cat family inhabit the forested interior, including the jaguar, ocelot, jaguarundi, puma, and margay. Several monkey species live in forested territories, among them the howler and spider monkey, the long-tailed capuchin, and the nocturnal durukuli. Other forest animals include two species of bear, two of peccary, two of deer, three of opossum, wild dog, and South American rodents, such as the agouti and the skunk. Among the more unusual breeds are the tapir, a large, prominently snouted, cloven-hoofed quadruped, and the aquatic, herbivorous manatee. Distinguished by its two flippers and spoon-shaped tail, the manatee is an inhabitant of coastal estuaries.

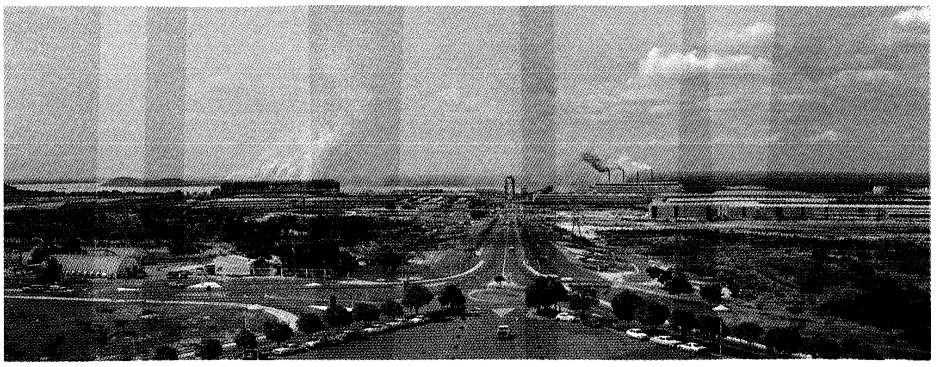
Reptilian life

Inhabiting the remoter rivers, coastal lagoons, and swamps are several reptilians, including caimans, alligators, lizards, and several species of turtle. Many species of snakes, too, inhabit the forested interior; these include such venomous ones as coral snakes, striped rattlesnakes, and bushmasters, as well as such nonvenomous varieties as the boa constrictor and anaconda.

Birds, both migratory and nonmigratory, are plentiful and diverse. The coastal swamps are the tropical venue for migratory cranes, herons, storks, and ducks. The ibis colonizes the Orinoco mangrove delta. Birds of prey are plentiful, thriving on the multitudes of migratory birds. The bellbird is prevalent in the Orinoco forests.

Pelagic and coral fish are plentiful off the Caribbean coast and the offshore waters of the Orinoco River, and the deltaic channels foster mollusks, young saltwater species, and shrimp. Swarming freshwater species in the interior rivers include electric eels and piranha. A wide array of catfish is caught for food.

Settlement patterns. *Regional disparities.* Agricultural development provided the stimulus for settlement and in-



Steel mill at Matanzas, part of the major industrial centre of Ciudad Guayana at the Orinoco–Caroní river junction.
Yoram Lehmann—Peter Arnold, Inc.

tegration of a national urban network during the Spanish colonial period, the towns established serving as market and distribution centres; occasional short-lived finds of gold in the mountainous interior also prompted some boomtown development. Rural populations were, however, always small and dispersed because of the limited amount of arable and pasture land. Insect-borne diseases severely hindered settlement in the Orinoco region and in other low-lying river basins; traditionally population densities were greater in the mountain valleys, where the climate is moderated by elevation and where malaria and other health hazards were less endemic. In particular, Andean towns served as administrative centres that fostered agricultural and commercial enterprise, cultural contact and change, and the centralization of authority and privilege; they prospered and grew on the profits of hides, cacao, and indigo.

Rural population dispersement

A century of independence beginning in 1830 consolidated the importance of the coastal ports, their hinterland administrative cities, and the growing dominance of Caracas as the hub of power, authority, and national wealth. The rural sector stagnated, while the northern urban network served as a conduit for the export of bulky raw materials, the importation of manufactured goods and foodstuffs, and the selective adoption of modern technologies in Caracas and La Guaira, including the telegraph and telephone, tramways, and railroads. Populations grew slowly in both rural and urban sectors, in part because of the prevalence of endemic diseases and the occurrence of natural disasters, and in part because of the economic stagnation that precluded mass immigration.

Urbanization. The uneven regional concentration of population has remained unchanged. With Caracas as its major focus, a major urbanized corridor consolidated in a relatively restricted zone of northern highland Venezuela during the period 1920–40. The oil boom in the Maracaibo Lowlands during the 1930s and '40s attracted some population from the Andean highlands, while elsewhere population concentrated in such flourishing agricultural and commercial centres as Barquisimeto, San Cristóbal, and Barcelona. Greater Caracas, however, was the primary focus of urban migration and immigration.

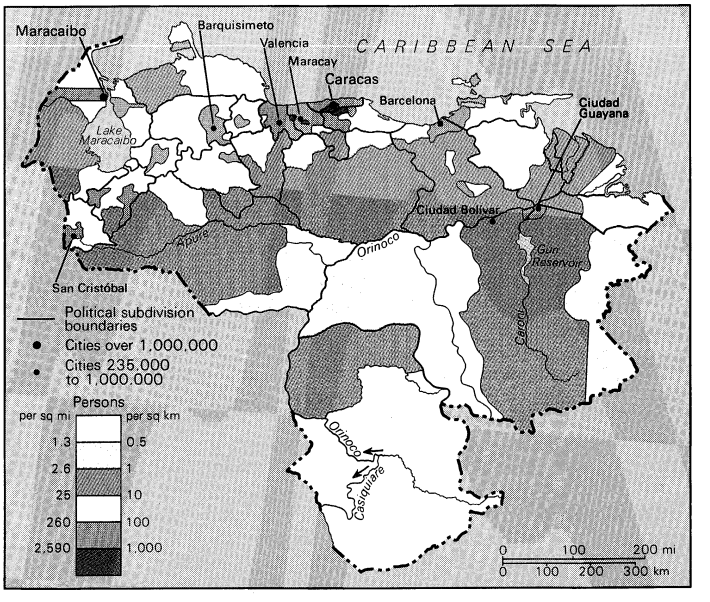
Major metropolitan areas

In modern Venezuela more than three-fourths of the people live in urban areas, with more than 40 percent residing in the four largest cities. A large and undocumented illegal immigration has contributed continuously to the urban growth rate. About one-sixth of the population lives in Greater Caracas, many of them recent migrants who inhabit the *ranchos,* or shantytowns; the second largest city is Maracaibo, followed by Valencia and Barquisimeto, all located in the north. In the eastern interior, Ciudad Guayana on the Orinoco has been growing rapidly.

The pace of urbanization in Venezuela between 1940 and 1970 was the most rapid in Latin America. In the 1940s this phenomenon caused mass rural depopulation. Rural outmigration was regionally concentrated. Whereas some rural communities, especially sparsely settled pioneer areas in the southern interior, experienced rapid growth after the eradication of many diseases, other rural communities in the densely settled highlands suffered sharp population declines. These latter regions became marginal productive sectors of agriculture, their stagnating communities offering few if any opportunities to the rural youth. Agricultural neglect was symptomatic of these early years of oil production. The apparent effects of material gains and prosperity in the cities and the persistence of inequalities in land ownership (large estates, *latifundios,* underutilized their lands, while small farms, *minifundios,* were overcultivated) demonstrated that agriculture was no longer regarded as the best vehicle for material advancement. Viewing farming and rural living as old-fashioned and devoid of modernity and opportunity, the youth and other elements of rural population chose migration to the city.

The doubling of the urban population in only nine years between 1940 and 1949 and continued rapid growth in the 1950s caused vast sums to be spent on public housing, although only a fraction of the demand for urban shelter could be met. The majority of the poor had to resort to self-help strategies to acquire shelter, and the rapid burgeoning of uncontrolled, illegal *ranchos* on the hillsides surrounding the large cities was viewed with alarm. These communities have come to be accepted, if not promoted, as inevitable consequences of the rapid urban growth. Many *ranchos* have undergone transformation to more permanent housing structures, with improvements in services and political representation, and with a general upgrading of neighbourhood facilities. Many *ranchos,* however, continue to suffer from high densities, low-quality housing, deficient services, crime, and malnutrition and disease. Rates of rural and urban migration from other parts of Venezuela to Caracas have declined, but illegal immigration from neighbouring Colombia, most of it city-bound, has aggravated the *rancho* problem.

Emergence of *ranchos*



Population density of Venezuela.

Decentralization. The spatial arrangement of population and production in Venezuela has changed profoundly since the 1950s. During that period modern road building began, extending the main Andean axis of population between San Cristóbal and Caracas westward to the Gulf of Maracaibo and eastward through Valencia to the Lower Orinoco. The Caracas–Valencia area, however, remained the population–production core, a position that was further entrenched by the building of *autopistas* (expressways) between Caracas and La Guaira and Caracas and Maracay. Paved highways have linked this northern axis to the interior and along the coast.

In the 1960s Venezuela embarked on an ambitious four-year plan, in which the main thrusts were economic diversification, decentralization, and planned urban growth. The administration of Rómulo Betancourt addressed the problems of regional inequalities and uneven development by promotion of a new urban growth pole on the southern bank of the Orinoco River at its confluence with the Caroní, a planned metropolitan region that would exploit the great natural resources there—hydroelectric power and iron ore—and counterbalance the Caracas–Valencia metropolitan core region. Centred on a new city, Ciudad Guayana, this government-controlled, integrated development region was incorporated into the national economic planning framework. Developed as a heavy industrial complex, it includes a huge government-owned steel mill, and the government undertook a project to connect the region with the north by rail. An oil refinery and petrochemical complex already have been established in this interior region.

Industrialization of the interior

National planning calls for the development of other new metropolitan regional complexes to further counter the centralizing tendencies of Greater Caracas. National parks, forest reserves, and the encouragement of Amazonian tourism provide balance to the emphases on industrial expansion and diversification. Progress has been made in the construction of a national electric power grid, and bridges and tunnels have been built to improve contact between regions. Airport development has helped to open up the interior, as well as to improve contacts with trade and business partners. Regional imbalances continue to exist, but there has been steady progress toward correcting them.

THE PEOPLE

Ethnic composition. Venezuela is a country of immigrants, with only 2 percent of its population being made up of indigenous Indian groups. The dominant ethnic type, sometimes called *pardo,* is of mixed African, European, and Indian ancestry; the *pardos* constitute 69 percent of the population, whites make up another 20 percent, and blacks 9 percent. Perhaps as many as one-fourth of the contemporary population are immigrants, many illegal.

The *pardos*

Prior to 1948, Venezuela had never openly encouraged non-Hispanic immigration, except for selective influxes of merchants, sailors, and entrepreneurs from neighbouring West Indian islands. In the late 1940s, however, stimulated by the development of a petroleum economy, a pro-immigration policy was adopted by the government. During a 10-year open immigration period, Venezuela recruited agricultural and skilled workers from Spain, Italy, and Portugal; at the same time emigration from Colombia to Venezuela also increased. Approximately one million immigrants entered the country between 1948 and 1958, although many of these eventually returned home. After 1958 the government tightened immigration controls, preferring a selective immigration that favoured foreigners with high-level skills. Colombians, however, continued to move into the rural sector during the 1960s as replacement labour for city-bound rural Venezuelans.

The mid-1970s petroleum boom marked another shift in immigration policies. A large demand for labour occurred, particularly semiskilled and skilled, in all sectors of the economy. At the same time, political instability and strife caused an exodus of professional and technical workers and their families from Argentina, Chile, and Uruguay, many of them relocating in Venezuela. After 1976 Venezuela again tightened its controls on immigration from other South American countries, favouring professionals from the United States, Italy, Spain, and Portugal. Throughout this period of relative prosperity and expansion of the Venezuelan economy, the volume of illegal immigration matched that of legal entry. Particularly serving the unskilled sectors of the domestic economy, illegal immigrants came in large numbers from Colombia, Ecuador, and Peru. This alien population is thought to exceed two million, many of them females employed in domestic service. The large and rapid influx of foreigners, especially illegal Colombians, has generated a xenophobic outcry that is the cause of considerable political debate and concern.

Illegal immigration

Ethnic and racial groups are regionally oriented. Whites and mestizos are found mainly in the major cities. The Indian minorities survive only in the remote interior—in the Guiana Highlands and in the forests west of Lake Maracaibo. Peoples of African ancestry and the mulatto-mestizo groups predominate along the Caribbean coast. There are physiological distinctions between the highland and lowland mestizos, which reflect the Hispanic mixing with the Indian populations of the two regions.

The Indian groups speak more than 25 different languages, most of which belong to three linguistic families—Cariban, Arawak, and Chibcha. Spanish is the national language of the majority. Local idioms, colloquial phrases, and simplified verb usage distinguish Venezuelan Spanish from other Latin-American and Iberian variants. In Caracas and other major commercial centres, English is often favoured in business communications, and private schools in Caracas have come to encourage bilingualism in their students. The presence of English-speaking professionals in the oil centres and in the major cities has encouraged the establishment of English as the country's most popular second language.

Freedom of religion in Venezuela is guaranteed by the constitution, and, although more than 90 percent of the people are adherents of Roman Catholicism, religious tolerance is generally observed. Various Protestant sects form the largest minority group, and there are small groups of Jews and Muslims. Some Indian peoples continue to practice their traditional religions, but many have converted to Roman Catholicism. The Roman Catholic church, while officially apolitical, has become increasingly an instrument for social progress.

Religion

Demographic trends. Venezuela's 20th-century population growth has been among the most rapid in Latin America, prompted by high birth rates, declines in mortality rates, and successive waves of immigration. After World War II Venezuela's mortality rate began to drop with advances in medicine and technology to combat malaria and yellow fever; progress in the treatment of tuberculosis, typhoid, bronchitis, and dysentery; improvements in hygiene and diet; and the upgrading of housing conditions. While birth rates continued to remain at high levels, mortality rates, which had been as high as 30 per thousand before 1920, dropped to below 10 per thousand by the 1960s. Since that time the mortality rate has stabilized, and demographic changes have been mostly influenced by reductions in fertility levels and immigration rates.

Fertility rates, nevertheless, have remained relatively high, a condition that maintains a young population, more than 60 percent being under 30 years of age. Life expectancy rates have also been rising, a trend contributing to rapid population growth, with its accompanying challenge of creating economic opportunities for the new generation.

THE ECONOMY

The Venezuelan economy is based primarily on the production and exploitation of petroleum. Until 1970 the country was the world's largest petroleum exporter, but it was overtaken in that year. The modernization and diversification of its economy have been predicated upon the application of petroleum sector earnings to other economic sectors; "sowing the oil" ("sembrando el petróleo") has been the slogan since the 1940s. The potential for such diversification and economic growth has been expanded as a result of the discovery of rich deposits of iron ore, nickel, coal, and bauxite, as well as the development of hydroelectric potential.

"Sowing the oil"

During the 1960s Venezuelan governments stressed import substitution, within the context of protective tariffs and government subsidies. This led to an expansion of new export-oriented enterprises. In the mid-1970s the multinational hold on the Venezuelan oil and gas industries was effectively broken, and a nationalized industry was developed and used to underwrite the massive programs to expand economic infrastructure and public works.

The progress of Venezuela's programs based on "sowing the oil" was considerably set back because of such external conditions as the oil glut beginning in the late 1970s and the global recession of 1980–83, as well as internal problems of inflation, inefficient management, corruption, and deficiencies in skilled personnel. A growing foreign debt, rising unemployment, and illegal immigration were other pressures on the economy. Although its economic problems were becoming increasingly severe, there were some advances. Investments increased farm output, created a more diversified heavy and light industry, and furthered development of the country's natural resources; and at the same time two new planned industrial cities (Ciudad Guayana and El Tablazo) were created.

Resources. The largest and richest petroleum deposits in Venezuela are in the Maracaibo Lowlands. Other deposits include those in the eastern part of the Llanos, the Orinoco delta, and offshore. At the time of nationalization in 1976, production was dominated by several multinational firms, which together accounted for more than 80 percent of production. Refining was primarily accomplished offshore in Aruba, Curaçao, and elsewhere in the Caribbean. After nationalization a state-owned company, Venezuela Petroleum South America (Petróleos de Venezuela S.A.), assumed responsibility for production, but this industrial effort still depended heavily upon foreign oil companies to refine, transport, and market the oil and natural gas and to provide technical assistance. Venezuela also has abundant natural gas deposits; those discovered north of the Paria Peninsula in eastern Venezuela have received priority government development.

Nationalization of oil

Venezuela is also a country rich in other mineral resources, which are largely underdeveloped. The major nonfuel mineral exploited is iron ore. The industry was developed on the basis of concessions granted for 50 years in 1950 to U.S. Steel and Bethlehem Steel to mine the ore in the region surrounding present-day Ciudad Guayana, at Mount Bolívar, and at El Pao. In 1975 the U.S.-owned mining operations were nationalized, and a government-owned corporation, Venezuelan Guayana Corporation, subsequently administered the operations and production.

In the mid-1970s large deposits of bauxite were discovered in the Guiana Highlands, much of this being high-grade ore suitable for alumina smelting in the Ciudad Guayana complex. Important mineral resources also include gold and diamonds in the Guiana Highlands, coal northwest of Lake Maracaibo, salt deposits in the Araya Peninsula, and scattered deposits of industrial-grade limestone. Other mineral resources include nickel, phosphates, copper, zinc, lead, titanium, and manganese, and surveys indicate the existence of substantial deposits of uranium and thorium. Inland and ocean fishing grounds and the inland forests are other less developed resources.

Apart from petroleum and natural gas, Venezuela's rivers constitute the most important source of power. The Caroní River, a major tributary of the Orinoco, has the largest hydroelectric potential. Also part of the Orinoco system, the Santo Domingo River, flowing out of the Mérida Range of the Andes, is the second most important power resource. There is also hydroelectric potential in the shorter course Andean rivers.

Hydroelectric power

Agriculture, fishing, and forestry. Historically, agriculture has been one of the weakest and most neglected sectors of the Venezuelan economy. Only a very small part of the country (less than 5 percent) is used for crop production. Most of the cropland is in the northern mountains or in their foothills. Extensive cattle grazing is practiced in the Llanos and in a more limited way in the Maracaibo Lowlands. South of the Orinoco, the interior forests are farmed by shifting cultivation and in small, cleared riverine plots. Corn (maize) and rice are significant food crops.

Agriculture in Venezuela is organized broadly in three basic types of holding. First, commercial crop farms, *fincas comercializados,* usually more than 50 acres (20 hectares), use wage labour, have some mechanization, and use fertilizers and insecticides. These modernized farms have benefited from government initiatives toward the provision of credit and from higher prices in local and export markets. They produce sugarcane, cotton, and rice, often grown as plantation crops. The second type of holding is the family farm, *conuco,* invariably leased by the farmer; it is usually small in size, concentrating on food crops, such as corn and beans, for local consumption and on tree crops, mainly coffee and cacao, for sale in urban or overseas markets. Venezuela's main agricultural export is coffee. Many of the *conucos* are being consolidated as larger commercial units using immigrant labour from neighbouring Colombia. The third type are the large pastoral farms, *fincas granderas,* often more than 6,000 acres in area. These are established in the Llanos, where unenclosed land is used for grazing cattle on the low-quality grasses. The cattle are herded and traded in yearly meetings called *rodeos* (roundups).

Land reform programs were launched in the late 1950s and early 1960s in an attempt to alter the backwardness and stagnation of the small-farming agricultural sector. Relics of the colonial encomienda system, which supported a type of feudal landholding, led to the uneven distribution of land ownership that allowed some 2 percent of the owners to control about 80 percent of the land. Most rural workers could not own enough land to support their families. Some reform has been implemented in the Andes foothill areas, and more than 500,000 farm families have been reorganized, but the agricultural sector is still struggling to recover from centuries of neglect and the smallness of its resource base. Production has been expanding generally, if irregularly, but many foods still need to be imported.

Land reform

Historically, fish have not been in great demand in Venezuela, and there has been only a moderate exploitation of the inland and ocean fishing grounds. A government enterprise established in the 1970s was designed to develop the fishing industry and increase the demand for fish, especially among the lower-income groups. Anchovies have been a major catch. A large part of the catch is sold fresh locally, but there is some exportation, particularly that of shrimp. Although more than 35 percent of the land is forested, the forestry industry has been slow to develop, mainly because of the remoteness of the richest forestlands. Also strict government conservation regulations have discouraged investment.

Industry. Until the 1950s Venezuela had little manufacturing industry apart from plants concerned with processing agricultural products and oil extraction. Huge oil revenues, combined with low tariffs on imported goods, permitted the importation of even the most simple of items. Since that time, and especially since the government's commitment to diversification of the economy in the 1960s, manufacturing has been transformed. Among the factors contributing to the industrial base are an abundant supply of oil, natural gas, and electrical power; a variety of raw materials; considerable capital available within the country; and a relatively high purchasing power per capita. The establishment of consumer goods and metalworking industries was accomplished under the protection of high tariffs and import quotas. As a result of the 1973–74 rise in world oil prices, government revenues expanded, and investment strategy was directed toward large-scale resource-based projects: iron and steel manufacturing, aluminum smelting, production of transport equipment, petrochemicals, and a foundry and metal engineering complex. This industrial progress slowed when oil prices went into decline several years later.

Transformation of manufacturing

Venezuela's modern industries fall into three groups. First are the oil refineries and associated petrochemical plants. These are centred at such places as Morón near Puerto Cabello, which supports a petrochemical complex, and in Zulia state, where a large-scale petrochemical industry has been built at El Tablazo, on Lake Maracaibo. A second industrial group is engaged in the production of consumer

goods, much of which takes place in the Valencia–Maracay–Caracas area, with some parallel development at Barquisimeto. Import substitution items are the focus of this industry, including textiles, leather, paper, tires, tobacco, light engineering products, and such modern appliance items as radios, television sets, and washing machines. The motor vehicle assembly industry is well established, but progress has been hindered by poorly conceived bilateral agreements with multinational manufacturers, necessitating a restructuring of this industry. A third industrial group is composed of the complex of heavy industries that are being developed at the Ciudad Guayana complex in the Orinoco–Caroní region. A large integrated iron and steel works, at Matanzas near Puerto Ordaz, serves domestic needs and supplies a growing export market. In the same industrial complex, at San Tomé, Venezuela's aluminum industry has expanded from virtually nothing at the beginning of the 1980s to become one of the world's leading producers. Expansion in this industry has been enhanced by technical efficiency that permits high-grade aluminum to be produced at much below the average world price.

Ciudad Guayana

Finance and trade. The government plays a key role in the operation of Venezuela's financial system. It directly supports national banks in their handling of medium- and long-term credit arrangements designed to curtail dependency on international financial institutions, and it also channels the petrodollar profits of the 1970s into productive sectors of the economy. The Central Bank exercises the customary functions of issuing the national currency, the bolivar, regulating the money supply, and granting loans to the commercial banks. Banking and capital markets laws were instituted in the 1970s to govern the financial operations of private and public commercial banks as a means of directing and assisting domestic enterprises. In the 1980s, however, the government had to take over several leading private banks. With a growing trade deficit among the country's other financial problems, successive devaluations of the bolivar and austerity measures also have been undertaken. Venezuela's financial well-being in part continues to depend upon world oil price fluctuations and the policies of the Organization of Petroleum Exporting Countries (OPEC).

Venezuela was a leader in founding OPEC, signing the agreement in 1960 that led to the creation of the organization. That organization's raising of oil prices by more than 400 percent in 1973–74 provided the country with considerable windfall profits—its oil income rising from $2 billion in 1972 to $14.5 billion in 1983. The huge oil revenues vastly increased Venezuelan influence in Latin America. The country has entered into favourable agreements with its neighbours, providing oil and gas at arranged prices. Venezuela has also helped to finance international cartels in such other commodities as banana and coffee production and marketing. Venezuela is a member of the Latin American Free Trade Association (LAFTA), and it joined the Andean Pact in 1973.

Despite the drop in oil prices and the emergence of such exports as steel and aluminum, the main feature of Venezuela's external trade remains the dominance of oil. Although efforts have been made to diversify trade, such as providing steel to China, there remains a high level of dependence upon the U.S. market both as an outlet for petroleum products and as a major source of imports, including machinery and transport equipment, pharmaceuticals and chemical products, food products, and beverages and tobacco. Venezuela has generally had a favourable balance of trade. Its chief trading partner by far is the United States; others include Japan, West Germany, France, Italy, Brazil, Colombia, and The Netherlands.

Major trading partners

Transportation. The nation's transportation system is well developed, especially in the densely populated northern and northwestern regions. Domestic travel depends largely on roads, while industrial transportation needs are served by coastal shipping routes as well as by inland waterways. Air services provide access to regions without other means of communication.

There are three major trunk roads—a section of the Pan-American Highway that runs southwestward from Caracas through Valencia and Barquisimeto; to San Cristóbal, and then into Colombia; the northwestern highway, which runs from Valencia to Coro and on to Lake Maracaibo; and the Llanos Highway, which extends eastward from Caracas to Barcelona, Cumaná, and beyond. A branch also runs from Barcelona across the Llanos to Ciudad Bolívar.

Railways, both for passenger and freight transport, are relatively unimportant. One public line runs northeastward from Barquisimeto to Puerto Cabello on the coast and on to Caracas. Private railways serve the iron and steel industry, connecting mines in the Guiana Highlands region to Ciudad Guayana.

Almost all of the nation's foreign commerce is carried by sea. There are a number of ports, of which several are used by international shipping; many small ports serve fishing or coastal trade purposes. General cargo is handled at eight major ports run by the government. Inland waterways are utilized principally around Lake Maracaibo and on the Orinoco River. A dredged channel between the Gulf of Venezuela and Lake Maracaibo allows seagoing vessels to dock at the ports of Maracaibo, Bobures, and La Salina. A dredged channel through the Orinoco delta permits seagoing vessels also to sail upriver to Ciudad Guayana.

Ocean shipping

Transoceanic air routes use Venezuelan international airports as a stopover, as do flights between North and South America. Simón Bolívar Airport, located at Maiquetía 10 miles by road from Caracas, is the busiest airport, providing international and domestic flights.

GOVERNMENT AND SOCIAL CONDITIONS

Government. The Venezuelan constitution prescribes a government based on republican, democratic, and federalist principles. It is divided into executive, legislative, and judicial branches, none of which may prevail over the others. The 1961 constitution provides for popular elections of the bicameral congress and of the president, to be held at five-year intervals. The Congress consists of a Senate and a Chamber of Deputies. The Senate is made up of two senators from each state and the Federal District; a few additional senators are elected on the basis of minority representation, and former presidents are awarded lifetime seats in the chamber. The number of deputies for each state and the two federal territories is based on proportional representation. Numerous offshore islands under Venezuelan control are federal dependencies. Executive power is exercised by a directly elected president, who is head of state and of the armed forces. He may not serve consecutive terms. The elections are contested by political parties, whose existence is guaranteed by the constitution; two major parties have dominated Venezuelan politics: Democratic Action (Acción Democrática; AD) and the Christian Social Party (Partido Social-Cristiano; COPEI).

Each state and federal territory is headed by a governor who is appointed by the president. State legislative assemblies are unicameral bodies composed of two elected representatives from each of the state's countylike districts. Regarding the day-to-day administration of local affairs, authority is vested in the municipal councils, which function in the capital of each administrative district. In a move to decentralize national authority, the municipal councils have been awarded considerable autonomy over revenue distribution to their districts although they are not empowered to raise revenue locally.

Local government

Venezuelans generally enjoy a high degree of individual liberty. Civil and human rights are protected by a strong and independent judiciary. National in character, the judicial structure has no autonomous state courts. At the highest court level is the Supreme Court of Justice, an independent body, which adjudicates civil, criminal, and political cases. Its members are elected by Congress. A judiciary committee appoints members of the lower courts. Protests have grown over the lack of equal civil and human rights protection for the nation's Indian population.

Education, health, and welfare. Modernization and the integration of the country's urban network in the 20th century has brought considerable improvement in the provision of education and of health and welfare services and in governmental support for training and technical skill acquisition, particularly in the major cities.

Education. Preschool and nine years of basic education

is free and compulsory. Private schools play a significant role at the primary and secondary levels. The number of institutions of higher education expanded rapidly between the late 1950s and the early 1980s. Higher education is provided by private and public institutions, and approximately 20 percent of secondary school graduates attend them. Caracas is an educational centre with several notable universities, including the Central University of Venezuela and the National Open University. Several of the states have major universities, among the largest of which are the University of Zulia and the University of Carabobo.

Health and welfare. The government has greatly expanded health and welfare services, again particularly in the cities. Medical assistance is both public (free) and private. The Ministry of Health is responsible for the organization and staffing of the public hospitals and rural medical centres. The Venezuelan Institute of Social Security Insurance offers medical and welfare assistance to urban workers and employees. In the area of housing, however, the metropolitan authorities have been unable to render sufficient support to supply the needs of the urban poor and the cityward migrants. The problems of the *ranchos* persist as public housing schemes meet mainly the needs of middle-income groups. Lower-income groups are left largely on their own to find employment.

The Ministry of Health

CULTURAL LIFE

Cultural traditions. In Venezuela, as in many Latin-American countries, the rich popular culture of the rural campesinos reflects a blending of Roman Catholicism and traditional folkways. The rural culture, however, was paralleled by the development of an elite culture, predominantly urban-based and European-oriented but distinctly Venezuelan in form. The rapid cityward migrations of the more dynamic and youthful segments of rural communities brought to the *ranchos* some of the local folk traditions, but much of the regional flavour was lost. Colonial Hispanic traditions continue to persist, however, in spite of the North Americanization of Venezuelan urban life. For example, compadrazgo (co-parenthood; a set of social and moral obligations between a child's parents and godparents) is a deeply held family tradition that is retained by the urban classes. The vast inequalities in Venezuelan society are reflected in the nation's two distinctive cultural milieus—one elite, modernized, and externally oriented; the other nationalistic and selectively folk-oriented.

The arts. In an effort to maintain cultural autonomy in the face of foreign influences, the government has, since the 1920s, supported nationalistic artistic expression. Greater freedom for writing and publishing prompted the flourishing development of national literature. Emerging in this wave of nationalistic influence was Rómulo Gallegos, who became Venezuela's best-known writer; he received international recognition for his novel *Doña Bárbara* (1929). Such painters as Armando Reverón and Manuel Cabré also expressed nationalistic fervour. The architect Carlos Raúl Villanueva won international acclaim for his design of asymmetrically arranged buildings, complemented by freestanding murals and sculpture, at the Central University in Caracas. The state-supported Venezuelan Symphony Orchestra is highly popular, and its repertoire reflects as well a spirit of nationalism. Each of the regions has its own distinctive musical expression.

Cultural institutions. Most of Venezuela's major cultural institutions are located in Caracas. The Museum of Fine Arts, founded in 1938, houses a large collection of national and foreign paintings and sculptures. Other museums include the Museum of Colonial Art, housed in an 18th-century mansion, and the Museum of Natural Science (founded in 1940), which contains more than 20,000 exhibits. Among the noteworthy museums outside Caracas are a pre-Columbian museum in Ciudad Bolívar and the Museum of Military History in Maracaibo. The most important library is the National Library in Caracas, with holdings of more than two million volumes, including a large collection of rare books.

Recreation. Venezuela continues to struggle against foreign influences to retain such traditional pastimes as the *toros coleados,* a form of bullfighting in which the bull is thrown by its tail. Nonetheless, there have been widescale adoptions of such North American pastimes as baseball, now the national sport. Carnival, the two days preceding Ash Wednesday, is the major holiday (fiesta) in Venezuela, particularly in Caracas. On these days business comes to a halt as games, races, and street celebrations prevail. Other important holidays are the New Year and, in rural areas, the local saints' days.

Carnival

Press and broadcasting. Caracas is the national press centre, and its newspapers are widely available throughout the interior. The majority of newspapers, including the two leading dailies, *Ultimas Noticias* and *Meridiano,* are independent and generally operate with little government interference. In addition, political parties and labour unions publish their own newspapers. Television broadcasting is available to more than 80 percent of Venezuelans. *Telenovelas* (soap operas) are the most important genre, followed by variety programs. Venezuela has decreased its dependence on American and European programs, an indication of the increasing trend to produce more programs domestically.

For statistical data on the land and people of Venezuela, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (D.Co./I.S. de S.)

History

The oldest inhabitants of Venezuela were primitive food-gathering Indians who arrived in the Late Paleolithic period. There followed, successively, invasions by other food-gathering groups, by community-dwelling Arawak, and by warlike, cannibalistic Caribs. The most advanced Venezuelan Indians were the farming tribes of the Andes; nomadic hunting and fishing groups roamed Lake Maracaibo, the Llanos, and the coast.

THE COLONIAL ERA

Christopher Columbus arrived in what is now Venezuela in 1498, during his third voyage to the New World. According to some historians, the region was named by Amerigo Vespucci, who, on seeing the native Indian houses built on stilts over water, called it "Little Venice," or Venezuela. The first quarter century of European contact was limited to the northeast coast and confined to slave hunting and pearl fishing; the first permanent Spanish settlement, Cumaná, was not made until 1523. In the second quarter of the 16th century, the centre of activity shifted to the northwest region, where the Welser banking house of Augsburg purchased exploration and colonization rights; German attempts to find precious metals and to occupy the area failed, however, and Spain repossessed the area in 1546.

European contact

In the latter half of the 16th century, Spanish agriculturalists, using Indian slave labour, began effective colonization. Caracas was founded in 1567, and by 1600 more than 20 settlements dotted the Venezuelan Andes and the Caribbean coast. During the 17th and 18th centuries, the Llanos and Maracaibo regions were taken over gradually by various Roman Catholic missionary orders.

The colonial economy was based on agriculture and stock raising. Corn (maize), beans, and beef were the domestic consumption staples; sugar, cacao, tobacco, and hides were the principal exports. Spain's European rivals, the French and English in the 16th century and the Dutch in the 17th century, succeeded in taking over most of Venezuela's commerce until the early 18th century, when Spain established a monopoly trading company. The interests of the latter, however, proved contrary to those of Venezuelan producers, who forced dissolution of the company during the 1780s.

Venezuelan society during the colonial era was headed by agents of the Spanish crown. Royal bureaucrats monopolized the top governing posts, and Spanish clergymen dominated the high church offices. Creoles (native-born whites), however, owned the colony's wealth, principally land, and used it to hold the nonwhite races in bondage: mestizos (persons of mixed ancestry) were generally without property, social status, or political influence; Indians performed forced labour on interior farms or were segre-

Creole domination

gated on marginal lands; blacks were slaves on the coastal plantations. In theory, Venezuela was governed by the Spanish crown through the Audiencia of Santo Domingo in the 16th and 17th centuries and through the Viceroy of New Granada (at Bogotá) from its incorporation in 1717. In practice, however, the Venezuelans exercised a great deal of local autonomy throughout the colonial era.

THE INDEPENDENCE MOVEMENT

A group of Venezuelan Creoles boldly proclaimed their country an independent republic in 1797. Although their effort failed, it forewarned of the revolutionary movements that were soon to inflame Latin America.

In 1806 Francisco Miranda—who had earlier fought under George Washington against the British, served as a general in the French Revolution, and fought with the French against Prussia and Russia—tried unsuccessfully to land on the Venezuelan coast with a group of mercenaries whom he had recruited in New York City. He was recalled to his country four years later, set up a revolutionary junta, had a constitution drafted, and established an independent nation with himself as dictator. In the ensuing war with royalist forces, however, he surrendered to the Spaniards, who sent him first to Puerto Rico and later to Spain, where he died in prison in 1816.

Simón Bolívar

Early in 1813 the revolutionary junta appointed Simón Bolívar commander of the Venezuelan forces. Born in Caracas in 1783 of a wealthy family of Creole landowners, Bolívar had long been a leader in the independence movement. He suffered many reverses in his war against the Spanish forces but was helped by the new Republic of Haiti and by a foreign legion of British and Irish soldiers. The Republic of Gran Colombia, with its capital at Bogotá, was proclaimed on Dec. 17, 1819, and on June 24, 1821, Bolívar defeated the royalist army in the Battle of Carabobo; the last of the royalist forces surrendered at Puerto Cabello on Oct. 9, 1823. In 1824 Bolívar marched south to liberate Peru, and the following year he freed Bolivia from Spanish rule.

During his absence regional rivalries broke out in Gran Colombia, and his prestige was not enough to hold the country together after his return. Venezuela broke away in 1829, and Ecuador soon after. Bolívar died in Santa Marta, Colombia, in 1830, penniless and disillusioned. Today he is honoured as the principal architect of Venezuelan and Latin-American independence.

THE CAUDILLOS (1830–1935)

After the destruction of the colonial system, Venezuela passed through an era of government-by-force that lasted well over a century, until the death of Juan Vicente Gómez in 1935. Backed by their personal armies, much on the order of the later Chinese warlords, a series of caudillos (leaders) assumed power, which they exercised for their personal benefit rather than for that of the nation.

Páez and the Conservatives. The first of the military dictators was General José Antonio Páez, who gave the country better government than it would see again for nearly a century. Bolívar had left Páez in charge of the armed forces of Venezuela, and he was soon in full control of the country. He led the separation movement from Gran Colombia in 1829 and in 1830 convoked a constitutional convention to draw up a separate constitution for Venezuela. He dominated Venezuelan politics throughout the period 1830–48; he was president from 1831 to 1835 and was elected to another four-year term in 1839. He established law and order by subduing ambitious provincial caudillos. Páez ruled in cooperation with the large landholders and leading merchants of the Conservative Party. The constitution that they enacted in 1830 reflected their social and political philosophy—a centralist state, property qualifications for voting, death penalty for political crimes, freedom of contracts, and continuance of slavery. The church lost its tax immunity and its educational monopoly, and the army was shorn of its autonomy; thus, state supremacy was achieved. Stability thus assured, reconstruction of the war-torn economy began. Government finances were put in order, the nation's credit was firmly established abroad, and amortization of the national debt was begun. Construction of new roads promoted interior commerce and the export of coffee and cacao.

The Conservative oligarchy, 1830–48

In contrast to the troubled times that preceded and followed it, the 1830–48 period of Conservative Party domination was an era of political stability, economic progress, and responsible administration. An opposition movement began to develop in 1840, however, when Antonio Leocadio Guzmán, the leading spokesman for dissident merchants and professional men, founded the Liberal Party. Guzmán's new liberal newspaper, *El Venezolano,* demanded abolition of slavery, extension of voting rights, and protection for the debtor classes. Declining demands in the world market for Venezuela's agricultural commodities during the 1840s produced economic difficulties, which in turn contributed to the increasing opposition to the Conservative oligarchy.

The Monagas and the civil wars. The growing political crisis was brought to a head in 1848 by General José Tadeo Monagas. Although elected president as a Conservative in 1846, he soon gravitated toward the Liberals. He intimidated the Conservative congress and appointed Liberal Party ministers. When Páez rebelled in 1848, Monagas defeated him and forced him into exile.

Liberalism and federalism, 1848–70

The decade 1848–58 was one of dictatorial rule by José Tadeo Monagas and his brother, General José Gregorio Monagas, who alternated as president during the period. Liberal Party laws were passed abolishing slavery, extending suffrage, outlawing capital punishment, and limiting interest rates, but they were not implemented. Integrity in government waned; heavy deficit financing ruined the nation's credit; the economy began to stagnate and decay. In 1857 the Monagas brothers attempted to impose a new constitution extending the presidential term from four years to six and removing all restrictions on reelection. The Liberal leaders thereupon joined the Conservative opposition, and in March 1858 they brought the Monagas dynasty to an end. This first successful rebellion in Venezuela's national history set off five years of revolutionary turmoil between the Liberals and Conservatives. The issues in these so-called Federalist Wars were, on the Liberal side, federalism, democracy, and social reform and, on the Conservative side, centralism and preservation of the political and social status quo. The conflicts were extremely bloody, and control of the central government changed hands several times. General Páez returned in 1861 to restore Conservative hegemony for two years, but in 1863 final victory went to the Liberals, led by the generals Juan Falcón and Antonio Guzmán Blanco.

A new constitution enacted in 1864 incorporated the federalist principles of the victors. Local freedoms quickly disappeared, however, at the hands of provincial caudillos. As president in 1864–68, Falcón appeared content to allow subordinates, many of them irresponsible, to rule at both the state and national levels. Liberal mismanagement and increasing political chaos provided an opportunity for the Conservatives, now led by José Tadeo Monagas, to return to power in 1868. But civil war followed. General Guzmán Blanco rallied the Liberals to his cause, overthrew the Conservatives, and assumed power in 1870.

The reigns of Guzmán Blanco and Crespo. Guzmán Blanco's triumphal entry into Caracas in April 1870 halted the political chaos and economic stagnation that had plagued the nation since 1858. The new president took the field himself and pacified the country in less than two years; he thereupon launched a broad program of reform and development. A new constitution in 1872 proclaimed representative government, universal suffrage, and direct election of the president. Economic reforms, such as restoration of the nation's credit by means of new bond issues and liberal concessions to foreign investors, gave further evidence of Guzmán Blanco's apparent devotion to Liberal Party principles. He ordered establishment of a nationwide system of public primary education and liberal state support for secondary and higher education. He not only abolished ecclesiastical privileges, cut off state subsidies to the Roman Catholic church, proclaimed religious liberty, and legalized civil marriage but also confiscated church properties, exiled the archbishop, and closed the convents.

Curtailment of church power

Guzmán Blanco was the popular choice for president in the 1873 election. He departed for Europe in 1877, leaving a puppet successor in charge; but when the opposition rebelled, he returned to crush it and resumed the presidency in 1878. The following year he returned to Europe, leaving General Joaquín Crespo in charge. Guzmán Blanco came back again in 1886 to serve a final two years in the face of growing popular opposition to his policies.

Unquestionably, Guzmán Blanco's regime had both positive and negative results for the nation. His admirers point to his political and military genius and to his administrative, economic, educational, and religious reforms. His detractors emphasize his tyrannical ruling methods, his financial chicanery, his monumental vanity, his superficial educational reforms, and his unwarranted attacks upon the church. For four years after the end of his regime, Venezuela floundered in new political chaos as various civilian political groups tried unsuccessfully to establish responsible representative government. In October 1892 Crespo seized power. His six-year rule was troubled by continued political turmoil, growing economic difficulties, and the nation's first serious diplomatic problem—concerning a dispute with Great Britain over the boundary between eastern Venezuela and western British Guiana. This jungled no-man's-land, in which gold was discovered in 1877, had been the object of alternating claims and counterclaims between Venezuela and Great Britain for more than half a century. Great Britain repeatedly refused Venezuela's requests to refer the matter to arbitration, and in 1887 Venezuela suspended diplomatic relations. President Crespo appealed to the United States, and in 1895 U.S. president Grover Cleveland pressured Britain to arbitrate. An international tribunal handed down a decision in 1899 that failed to satisfy Venezuela's demands.

Castro and Gómez, 1899–1935

The Andinos. The turn of the century was a turning point in Venezuelan history. In 1899 General Cipriano Castro, a caudillo from the Andean state of Táchira, descended with his provincial army upon Caracas and seized the presidency. For the next 59 years, except for an interlude in 1945–48, five successive military strongmen from Táchira, known as Andinos, controlled the nation. Castro ruled from 1899 to 1909. His regime was characterized by administrative tyranny, financial irresponsibility, almost constant domestic revolt, and frequent foreign intervention. The most serious internal uprising occurred in eastern Venezuela in 1902–03. This and subsequent revolts of the early part of the century were put down by General Juan Vicente Gómez. Castro's cavalier treatment of foreign businessmen and diplomats and his refusal to pay for foreign properties damaged in domestic insurrections resulted in a British-German-Italian blockade of the Venezuelan coast in 1902–03 and a Dutch attack upon Venezuela's navy in 1908. Ill health forced Castro's departure for Europe for medical attention in 1908, whereupon Gómez usurped the presidential powers and did not relinquish them until his death 27 years later.

Gómez was an effective dictator. By manipulating elections, abolishing all organized political activity, and monopolizing appointive powers, he was able to establish a completely subservient legislative and judicial structure. He muzzled the press and stifled the opposition with an elaborate spy service, and he used arbitrary arrests, exiles, long imprisonments, and assassinations to insure his control. Efficient police and army organizations maintained his power through unrestricted use of force.

Political order and liberal concessions attracted foreign petroleum investors. Dutch and British petroleum interests entered Venezuela just before World War I; immediately after the war Standard Oil interests from the United States arrived to compete with the British and Dutch. By 1928 Venezuela had become the world's leading exporter of oil and was second only to the United States in oil production. The oil industry brought the nation such benefits as high-paying jobs, subsidies to agriculture, expanded government revenues, and increased trade. Networks of roads, railroads, and port facilities were constructed; the entire foreign debt was paid off; and the large domestic debt was drastically reduced. Yet the oil prosperity was unevenly distributed; most Venezuelans continued to live in poverty, and their health, housing, and education needs were ignored by the state. Meanwhile, Gómez and the top bureaucrats and army officers enriched themselves; the dictator became the nation's wealthiest citizen, retaining power until his death, from natural causes, in 1935.

VENEZUELA SINCE 1935

Prosperity, reform, and military rule, 1935–58

Eleazar López Conteras, who had been war minister under Gómez, succeeded him and served as president until 1941. López restored civil liberties, sanctioned political activity, and permitted labour to organize during 1936; but he restored the dictatorship in 1937, when the opposition became too threatening. In 1938 he inaugurated a three-year development plan that included construction of public schools and hospitals and support for agriculture and private industry.

Isaias Medina Angarita, a fellow Táchira general, was president in 1941–45; he continued this development program and also restored political liberties. A World War II transportation squeeze resulted in a sharp decline in petroleum revenues during 1941 and 1942, and President Medina revised upward—under a 1943 oil law—the nation's share in the profits of the petroleum industry. As the transportation shortage eased and new concessions were granted, a petroleum boom was stimulated.

In October 1945, at the height of the wartime prosperity, the Medina administration was suddenly overthrown. This revolution marked the first assumption of power in Venezuela by a political party (Democratic Action) that had the support of a majority of the people. Party leader Rómulo Betancourt headed a civilian-military junta that ruled the nation for 28 months. On July 5, 1947, a new constitution reflecting the labour-leftist philosophy of the party was adopted, and in December 1947 novelist Rómulo Gallegos was elected to the presidency.

Democratic Action promptly launched a program of reform: a fifty-fifty tax decree assured the nation of at least half the profits of the petroleum industry; labour was encouraged to organize and to bargain for its rights; and governmental support was granted for health, housing, and education and for agricultural and industrial development. These reforms provoked strong opposition from conservative forces that culminated in a November 1948 military coup. The new ruling junta was headed by Lieutenant Colonel Carlos Delgado Chalbaud and Major Marcos Pérez Jiménez; two years later the former was assassinated, and the latter took power.

Thus, from 1951 to 1957 the nation was again controlled by a Táchira military dictator. Pérez Jiménez outlawed political activity, crushed the labour movement, closed down the universities, and muzzled the press. Democratic Action's nationwide reform programs were abandoned in favour of modernizing Caracas and enriching the dictator and his army associates. Finally, popular opposition grew so great that the navy and air force joined to overthrow Pérez Jiménez in January 1958. A civilian-military junta ran the country for one year, after which Rómulo Betancourt was elected president.

Democracy and economic nationalism, 1959 to the present

The second Betancourt administration (1959–64) was considerably more moderate than the first. This time Democratic Action, in contrast to its earlier exclusivism, cooperated with the next largest party, the middle-of-the-road Christian Democrats, and set up a coalition government. This government launched programs designed to modernize agriculture, develop domestic industry, improve the nation's health, and eliminate illiteracy. In 1960 it passed an agrarian reform law, and in 1962 it inaugurated a national steel industry.

Despite broad developmental progress, the Betancourt administration was troubled by political unrest and economic crisis. To complicate matters, a sharp depression occurred in 1960–63. In foreign affairs Venezuela severed diplomatic relations with the Dominican Republic in 1960 (after Dominican agents attempted to assassinate Betancourt) and broke relations with Cuba in 1961 (following repeated Cuban attempts to aid the Venezuelan Communists).

The 1963 presidential elections, held in an atmosphere of great political tension, were narrowly won by the

Democratic Action candidate Raúl Leoni. The Christian Democrats thereupon withdrew from the governing coalition, but they were replaced by the labour-leftist Democratic Republican Union. The oil and iron ore industries began to boom once more, and a new petrochemical industry was launched. Although prosperity had returned, growing popular dissatisfaction strengthened the opposition Christian Democrats, whose presidential candidate, Rafael Caldera, won the 1968 elections.

Caldera's inauguration in 1969 marked the first time in Venezuela's history that an incumbent government peacefully surrendered power to an opposition electoral victor. The programs of the Christian Democrats were similar to those of Democratic Action. Caldera improved relations with Cuba, the Soviet Union, and the Latin-American military dictatorships. In the early 1970s Venezuela took majority ownership of foreign banks, took control of the natural gas industry, and declared a moratorium on the granting of oil concessions.

President Carlos Andrés Pérez Rodriguez, the Democratic Action victor in the 1973 elections, nationalized the iron ore industry in 1975 and the petroleum industry in 1976. Following the Arab-Israeli War of 1973, Venezuela, as a founding member of OPEC, more than quadrupled the price of its oil. The result was a spending orgy that attracted a wave of South American immigrants, increased food and luxury imports, produced growing waste and corruption, and created a privileged economic elite while doing little to alleviate poverty. The economic boom did not last, however. An international recession and oil glut beginning in the late 1970s plunged the country into economic stagnation. This condition, continuing into the late 1980s, was reflected in a downward trend in the gross domestic product and a steady increase in inflation; exports declined and unemployment became a major concern. The accompanying loss of confidence in the economy caused an enormous increase in capital flight (Venezuelan investment in foreign markets). That problem and the inability to repay foreign debt reached crisis proportions during the administrations of the Christian Democrat Luis Herrera Campins, elected in 1978, and Democratic Action's Jaime Lusinchi, elected in 1983. Lusinchi adopted austerity measures that would slow capital flight and encourage lenders to reschedule foreign debt; the latter was accomplished in 1985. An economic recovery plan presented by Lusinchi with the 1987 budget was designed to reduce dependency on oil and to encourage industrial diversification. Lusinchi subsequently announced a new economic course involving partial devaluation.

Oil wealth

The Lusinchi government's economic policies did little more than soften the impact of external forces. By 1988 the drastic decline in world oil prices had cut government income in half, and payment of the foreign debt—which Lusinchi had continued as scheduled—became increasingly difficult. In December of that year the electorate returned former president Carlos Andrés Pérez to office. The nation's most popular politician and a leader of hemispheric democracy, Pérez was pledged to developing a regionwide plan to deal with the foreign debt. Domestically he sought to stimulate new growth within existing economic constraints. His succession of President Lusinchi marked the first time in 25 years that the governing party had retained the presidency in an election. President Pérez's popularity was short-lived, however, as riots broke out across the country in reaction to a rise in bus fares, part of a package of austerity measures that Pérez had announced early in 1989. Massive looting took place, and hundreds of people were killed as troops were called out to put down the disturbances.

Return of Carlos Andrés Pérez

For later developments in the history of Venezuela, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

For coverage of related topics in the *Macropædia* and *Micropædia,* see the *Propædia,* sections 964, 966, and 974, and the *Index.* (E.Li./J.D.Ma.)

BIBLIOGRAPHY

Physical and human geography: General introductions are provided by HOWARD I. BLUTSTEIN *et al., Area Handbook for Venezuela,* 3rd ed. (1977); and *Esto es Venezuela* (annual). The country's geography is discussed in DAVID ROBINSON, "Colombia and Venezuela," ch. 5 in HAROLD BLAKEMORE and CLIFFORD T. SMITH (eds.), *Latin America—Geographical Perspectives,* 2nd ed. (1983), pp. 187–240; and PABLO VILA (PAU VILA), *Geografía de Venezuela,* vol. 1, *El territorio nacional y su ambiente físico* (1960), which also includes information on culture. EDGAR BUSTAMENTE (ed.), *Maravillosa Venezuela* (1987), portrays the splendour and diversity of regional landscapes. Geographic, economic, and historical themes are presented by VENEZUELA, DIRECCIÓN DE CARTOGRAFÍA NACIONAL, *Atlas de Venezuela,* 2nd ed. (1979). Studies of the population include JESÚS A. AGUILERA, *La población de Venezuela: dinámica histórica, socioeconómica, y geográfica* (1980); and ANGELINA POLLAK-ELTZ, "The Family in Venezuela," in MAN SINGH DAS and CLINTON J. JESSER (eds.), *The Family in Latin America* (1980), pp. 12–45.

Economic development issues are addressed in THE ECONOMIST INTELLIGENCE UNIT, *South America* (1987), a concise introduction; and LORING ALLEN, *Venezuelan Economic Development: A Politico-Economic Analysis* (1977). Regional disparities and urban and regional planning strategies are discussed by JOHN FRIEDMANN, *Regional Development Policy: A Case Study of Venezuela* (1966); and LLOYD RODWIN *et al., Planning Urban Growth and Regional Development: The Experience of the Guayana Program of Venezuela* (1969).

A systematic history of Venezuelan politics and economy since 1958 is found in JOHN D. MARTZ and DAVID J. MYERS (eds.), *Venezuela: The Democratic Experience,* rev. ed. (1986), with an excellent treatment of the political system. See also JOHN A. PEELER, *Latin American Democracies: Colombia, Costa Rica, Venezuela* (1985). Studies of political parties include JOHN D. MARTZ, *Acción Democrática: Evolution of a Modern Political Party in Venezuela* (1966), a sympathetic history and analysis of the period 1941–64; and DONALD L. HERMAN, *Christian Democracy in Venezuela* (1980). For the political left, see STEVE ELLNER, *Venezuela's Movimiento al Socialismo: From Guerrilla Defeat to Innovative Politics* (1988). ENRIQUE A. BALOYRA and JOHN D. MARTZ, *Political Attitudes in Venezuela: Societal Cleavages and Political Opinion* (1979), examines democratic public opinion. DANIEL H. LEVINE, *Religion and Politics in Latin America: The Catholic Church in Venezuela and Colombia* (1981), analyzes historical and current relations.

The Venezuelan oil industry is treated in EDWIN LIEUWEN, *Petroleum in Venezuela: A History* (1954, reprinted 1967), a view of the industry from 1907 to 1950; FRANKLIN TUGWELL, *The Politics of Oil in Venezuela* (1975); JAMES F. PETRAS, MORRIS MORLEY, and STEVEN SMITH, *The Nationalization of Venezuelan Oil* (1977); RÓMULO BETANCOURT, *Venezuela: Oil and Politics* (1979; originally published in Spanish, 2nd ed., 1967), a review by a former president; GUSTAVO CORONEL, *The Nationalization of the Venezuelan Oil Industry from Technocratic Success to Political Failure* (1983); and DAVID EUGENE BLANK, *Venezuela: Politics in a Petroleum Republic* (1984).

Venezuela's art and music are discussed in ALFREDO BOULTON *et al., Arte de Venezuela* (1977) and LUIS FELIPE RAMÓN Y RIVERA, *La música popular de Venezuela* (1976).

History: Overviews of Venezuela's history are found in EDWIN LIEUWEN, *Venezuela,* 2nd ed. (1965, reprinted 1985); J.M. SISO MARTÍNEZ, *Historia de Venezuela,* 9th ed. (1967); J.L. SALCEDO-BASTARDO, *Historia fundamental de Venezuela,* 9th rev. ed. (1982); JOHN V. LOMBARDI, *Venezuela: The Search for Order, the Dream of Progress* (1982); and JUDITH EWELL, *Venezuela: A Century of Change* (1984). See also DONNA KEYSE RUDOLPH and G.A. RUDOLPH, *Historical Dictionary of Venezuela* (1971), succinct information on major events and persons through 1969. JOSÉ DE OVIEDO Y BAÑOS, *The Conquest and Settlement of Venezuela* (1987; originally published in Spanish, 1723), details events from the time of Columbus to 1600. FRANCISCO GONZÁLEZ GUINÁN, *Historia contemporánea de Venezuela,* 15 vol. (1909–25, reissued 1954), contains an encyclopaedic treatment of the 19th century. Specific events and periods are analyzed in BENJAMÍN A. FRANKEL, *Venezuela y los Estados Unidos, 1810–1888,* trans. from English (1977), a fine account of 19th-century diplomatic relations; ROBERT L. GILMORE, *Caudillism and Militarism in Venezuela, 1810–1910* (1964), on the evolution from military personalism to military professionalism; MARIANO PICÓN-SALAS *et al., Venezuela independiente, 1810–1960* (1962), which includes essays on the evolution of society, culture, the economy, and the political system; WINFIELD J. BURGGRAAFF, *The Venezuelan Armed Forces in Politics, 1935–1959* (1972), an account of the 20th-century role of the military; and JACQUELINE ANNE BRAVEBOY-WAGNER, *The Venezuela-Guyana Border Dispute: Britain's Colonial Legacy in Latin America* (1984), which elaborates the arguments. For further historical information, consult JOHN V. LOMBARDI, GERMÁN CARRERA DAMAS, and ROBERTA E. ADAMS, *Venezuelan History: A Comprehensive Working Bibliography* (1977). (D.Co./J.D.Ma.)

Venice

An island city, the greatest seaport in late medieval Europe and its commercial and cultural link with Asia, Venice (Italian: Venezia) is unique environmentally, architecturally, and historically. In its days as a republic, the city was styled *la serenissima* ("the most serene" or "sublime"). It remains a major Italian port in the northern Adriatic and is the capital of Venezia *provincia* (province) and of the Veneto *regione* (region).

Venice is one of the world's oldest tourist and cultural centres. Today the city is recognized as part of the artistic and architectural patrimony of all humanity, a fitting role for a city whose thousand-year economic and political independence was sustained by its role in global trading. The situation of the city on islands has limited modern suburban spread beyond the historic centre; its framework of canals and narrow streets has prevented the intrusion of automobiles; and its unmatched wealth of fine buildings and monuments dating from the period of commercial dominance has ensured a keen and almost universal desire for sensitive conservation.

Since the fall of the Venetian republic in 1797, the city has held an unrivaled place in the Western imagination and has been endlessly described in prose and verse. The luminous spectacle of ornate marbled and frescoed palaces, towers, and domes reflected in the sparkling waters of the lagoon under a blue Adriatic sky has been painted, photographed, and filmed to such an extent that it is difficult to distinguish the real city from its romantic representations. The visitor arriving in Venice is still transported into another world, one whose atmosphere and beauty remain incomparable.

This article is divided into the following sections:

Physical and human geography 480
- The landscape 480
  - Site
  - Climate
  - Layout
  - Canal boats and bridges
  - Palaces
- Legacy of maritime commerce 482
  - The port of Venice
  - Traditional industries
  - Tourism
- Cultural heritage: the myth of Venice 483
  - San Marco
  - Trade-guild buildings
  - Churches
  - Painting
  - Music
- Contemporary life 484

History 485
- The early period 485
  - Origin of the city
  - The new order
- Growth of trade and power 485
  - Relations with the Byzantine Empire
  - The commune
  - Trade conflicts
  - The patriciate
  - Struggle for naval supremacy
  - Zenith of power
- Political and economic decline 486
  - Europeans and Turks against Venice
  - End of the Venetian republic
- After incorporation into Italy 487

Bibliography 487

Physical and human geography

THE LANDSCAPE

Site. Situated at the northwestern end of the Adriatic Sea, Venice lies on an archipelago in a crescent-shaped lagoon that stretches some 32 miles (51 kilometres) from the reclaimed marshes of Jesolo in the north to the drained lands beyond Chioggia at the southern end. The shallow waters of the lagoon are protected by a line of sandbanks, or *lidi* (singular *lido*) whose three gaps, or *porti* (singular *porto*) allow passage of the three-foot (one-metre) tides and the city's maritime traffic. On the sandbanks are many small settlements, some of them centuries old. The best known is Lido itself, built as a fashionable seaside resort in the 19th century.

Lagoon and tides. Originally formed by the interaction of Adriatic tidal currents and the waters of several Alpine rivers (Piave, Sile, Bacchiglione, and Brenta), the lagoon has always been crucial to the survival of Venice. Its mudbanks, shallows, and channels are a source of income from marine and bird life and from salt pans. The lagoon has served as protection (the Venetians defeated the Genoese in 1380 through their superior knowledge of the navigable channels) and as a natural sewerage system, with the tides flushing out the city's canals twice daily.

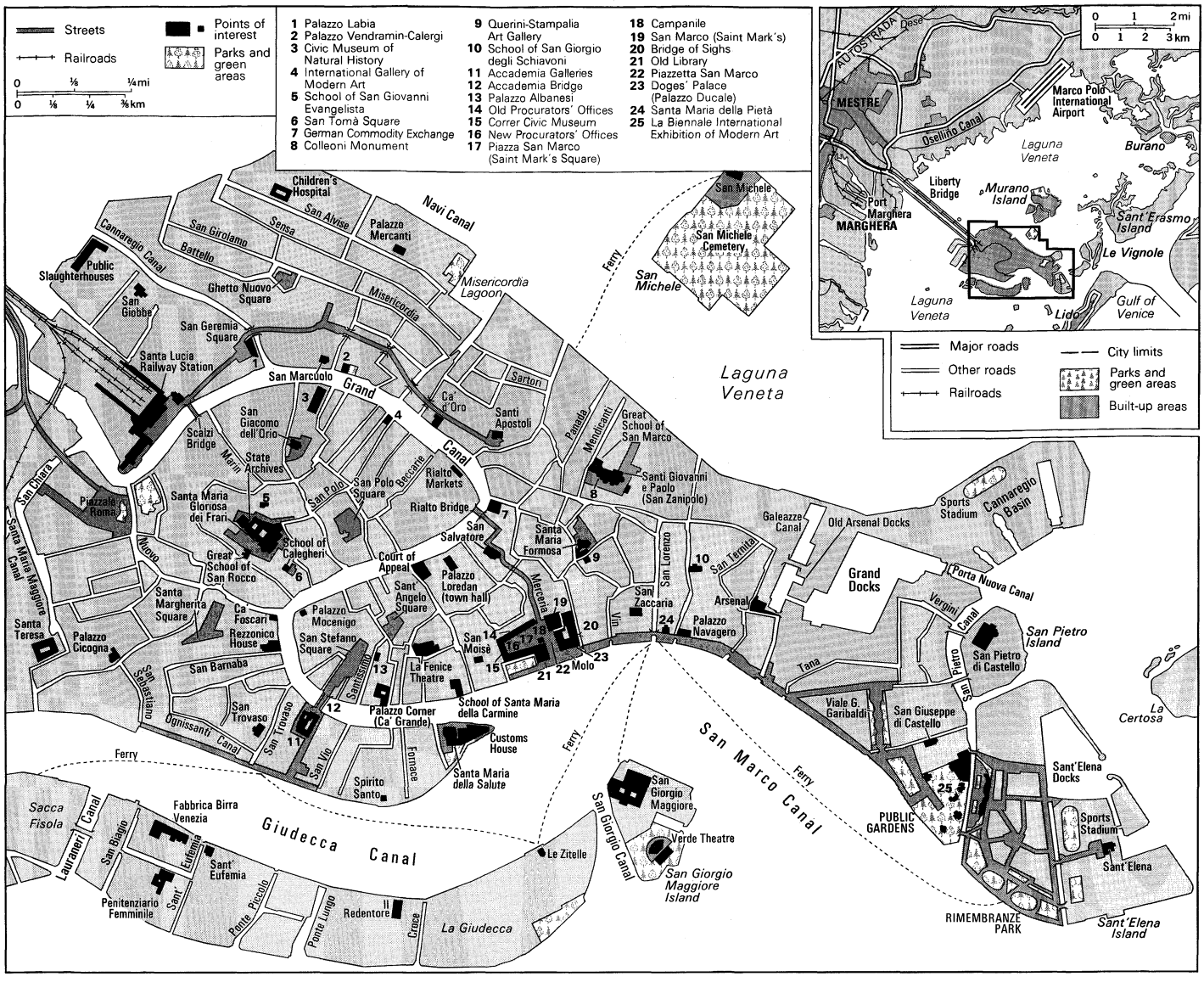
Flooding

But the lagoon requires careful husbandry to prevent it from threatening the very existence of Venice. When high tides combine with storm winds from the south and east, the waters in the lagoon rise and flood the city. The deepening of channels in the 20th century, the overextraction of fresh water from mainland aquifers, and the geologic sinking of the Po River basin have all combined to lower the land level, adding to Venice's flooding problem. The need to protect the lagoon is not new. In the 15th and 16th centuries the Venetian ministry for water matters was already diverting rivers from the lagoon and passing laws to safeguard the urban environment.

Terraferma. Although Venice may aptly be regarded as an isolated sea city, it has always had close links with the surrounding marshlands and the mainland of northern Italy. The Venetian republic included the perimeter of the lagoon, the *dogado;* and, from the early 15th century, amassed a large land empire, known as *terraferma* ("dry land"), stretching from Istria, in modern Yugoslavia, to the borders of Milan and from the Po River to the high Alps. From the 16th century onward, the Venetians invested heavily in the purchase, reclamation, and drainage of *terraferma* lands. The imprint of the republic may still be seen in former subject cities, such as Padua, Verona, and Vicenza, where Venetian Gothic palaces line the streets and the symbol of Venice, the lion of San Marco, stands over the city squares.

Today the administrative city of Venice embraces the 90-mile perimeter of the lagoon, including the mainland urban and industrial complex of Mestre and Marghera and the international airport.

Climate. Most visitors experience Venice in summer, when average daytime temperatures are about 75° to 80° F (about 24° to 27° C), with a haze caused by high humidity frequently obscuring the view of the Alps across the lagoon. Spring and autumn bring clear, bright light, especially when winds are northerly, giving relief from mosquitoes and the exhausting heat of the southerly sirocco. In January the mean average temperature is 36° F (2.2° C), and wintertime Venice is dulled and chilled by mists, lending the city an especially mysterious appearance. Annual rainfall averages 33.6 inches (854 millimetres), of which more than 7 inches (185 millimetres) falls in October and November and about 6.5 inches (168 millimetres) in May and June.



Venice and (inset) its metropolitan area.

Layout. Settlement in the lagoon predates Roman times, but the present urban structure took shape in the early 7th century when migrants from the mainland swelled existing fishing communities on the higher mud flats and sandbanks. Among these early settlements Rivo Alto, its name corrupted over time to Rialto, was the most central and became the heart of Venice, linking together 118 separate islands with bridges and canals and subordinating all other settlements to the rule of its elected doge (duke). In all these lagoonal settlements the characteristic plan, still detectable in the street patterns, was dominated by a navigable channel from which side channels branched at intervals.

The Grand Canal

More than 200 original channels have been linked together to form a dense urban network on either side of the curving Grand Canal, which describes a great backward S more than two miles long, from the railway station to San Marco Basin in front of the Doges' Palace. Its width varies from about 100 to 225 feet (30 to 70 metres), and it is lined by buildings that once were the palaces of great merchant families and the public warehouses, or *fondaci* (singular *fondaco*), used in foreign trade.

The original pattern of separate islands surrounding the Rialto is evident in the parishes of Venice. In many respects they remain distinct communities, with life centred on the square, or *campo* (site of the community well), and its parish church. Perhaps the most clearly recognizable such area today is the Ghetto, the islet on which from 1516 to 1797 Venice's Jews were confined. The Ghetto is located in the northwestern part of the city and is surrounded by canals whose bridges were once raised and guarded at night. Because this was the only area in which Jews could live in Venice, houses are densely packed and rise to seven stories; alleyways are almost too narrow for two people to pass.

Many parishes had their own minor guild or fraternity, and at festivals their representatives competed with one another to provide floats or oarsmen, a ritualized rivalry encouraged by the ruling patricians to promote civic stability. Over time the patchwork of local streets, canals, and quays has been modified to improve the overall structure of the city. Quayside paths have been widened to form canalside walkways, or *fondamenti* (singular *fondamento*), canals have been filled in (*rio terra*), and streets have been joined by passages under the houses. (For the visitor, trying to find an address in Venice is not made any easier by the practice of numbering houses consecutively through a whole district rather than along each street.)

Canal boats and bridges. The best-known form of transport on the waterways of Venice is the gondola. Today there are fewer than 400 of these unique, keelless boats left, and they have long been outnumbered by other vessels. But their elegant, sleek shape and gleaming black paintwork have made them a symbol of Venice. Many writers have described the romance of Venice by gondola, and many tourists are still willing to pay outrageous prices to be rowed at twilight through the canals to the singing of a gondolier. But it is many years since gondoliers could

Gondolas

recite verses from such Italian poets as Ariosto or Tasso while maneuvering their amazingly flexible craft around the sharp bends of the minor canals. A number of gondolas still serve as ferries across the Grand Canal, but the cost of maintenance makes their ultimate disappearance likely.

The canals are filled with a variety of motor-powered boats. They range from the *vaporetti* (singular *vaporetto*), public water buses run by the municipal transport system, to the private motor-launch taxis. Other specialized craft, such as the barges carrying fruits and vegetables, the garbage barges, ambulance and police launches, and the boats filled with tourists' baggage, make up a water scene of endless colour and variety.

Venice is a walking city. Other than at the great parking lots at Rome Square and on the Lido, automobiles are banned from the city. At the risk of encountering frequent detours and dead ends, one can reach any point in Venice by foot; along the banks of the canals, on the paved streets, through the neighbourhood squares, and over the 400 or so canal bridges (*ponti*). Many of the traditional arched marble bridges remain, but large numbers of old bridges were replaced by wrought iron structures in the 19th century.

The Grand Canal is spanned by three bridges. At its most dramatic bend is the famous Rialto Bridge, designed by the aptly named Antonio da Ponte (*c.* 1590). The other two bridges are of more recent origin: the one at the railway station was built of marble in 1932, and the Accademia Bridge, a high-arched wooden structure with a temporary look, has withstood a half-century of foot traffic.

Palaces. The houses (*case,* or, in Venetian, *ca'*) that line the streets and canals of the city range from the poorest blocks to the great palaces (*palazzi*). Ordinary houses generally rise three or four stories. They originally had external staircases and are grouped around a communal courtyard and well. Their simple rectangular doorways and window lights may be framed in unpolished marble; otherwise they are unornamented, their red brick or ochre-painted stucco walls giving a comfortable warmth to the townscape. But it is the palaces, not the ordinary dwellings, that front directly onto the larger canals, particularly the Grand Canal, with gaudily painted mooring posts marking their water entrances.

Although architectural styles changed over time, the structure of these merchants' buildings remained fairly constant. The water story served as the merchant's offices and storerooms, while the floor above was occupied by the patrician family. The piano nobile (principal room), or *salone,* overlooking the canal is lit by a central window group of five or six lights and, characteristically, two single lateral windows. This tripartite fenestration reflects the internal arrangement of rooms running back from the canal to a courtyard at the rear, which frequently contains gardens and greenery. Walls are often of brick faced with dressed stone. The most characteristic building stone is the brilliant white marble shipped cheaply into Venice from Istrian quarries. It is easy to carve but remarkably resistant to weathering, which is severe in the humid, saline—and now acidic—air of Venice.

Marble building stone

The facades of the palaces evolved stylistically from their original Byzantine form, characterized by tall, narrow arches—those of the early Gothic period (13th to early 14th century) pointed and Moorish-looking and those of the 15th century adorned with fantastic trefoil and quatrefoil tracery. In the most ornate late Gothic palaces, such as the Ca' d'Oro (1425–*c.* 1440), the central panel extends across the whole facade and is repeated on two upper stories. In the late 15th century, Renaissance forms began to influence palace architecture, as in the Palazzo Corner, also called Ca' Grande (*c.* 1533–*c.* 1545, designed by Jacopo Sansovino), and the Palazzo Grimani (*c.* 1556, by Michele Sanmicheli, completed 1575). Buildings such as these introduced a measured proportion, tight symmetry, and classical vocabulary to the facade. Mannerist and Baroque palaces built in the 17th century present a decorated classical style with heavy moldings and grotesques, as in the Palazzo Pesaro (1659–1710, by Baldassare Longhena). The variety of styles along the larger canals, unified by the chiaroscuro of deep-set windows, decorative paneling, and building materials, provides much of the visual excitement of the Venetian landscape.

LEGACY OF MARITIME COMMERCE

The landscape of Venice is as much a product of its economic activities, past and present, as of its physical environment. The enduring foundation of Venetian wealth was maritime commerce, initially in local products such as fish and salt from the lagoon, but rapidly expanding to include rich stores of merchandise as Venice became the entrepôt between Europe and the Middle East and Asia. The Rialto remains the core of Venetian commercial and mercantile activity. Fruit, fish, and other markets are concentrated under the open arcades of the Rialto New Building (1554, by Sansovino) and associated buildings. The Rialto Bridge and surrounding streets remain crowded with market stalls. Along the Merceria, the route from the Rialto Bridge to the Piazza San Marco (St. Mark's Square), are the offices of the major banks, still in the traditional banking quarter.

The Rialto

Venetian trade required well-constructed vessels both for transport and for protection from pirates, rivals, and Turkish military forces. Shipbuilding inevitably became a major industry. It occupied a whole sector in the northeast of the city, the Arsenal—a vast assemblage of basins, yards, and workshops for making sails, ropes, and ordnance. At its entrance is an elaborately decorated gateway with a fine group of stone lions guarding what was until the 18th century Europe's largest industrial complex. Parts of the Arsenal are still used for Italian military purposes.

The port of Venice. The main port and related activities have now shifted to the parish of Mendigola in the west. Here the main cruise liners dock, and the offices of shipping lines occupy former palaces. But the real focus of commercial shipping today is Port Marghera, developed on the mainland shore west of Venice. Marco Polo International Airport (1960) was built on reclaimed land to the northwest of the city. Although they are incorporated into the city administration, the chief port activities are now largely separate from Venice proper. Their impact on the old city, however, is considerable. Marghera is the site of a huge oil-refining and petrochemical complex, easily visible from Venice and a source of air pollution that has severely damaged its architecture.

Traditional industries. Scattered throughout the urban area are small boatyards and other traditional luxury craft workshops producing lace, textiles, and furniture. One of Venice's oldest specialties is glassworking. The finest products are of exquisite quality, but most of the present-day glass goods are trinkets for the tourist trade. In 1291 many of the glassworking furnaces were relocated on the island of Murano to the north as a precaution against fire, and Murano remains the focus of present-day glass production. Exhaust fumes from this ancient industry also contribute to the corrosion of Venice's stonework.

Glass-working

Other small island settlements such as Burano, Caorle, Marano, and Malamocco traditionally depended on the local economic activities of the lagoon: fishing and fowling, salt production, and horticulture. Some settlements are swamped by seaside tourist developments, but the ancient trades are still carried on. Fishermen in small craft continue to be common sights in the lagoon.

Tourism. Since the end of the 18th century, tourism has been at the heart of the Venetian economy. Luxury establishments such as the Danieli hotel and the celebrated Caffè Florian were developed in the 19th century for wealthy foreigners. Pensions, food emporiums, and souvenir shops line each major street and square along the routes from the station and parking lots to the Rialto and San Marco. Nearly a third of the city's workers find employment in tourism, now continuous through all seasons.

The tourist industry has been actively encouraged by the authorities. In the early 1980s they revived the ancient Carnival during February, a complement to the round of events like the Biennale (an international art festival held in even years) and international festivals of film, drama, and contemporary music. These, together with the promotion of Venice as an international conference centre, bind the city's economy ever more firmly to tourism.

CULTURAL HERITAGE: THE MYTH OF VENICE

Reacting to their physical environment and to a variety of cultural influences—from Italy, northern Europe, and the East—the Venetians consciously designed their city as an exceptional place. They regarded it as a divinely ordained centre of religious, civic, and commercial life, a community blessed by St. Mark, protected by its lagoon, and governed by a balanced constitution incorporating monarchy, aristocracy, and republican liberty. Historians refer to this perception as the "myth of Venice." The architecture of the city, especially in the Renaissance, purposely emulated republican Rome, and the great rituals of state—the doge's procession from his palace to the basilica or the annual Marriage with the Sea, when the doge cast a gold ring into the lagoon as a "sign of true and perpetual dominion"—publicly expressed the myth.

Historic rituals

San Marco. The administrative heart of the Venetian republic was at San Marco in the buildings surrounding Piazza San Marco and the Piazzetta ("Little Square"). This spectacular piece of town planning depends for its impact on the articulation of paved open spaces, monumental buildings, carefully situated monuments, and the reflective waters of the lagoon basin.

The Doges' Palace. The core of political life in Venice was the Doges' Palace (Palazzo Ducale), whose crenellated mass appears to float upon the waters of the lagoon. Its plan, typical of Venetian palaces, is centred on an internal courtyard with a great staircase (Scala dei Giganti), and it incorporates three great architectural traditions—Gothic, Moorish, and Renaissance. Erected over many years after the burning of the original 9th-century structure in 976, most of the present building dates from the 14th to the 16th century. The palace was not only the residence of the elected doge but also the meeting place of the republic's governing councils and ministries. Its chambers and staircases were richly decorated by a succession of Venetian painters and craftsmen. On the east side of the palace runs a narrow canal spanned by the Bridge of Sighs, which led to the state prisons and is immortalized in Lord Byron's *Childe Harold.*

The Molo. At the entrance to the Piazzetta was the ceremonial landing spot for great officials and distinguished visitors. This "front door" to Venice, the Molo, is marked by two massive granite columns brought from the Orient in the 12th century; one supports the winged lion of St. Mark supporting a book and the other St. Theodore, Venice's first patron, standing on a crocodile.

San Marco Basilica. Sculptured lions are to be found at many points within the square and on its buildings. The key to Venetian political iconography, they symbolize the evangelist St. Mark, whose body was said to be buried in San Marco Basilica, attached to the Doges' Palace. This splendid church, begun in 829 and completed in about 1071, was traditionally the private chapel of the doges and effectively a political building. Its architectural design is Byzantine, with five vaulted domes set in a Greek cross. The interior glows with light reflected from its undulating marbled pavements, its columns and polished stone panels, and its golden mosaics.

Piazza San Marco. Before the five arched portals of the basilica lies the Piazza San Marco, a vast paved and arcaded square. Napoleon called it the finest drawing room in Europe. The northern and southern wings of the square are formed by two official buildings, the Old Procurators' Offices and the New Procurators' Offices. The buildings now house fashionable shops and elegant cafés, whose string ensembles compete with each other in summer months to attract customers to their open-air tables. At the basilica end of the Old Procurators' building stands the Clock Tower, a late 14th-century structure where the hours are struck by two Moorish figures.

Tourists throng the square at all hours, outnumbered only by gluttonous pigeons. The Clock Tower rises over the entrance to the Merceria, the main shopping street leading to the Rialto, and stands in a direct line of sight to the columns at the end of the Piazzetta. This sightline is emphasized by three flagpoles fronting the basilica and by Sansovino's Loggetta ("Small Loggia," or "Small Gallery"), at the base of the Campanile, where council members met. The Campanile, the massive 324-foot (99-metre) bell tower of St. Mark's, dominates the townscape, visible for miles across the lagoon. In 1902 it collapsed, making a fortune for the photographer who captured the event. The city council decided immediately to rebuild it exactly as it was, and the work was completed by 1912.

The Campanile

The Campanile stands close to the 21 bays of the Old Library (1529, also called the Library of St. Mark's or the National Marcian Library), on the western side of the Piazzetta. The library was designed by Sansovino to house a great collection of humanist texts and manuscripts bequeathed in 1468 to the republic. Now a major research library, it numbers among its treasures Marco Polo's will, a manuscript in Petrarch's hand, and many books and maps printed when Venice was a great publishing and cartographic centre.

Trade-guild buildings. The oligarchic form of government during the republic excluded from power all non-noble Venetian families. There were, however, other ways in which ordinary Venetians could participate in public life. One of these was through the *scuole* (singular *scuola*), six major and numerous minor philanthropic confraternities and guilds that originated in the 13th century. Each school had a two-story meeting hall used for gatherings of its members and for discharging its charitable functions. The six great schools became enormously wealthy, enriching their buildings with splendid architectural decoration, as at the Great School of San Marco (founded *c.* 1260, rebuilt after a fire 1487–95; now a hospital), with its trompe l'oeil marble panels. The painted panels and ceilings of the Great School of San Rocco (instituted 1478, completed 1560) are masterpieces by Tintoretto. The School of San Giorgio degli Schiavoni (for Slavic merchants) has the finest collection of Vittore Carpaccio's works outside Venice's chief gallery, the Academy of Fine

Ralph Lieberman

Venice, looking toward the Molo and the Piazzetta, with (left of the Piazzetta) the Old Library and the Campanile and (right of the Piazzetta) the Doges' Palace. The domes of San Marco Basilica appear behind the Doges' Palace.

Arts, whose own collection came in part from a confraternity of flagellants, the school of San Giovanni Evangelista (founded 1261).

Churches. The San Marco Basilica was the focus of public religious life, but the scores of other Venetian churches are an essential element of the city's landscape. Their campaniles, rarely perpendicular, punctuate the skyline; their ornate facades grace the squares, from the delicate Gothic of Madonna dell' Orto (*c.* 1350, rebuilt in the early 15th century) and the restrained elegance of the early Renaissance at Santa Maria dei Miracoli (1481–89) to the Baroque flamboyance of San Moisè (1668).

The most impressive churches are those of the medieval mendicant orders, the Dominicans and Franciscans. The Dominican church of Santi Giovanni e Paolo (San Zanipolo in the Venetian dialect; founded in 1246 and consecrated in 1430), is of rose-coloured stone, its vast interior designed for the large congregations of urban poor whom it served. As a burial place, it was favoured by noble families; a number of doges lie there, commemorated by richly wrought sepulchral monuments. The church's altarpieces, painted by Titian and Giovanni Bellini, were partially destroyed in a fire in 1867, and for masterworks it can no longer rival the Franciscan Santa Maria Gloriosa dei Frari (founded *c.* 1250, completed *c.* 1443), whose enormous Gothic mass rises in the densely settled area west of the Rialto. Titian's "Assumption" (1516–18) stands over its high altar, and the church and sacristy display a magnificent collection of Venetian religious paintings from the High Renaissance.

In contrast to these great Gothic churches, and indeed to the small parish churches enmeshed in the urban fabric, are the church of La Salute and the Palladian churches seen across the water from San Marco. All serve a monumental as much as a religious function. Santa Maria della Salute (begun in 1631/2, Longhena, and consecrated in 1687), erected in thanksgiving for release from plague, occupies a spectacular site where the Grand Canal opens into the San Marco Basin. Its mass of brilliant white marble rises majestically above the Gothic palaces of the Grand Canal. The churches of San Giorgio Maggiore (1566, completed in 1610), Il Redentore (1577–92), and Le Zitelle (1582–86) were all designed by Andrea Palladio; their Roman classical facades were intended to be seen across the waters of the Giudecca Canal. San Giorgio and La Salute turn the open lagoon in front of San Marco into an aquatic extension of the Piazza. Il Redentore is linked to Venice proper by a temporary bridge every July on the Feast of the Redeemer, when illuminated boats fill the Giudecca Canal and there is a display of fireworks.

Palladian churches

Painting. Just as the city's architecture reflects notions of Venice as a place for public ritual, so, too, Venetian painting evokes the "myth of Venice." The magnificent art treasures of the republic now grace churches, palaces, and galleries throughout the city. Early paintings were heavily influenced by Byzantine traditions, as can be seen in the religious icons of Paolo and Lorenzo Veneziano and in the taste for mosaic patternings, vibrant colour, and the love of light that are characteristic of the Venetian school. Painting styles evolved in concert with broader European tastes, and in the 18th century much of the work became more frivolous, even voluptuous, no longer addressing serious themes. Nonetheless, the sense of sparkling colour, the contrasts of light and shade, and the reflective intensity of sky, stone, and water so immediately apparent on the streets and canals of Venice reverberate through all its artistic productions.

The enduring theme of Venetian culture is Venice itself. From the late 15th-century townscape paintings by Carpaccio and Gentile Bellini to the High Renaissance Madonnas and saints in landscapes by Giovanni Bellini, Giorgione, and Titian, the Mannerist canvases of Paolo Veronese and Tintoretto, and the rich early 18th-century townscapes of Canaletto, G.B. Piazzetta, and Francesco Guardi, Venetian painting returned constantly to themes celebrating the city, its great families, its legends, its saints, and its victories. Venetian art was more often than not political art; like all cultural life in Venice, it was subordinated to the interests of the state.

Music. Colour and splendour reflecting civic pride are evident too in Venetian music. The works written for several separate choirs by Giovanni Gabrieli and Claudio Monteverdi for San Marco echoed around its Byzantine interior with stirring effect. After the opening in 1637 of the San Cassiano Theatre (Europe's first public opera house), the commercial flair of Venice's patricians, allied to the secular ambitions of choirmasters of San Marco like Monteverdi and Francesco Cavalli (both noted opera composers), and Giovanni Legrenzi, made Venice the operatic capital of Europe.

Vocal and instrumental traditions were strengthened in the 18th century when four *ospedali,* orphanages run by churches, incorporated conservatories of music. Antonio Vivaldi was master of music at the Santa Maria della Pietà Hospice between 1703 and 1741. Venice's opera house, La Fenice Theatre, built in 1792, is a major Italian music centre, and its winter opera season is one of the city's great attractions. There the premieres of Gioacchino Rossini's *Tancredi* (1813) and Guiseppe Verdi's *Rigoletto* (1851) and *La Traviata* (1853) were witnessed by their composers. Many foreign composers also developed a special attachment to the city.

La Fenice

CONTEMPORARY LIFE

For Venice, adaptation to the demands of the modern world is often a painful process. While the era of vast refining and petrochemical development along the lagoon shore may be over, its damage to Venice in visual and environmental terms has been immense. After the disastrous floods of 1966, the United Nations Educational, Scientific and Cultural Organization (UNESCO) began to coordinate an international effort to preserve the city. A number of national committees, too, now exist to save Venice and its art treasures from the combined effects of corrosive air pollution, rising damp, flooding in high-water periods, sheer age, and even defacement by pigeons. The completion of an aqueduct from the nearby Alps to Marghera has prevented further aquifer exhaustion, and the effects of international funding are increasingly visible in renovated buildings and monuments.

But the most intractable environmental problem remains that of the high tides, or *acqua alta.* A scheme to build a mechanical barrage that can be raised in times of flooding to close the lagoon at the Lido was initiated in 1988, when engineers began testing a prototype. More rapid progress has been hampered by the overlapping local, regional, and national bureaucratic concerns. Venice has its own city council with ultimate responsibility for the daily life of the city, but the regional authority, for example, administers the Veneto area, whose hydrographic and industrial management bears heavily on the problems of the lagoon and city. The ministries of the national government have a direct stake in museums and galleries, in port activities, and in historic buildings. Many of the great industries are nationalized or partly owned by the state, including the railways, the airport, and the oil and chemical industries located in Marghera. The often concurrent jurisdiction and resulting conflicts of interest tend to produce administrative stalemate.

Venice still sustains a distinct urban life. Away from the main tourist routes, children use the squares as playing fields, a poor substitute for the chronic shortage of parks, playgrounds, and modern school amenities. The university at Ca' Foscari, a modern foundation, has highly reputable schools of architecture and planning. The renovated State Archives is an international centre of scholarship, its documentary collections covering a thousand years of the Venetian republic. It is complemented by other scholarly centres such as the Old Library, the Correr Civic Museum, and the Cini Foundation.

Although planning regulations severely restrict alterations to buildings, there are examples of modern additions and structures. The fire station of Foscari Canal makes use of the traditional architectural vocabulary; other buildings, such as the headquarters of the Venetian Savings Bank in the Campo Manin, are uncompromisingly modern.

A growing problem for Venice today is the loss of population from the city core. Of some 340,000 residents,

Population loss

fewer than 100,000 still live in the historic centre. Faced with poor social amenities and old, decaying, often damp buildings with rents inflated by the costs of renovation, demands of the tourist industry, and wealthy foreign residents, Venetians have elected in ever-increasing numbers to move into modern apartments in the mainland boroughs of Mestre and Marghera or on the Lido. This exodus has produced a daily commuting problem and left the city of Venice with a smaller resident population than many of its formerly subject towns. It threatens to turn Venice into a museum city—a glorious spectacle whose architectural and artistic heritage is preserved, as it should be, but whose daily life is almost a parody of the vital unity of commerce, piety, politics, and ritual that was the pride of *la serenissima.* (D.E.C.)

History

THE EARLY PERIOD

Origin of the city. Uniquely among Italy's chief cities, Venice came into being after the fall of the Roman Empire in the West. The Lombard hordes, whose incursions into northern Italy began in AD 568, drove great numbers of mainlanders onto the islands of the lagoon, previously the homes of itinerant fishermen and salt workers. The isolated communities, literally islands of Veneto-Byzantine civilization, became part of the Exarchate of Ravenna when it was created in 584. When the mainland Byzantine city of Oderzo fell to the Lombards in 641, political authority was shifted to one of the islands in the Venetian lagoon.

The first elected doge, or duke, was Orso, chosen in an anti-Byzantine military declaration in 727, but he was succeeded by Byzantine officials until about 751, when the Exarchate of Ravenna came to an end. There followed decades of internal political strife among various settlements vying for supremacy and between pro- and anti-Byzantine factions; also involved were attempts by church authorities to acquire temporal influence. Finally the doge Obelerio and his brother Beato formed an alliance with the Franks of Italy and placed Venice under the authority of the Italian king Pepin (died 810), in order to free themselves from Byzantine control.

Pro-Byzantine reaction to this event under the doges of the Parteciaco family led to the transfer of the seat of government to the Rialto group of islands, by then the centre for exiles in the factional fighting. Though a Franco-Byzantine treaty of 814 guaranteed to Venice political and juridical independence from the rule of the Western Empire, it did not confirm any effective dependence on the Byzantine Empire, and by 840–841 the doge was negotiating international agreements in his own name. The unusual legal and political position of the small independent situated duchy, situated in territorial isolation between two great empires, contributed greatly to its function as a trading intermediary.

Political independence

A long succession of serious disputes between leading families concerning the office of doge did not halt the rapid development of trade. Increase in private wealth led to the gradual achievement of internal stability by creating a broader ruling class that was capable of putting a limit to the power of the doge. Gradually a national consciousness developed. Beginning in the late 9th century, the doges were chosen by popular election, though the right was frequently abused during times of civil strife. Finally the group of Rialto islands was solemnly transformed into the city of Venice (*civitas Venetiarum*).

The new order. The final collapse of family faction rule led to a change in the system of government, inaugurated by Doge Domenico Flabanico (1032–42). He restored to the people the sovereign right to elect the doge, but the term *populus* was in practice restricted to the residents of the Rialto and, more narrowly, to a select group of nobles. The executive organ was the ducal curia, and the legislative assembly was summoned to approve the doge's acts. A new church was built for St. Mark, symbol of the Venetian spirit, under Doge Domenico Contarini (1043–70), an energetic defender of the religious independence of the duchy.

GROWTH OF TRADE AND POWER

In the conflict between papacy and empire, Contarini and his successors remained neutral (despite the complaints of Pope Gregory VII), while safeguarding Venetian economic interests in the Adriatic when the conflict began to be reflected on the Dalmatian coast. But the greatest danger to Venetian interests was the 11th-century Norman expansion under Robert Guiscard, which threatened to cut Venetian communications to the south. The successful action taken against the Normans by Doge Domenico Selvo and his successor Vitale Falier helped to assure Venetian freedom on the Mediterranean Sea.

Relations with the Byzantine Empire. In gratitude for Venetian aid against the Normans, the Byzantine emperor Alexius I Comnenus granted Venice unrestricted trade throughout the Byzantine Empire, with no customs dues, a privilege that marked the beginning of Venetian activity in the East (1082). The Adriatic was not yet secured, however; Dalmatian ports were threatened by the Hungarians and Slavs, with whom it was difficult to come to agreement.

Toward the end of the 11th century the Crusades focused the newly awakened trading interests of the West on the Mediterranean. At first Venice was chiefly concerned with gaining control of the European trading ports of the Byzantine Empire, leaving to private interests the commercial opportunities in Syria and Asia Minor. Although they had been the first to win trade concessions and a commercial quarter in Constantinople, the Venetians antagonized the Byzantines by their arrogance and lawlessness as well as by their superior enterprise. In helping the emperor Manuel I Comnenus drive the Normans out of Corfu (1147–49), they offended him by their aggressive behaviour.

Enmity of the emperor

Soon the mutual dislike between Venetians and Byzantines ripened into hatred. The emperor encouraged merchants from the Italian republics of Genoa and Pisa to compete in Byzantine markets, and the Venetians responded by destroying the establishments of their rivals. In 1171, to maintain order in his dominions, the emperor arrested all Venetian residents in Constantinople and the provinces and confiscated their goods. Relations were patched up in 1187 and again in 1198, but the Venetians remained embittered.

The commune. All this time the expansion of Venice along the borders of the lagoon and across the Adriatic on the Dalmatian coast not only enriched its patrimony but also created an awareness of its own political power. Between 1140 and 1160, in the response to the needs of its increased territory and growing economy, Venice underwent a revolutionary change in its political structure, reorganizing itself as a republic. The doge lost his monarchic character, becoming a mere official (though he still assumed resounding titles), and a commune took over the powers, functions, and prerogatives of the state. All political and administrative matters were placed in the hands of the Great Council of 45. A Minor Council of six members exercised executive powers alongside the doge, and magistrates were granted administrative and judicial functions.

Trade conflicts. Venetian bitterness against the Byzantines found an outlet in the Fourth Crusade, which captured and sacked Constantinople in 1204 with the doge Enrico Dandolo among its leaders. In the subsequent partition of Byzantine territory between Venetians and crusaders, Venice acquired a commercial empire in the eastern Mediterranean. It included many of the Aegean islands, most importantly Crete and parts of Euboea, with valuable trading stations and fortified lookout posts on the Greek mainland. The doges adopted the title of Lord of One-Quarter and One-Eighth of the Entire Byzantine Empire (*Quartae Partis et Dimidiae Totius Imperii Romaniae Dominator*). A special magistrate, appointed from Venice, administered the substantial Venetian colony in Constantinople.

In 1261 the Byzantine emperor in exile at Nicaea, with the support of the Genoese, recovered the city and evicted the Venetians. The emperor rewarded the Genoese with privileges that challenged the Venetian monopoly of trade and opened up to Genoa the Black Sea markets. The Venetians retained control of many of the Greek islands,

however, and gradually found their way back to partial favour in Byzantium through a series of treaties. But when the last of the crusader strongholds in Syria fell to the Muslims in 1291, Venetian merchants who had been dispossessed moved north to dispute the Black Sea trade with the Genoese. For nearly two centuries thereafter, Venice and Genoa were periodically at war.

The patriciate. Meanwhile, at home the Venetian state was being built up. In 1242 the civil statutes of Jacopo Tiepolo regulated civil and economic relations; maritime statutes had been established in 1239. The number of elected members of the Great Council was raised from 45 to 60 and then to 100. The council of 40 (first mentioned in 1223) received powers of jurisdiction, and the Consiglio dei Rogati (60 members; founded mid-13th century), invested with the control of economic affairs, in time assumed all legislative functions and the honorific title of Senate.

Rivalry of family factions

In the 11th and 12th centuries the Michiel and Falier families had tried in vain to perpetuate their ducal power, and restrictive electoral systems were instituted to prevent the formation of committed family factions. In the 13th century similar attempts by the Ziani and Tiepolo families also failed. In 1268 an interlocking process of choice by lot and voting alternately among the members of the Great Council was introduced to select the next doge.

Between 1290 and 1300, new laws restricted the right to take part in the government to families traditionally performing magistrate's duties. The patrician class was not created by the "closing of the Great Council" (*serrata del Maggior Consiglio*) achieved by these laws, but it received its legal status from them. Henceforward anyone claiming personal power had to act outside the patrician order and rely on the people; and the people were linked so closely to the patricians by their economic needs that sufficient support was always lacking. Thus the conspiracy of Marin Bocconio failed (1299), as did those of Bajamonte Tiepolo and the Querini brothers (1310) and later of Marin Falier (1354). The special character of Venetian society created a governing class very different from that of the other Italian communes or of the continental states. To counter any attempts at sole personal rule, the Council of Ten was established (1310) to police the patrician order and defend the existing regime.

Struggle for naval supremacy. By the beginning of the 14th century the republic was swept into struggles on the mainland of Italy and in the Adriatic and Mediterranean seas. When the Scaligeri came to power in Verona, the republic made alliance with the Carraresi of Padua, with the Florentines, and with the Visconti of Milan, who feared the rise of a strong territorial lordship in the heart of northern Italy. Deviating from its strictly maritime policy, Venice established sovereignty over Treviso, thereby ensuring its own food supply but also taking on the defense of a land frontier.

The antagonism and rivalry with Genoa were rekindled. The conflict, carried on mainly in Dalmatia, was made more difficult for all by the spread of the Black Death (1348), by the economic and financial crisis caused by the war itself, and by the ineptitude of the military operations. In the alternation of victories and defeats, both sides exhausted their energies and resources. At last a second anti-Venetian coalition brought the war almost into Venice itself; at Pula (Pola) and at Chioggia, Venice first was defeated and then won the victory (1380–81). The Peace of Turin (1381) eliminated Genoese political influence from the Mediterranean and the East, leaving the Venetian government as arbiter of the sea routes.

Zenith of power. The Venetian victory over Genoa took place under the threat of Turkish advance in the East. The Venetians had to negotiate a state of neutrality with the Turks and find another economic base to compensate for the smaller yield now to be expected from trade with the East. So they turned to the Italian mainland, first to rid themselves of neighbouring lordships and then to

Territorial acquisitions

defend and exploit the rich lands they had acquired. For a time Venetian territorial rule went no further than the Mincio and Livenza rivers, but beyond the Livenza lay the politically and economically important principality of the patriarch of Aquileia, through which passed the main routes to Germany and to Istria. Because the patriarch could not guarantee peace and order, Venice incorporated the principality in the Venetian domains (1420).

Venetian territory now covered roughly the areas of the modern regions of Veneto and Friuli-Venezia Giulia, together with the Istrian Peninsula. The doge Tommaso Mocenigo maintained that his city had reached its political and economic zenith; it had a solid base in Italy that could compensate for its losses in the East, and it should not expect indefinite progress. In fact, efforts to enlarge its conquests might be dangerous, and it was better to preserve, not to risk, its accumulated wealth. Mocenigo's successors, however, did not heed his warning.

POLITICAL AND ECONOMIC DECLINE

When he became Venice's doge in 1423, Francesco Foscari embarked upon a series of wars in mainland Italy, particularly against Milan. Greed for conquering new territory involved the Venetians in the tangled web of Italian balance-of-power politics and in conflicts between the great powers of Europe on the scale out of proportion to Venetian forces and direct interests. The Peace of Lodi (1454) was followed by the formation of the Italian League to restore political balance among the Italian states, but the accord was ephemeral and Italy was threatened with foreign intervention.

Meanwhile, the Turks were encroaching upon the Byzantine Empire in the East; Thessalonica fell in 1430 and Constantinople in 1453. Further Turkish moves prompted Venice to defend its eastern territories, but in 1470 Euboea fell into Turkish hands. Peace with the Turks was finally achieved in 1479. The Venetians, however, soon became involved in another war, this time with Ferrara. Venice's conquest of the Polesine region (1484) increased the opposition of the other Italian states to Venetian territorial expansion.

Europeans and Turks against Venice. This internal discord made Italy a prey to invading foreigners, Spanish, French, and German. By 1508 these powers, together with the pope, the Hungarians, the Savoyards, and the Ferrarese, united to form the League of Cambrai against the Venetians, who were defeated at the Battle of Agnadello. Venice was saved from the worst results of this event by internal discords within the League of Cambrai, but Venetian territories on the mainland were diminished. At the same time, the republic was experiencing an economic crisis. Not only was the Eastern market lost but the discovery of new lands in the West and new trade routes to the East released Europe from dependence on Venetian merchants. Venice ceased to be a Mediterranean power, and, as a European power, it lacked the advantage that the Atlantic countries had of direct access to the New World.

Venetian policy in the 16th century was dictated by the need to keep intact its political, economic, and territorial heritage against the advance of the Turks on the one side and the pressure of the great western European powers on the other. This need supplied the reason for Venice's intervention in the Italian crisis of the emperor Charles V; for its struggle against the Turks, from the defeat of Préveza in 1538 to the victory of Lepanto and the loss of Cyprus in 1571; and for its tenacious resistance to pressure from the pope. So Venice declined into economic stagnation, embittered by a constitutional conflict between the Consiglio dei Rogati and the Council of Ten for control of the public finances. Venetian peace and neutrality meant defending the immediate interests of the nation but ceasing to take part in problems in which it was not directly concerned. Thus the spirit of political and religious conservatism grew increasingly tenacious in Venice.

Opposition from the pope

A political crisis was created by the papal interdict of Venice in 1606, concerned not with heresy or reform but with temporal prerogatives of the papacy. Paolo Sarpi, the energetic defender of Doge Leonardo Doná's policy, which had provoked the Roman Curia, never contested the legitimacy of papal power, but in the temporal sphere he denied that it carried any prerogatives superior to the sovereign rights of the state.

After a long campaign (1645–69) Crete, Venice's last

possession in the eastern Mediterranean, fell to the Turks—the Venetians being allowed to retain only a few strongholds. This blow to morale was mitigated, however, by the preservation of Dalmatia, and the government, after allying itself with Austria, attempted to reestablish itself in the eastern Mediterranean by liberating the Morea (Peloponnese) from the Turks. There the brilliant campaign of Francesco Morosini in 1684–88 assured Venice of this new Greek territory, which was finally handed over in 1699. But the conquest proved profitless and became an expensive burden, and in 1718 the Morea was returned to the Turks. Thus ended Venetian activity in the eastern and southern Mediterranean, save for an unsuccessful attempt in 1769 on Algerian and Tunisian pirates under Angelo Emo.

End of the Venetian republic. During its later years the Venetian republic was estranged from the fervour of new ideas germinating in other nations. Venetian life had crystallized inescapably. The plans of Angelo Querini, Giorgio Pisani, and Carlo Contarini, who in the 18th century called themselves reformers, did not go beyond those of the noble class that for three centuries had controlled the government and that existed to uphold ancestral tradition or to satisfy personal ambition.

The end of the republic came after the outbreak of the French Revolution. Napoleon, determined to destroy the Venetian oligarchy, claimed as a pretext that Venice was hostile to him and a menace to his line of retreat during his Austrian campaign of 1797. The Peace of Leoben left Venice without an ally, and Ludovico Manin, the last doge, was deposed on May 12, 1797. A provisional democratic municipality was set up in place of the republican government, but later in the same year Venice was handed over to Austria.

In 1848 the revolutionary leader Daniele Manin set up a provisional republican government, but it fell the following year. After the defeat of Austria by the Prussians in 1866, Venice was ceded to Italy, which had been a united kingdom since 1861.

AFTER INCORPORATION INTO ITALY

The subsequent growth of Venice was attendant upon its role in the commercial life of Italy and upon exploitation of its inherent physical and aesthetic attributes. The city had lost a bit of its island character and some of its insular mentality in 1846, when a causeway nearly two miles (three kilometres) in length brought the railway across 222 arches from the mainland. It lost even more in 1932 when a parallel road was built to give access to motor vehicles. Each link was stoutly resisted by persons who wished to leave the city unchanged, and they succeeded in forcing wheeled vehicles to be garaged at the landward edge of the island. Similar battles continue between traditionalists and modernists.

The environmental crisis

The most urgent problems confronting present-day Venice, however, are environmental. Since the 1950s the deterioration of ancient buildings and art treasures, which had long been associated with natural phenomena such as flooding and subsidence, has been intensified by an atmosphere laden with sulfuric acid, much of it generated by industrial and domestic smoke. As in other cities in which monuments and works of art stand exposed, air pollution has already corroded and defaced many priceless examples of stonework from the Venetian past. Severely damaging storms and floods in November 1966 stimulated increased efforts to rescue the historic city from environmental destruction. (R.Ce./Ed.)

BIBLIOGRAPHY

General: Views and descriptions of the modern city are found in many well-illustrated guides: HUGH HONOUR, *The Companion Guide to Venice* (1983); GUIDO ALBERTO ROSSI and FRANCO MASIERO, *Venice from the Air* (1988); FULVIO ROITER and OLIVIER BERNIER, *Venice II* (1985); and ALTA MACADAM, *Venice,* 3rd ed. (1986), in the *Blue Guide Series.* Life in Venice is surveyed in SHIRLEY GUITON, *A World by Itself: Tradition and Change in the Venetian Lagoon* (1977); JAMES MORRIS, *Venice,* 2nd rev. ed. (1983); ALESSANDRO SAVELLA and NANTAS SALVALAGGIO, *The Carnival of Venice,* trans. from Italian (1984); GEORGE BULL, *Venice: The Most Triumphant City* (1980); and CHRISTOPHER HIBBERT, *Venice: the Biography of a City* (1988). Observations and reflections on Venice can also be found in a number of works by individuals who lived or traveled in Venice; these include chapters of GIACOMO CASANOVA, *History of My Life,* 12 vol. in 6 (1966–71; originally published in French, 1826–38); HORATIO F. BROWN, *Life on the Lagoons* (1884), and *Venetian Studies* (1887); W.D. HOWELLS, *Venetian Life,* rev. ed. (1907), written after his posting as U.S. consul to Venice in 1860–65; CHARLES DICKENS, *Pictures from Italy* (1846), available in many later editions; HENRY JAMES, *Italian Hours* (1909, reprinted 1987); and MARY MCCARTHY, *Venice Observed* (1956).

History: JOHN JULIUS NORWICH, *Venice,* 2 vol. (1977–81; also published as *A History of Venice,* 1982), treats the history of the city from its beginnings to 1797. Another comprehensive survey of the same period is W. CAREW HAZLITT, *The Venetian Republic: Its Rise, Its Growth, and Its Fall, A.D. 409–1797,* 2 vol., 4th ed. (1915, reprinted 1966). See also ROBERTO CESSI, *Storia della repubblica di Venezia,* 2 vol. (1944–46, reissued in 1 vol., 1981), a history by Venice's foremost historian; and WILLIAM H. MCNEILL, *Venice: The Hinge of Europe, 1081–1797* (1974, reprinted 1986). FREDDY THIRIET, *La Romanie vénitienne au Moyen Age* (1959; reprinted with additions, 1975), describes the Venetian commercial empire of the 12th–15th century that was founded by the Fourth Crusade. Other historical studies include GUIDO RUGGIERO, *Violence in Early Renaissance Venice* (1980); ALBERTO TENENTI, *Piracy and the Decline of Venice, 1580–1615* (1967; originally published in Italian); DONALD E. QUELLER, *The Venetian Patriciate: Reality Versus Myth* (1986); DENNIS ROMANO, *Patricians and Popolani: The Social Foundations of the Venetian Renaissance State* (1987); MARGARET L. KING, *Venetian Humanism in an Age of Patrician Dominance* (1986); D.S. CHAMBERS, *The Imperial Age of Venice, 1380–1580* (1970); EDWARD MUIR, *Civic Ritual in Renaissance Venice* (1981), on the role of public display in the life of the republic; M.E. MALLETT and J.R. HALE, *The Military Organization of a Renaissance State: Venice, c. 1400 to 1617* (1984), on the armed forces and power politics; ROBERT FINLAY, *Politics in Renaissance Venice* (1980), on the complex political structure that was in place at the height of the Republic; and WILLIAM J. BOUWSMA, *Venice and the Defense of Republican Liberty: Renaissance Values in the Age of the Counter Reformation* (1968, reprinted 1984), on the evolution and political uses of the "myth of Venice."

Histories of the economic and social conditions and commerce include FREDERIC C. LANE, *Venice: A Maritime Republic* (1973); FREDERIC C. LANE and REINHOLD C. MUELLER, *Money and Banking in Medieval and Renaissance Venice: Coins and Moneys of Account* (1985); PETER LAURITZEN, *Venice: A Thousand Years of Culture and Civilization* (1978); and a collection of earlier works by FREDERIC C. LANE, *Studies in Venetian Social and Economic History,* ed. by BENJAMIN G. KOHL and REINHOLD C. MUELLER (1987). See also JAMES C. DAVIS, *A Venetian Family and Its Fortune, 1500–1900* (1975); SUSAN CONNELL, *The Employment of Sculptors and Stonemasons in Venice in the Fifteenth Century* (1988); RICHARD J. GOY, *Chioggia and the Villages of the Venetian Lagoon* (1985); and RICCARDO CALIMANI, *The Ghetto of Venice* (1987; originally published in Italian, 1985).

Arts and architecture: JOHN RUSKIN, *The Stones of Venice,* 3 vol. (1851–53; reprinted 1979), is a major historical study. DEBORAH HOWARD, *The Architectural History of Venice* (1980), provides a detailed treatment of Venetian buildings and their social context. See also MICHELANGELO MURARO, *Venetian Villas: The History and Culture,* trans. from Italian (1986); PETER LAURITZEN and ALEXANDER ZIELCKE, *Palaces of Venice* (1978, reissued 1985); DOUGLAS LEWIS, *The Late Baroque Churches of Venice* (1979); and RICHARD J. GOY, *Venetian Vernacular Architecture: Traditional Housing in the Venetian Lagoon* (1988). Developments of the Renaissance period are studied in JOHN MCANDREW, *Venetian Architecture of the Early Renaissance* (1980); RALPH LIEBERMAN, *Renaissance Architecture in Venice, 1450–1540* (1982); and BERNHARD BERENSON, *The Venetian Painters of the Renaissance,* 3rd ed. (1897). Essays on many aspects of Venetian art and politics of the Renaissance are collected in an illustrated catalog of a major exhibition, JANE MARTINEAU and CHARLES HOPE (eds.), *The Genius of Venice, 1500–1600* (1983). Other special studies include SIMON TOWNELEY WORSTHORNE, *Venetian Opera in the Seventeenth Century* (1954, reprinted 1984); MICHAEL LEVEY, *Painting in Eighteenth-Century Venice,* 2nd rev. ed. (1980); and PETER LAURITZEN, *Venice Preserved* (1986), a concise survey of restoration work.

(D.E.C.)

Victoria and the Victorian Age

Victoria served as the queen of the United Kingdom of Great Britain and Ireland from 1837 until 1901. In 1876 she also became empress of India. During her reign the English monarchy took on its modern ceremonial character.

Victoria first learned of her future role as a young princess during a history lesson when she was 10 years old. Almost four decades later Victoria's governess recalled that the future queen reacted to the discovery by declaring, "I will be good." This combination of earnestness and egotism marked Victoria as a child of the age that bears her name. The queen, however, rejected important Victorian values and developments. A mother of nine who hated pregnancy and childbirth, detested babies, and was uncomfortable in the presence of children, Victoria reigned in a society that idealized both motherhood and the family. She had no interest in social issues, yet the 19th century in Britain was an age of reform. She resisted technological change even while mechanical and technological innovations reshaped the face of European civilization.

Most significantly, Victoria was a queen determined to retain political power; yet unwillingly and unwittingly she presided over the transformation of the sovereign's political role into a ceremonial one and thus preserved the English monarchy. When Victoria became queen, the political role of the crown was by no means clear; nor was the permanence of the throne itself. When she died and her son Edward VII moved from Marlborough House to Buckingham Palace, the change was one of social rather than of political focus; there was no doubt about the monarchy's continuance. That was the measure of her reign.

Transformation of the sovereign's role

Lineage and early life. On the death in 1817 of Princess Charlotte, daughter of the prince regent (later George IV), there was no surviving legitimate offspring of George III's 15 children. In 1818, therefore, three of his sons, the dukes of Clarence, Kent, and Cambridge, married to provide for the succession. The winner in the race to father the next ruler of Britain was Edward, duke of Kent, fourth son of King George III. His only child, christened Alexandrina Victoria, was born at Kensington Palace, London, on May 24, 1819. After his death and George IV's accession in 1820, Victoria became third in the line of succession to the throne after the Duke of York (died 1827) and the Duke of Clarence (subsequently William IV), whose own children died in infancy.

EB Inc.

Queen Victoria.

Victoria, by her own account, "was brought up very simply," principally at Kensington Palace, where her closest companions, other than her German-born mother, the Duchess of Kent, were her half sister, Féodore, and her governess, Louise (afterward the Baroness) Lehzen, a native of Coburg. An important father figure to the orphaned princess was her uncle Leopold, her mother's brother, who lived at Claremont, near Esher, Surrey, until he became king of the Belgians in 1831.

Victoria's childhood was made increasingly unhappy by the machinations of the Duchess of Kent's advisor, Sir John Conroy. In control of the pliable duchess, Conroy hoped to dominate the future queen of Britain as well. Persuaded by Conroy that the royal dukes, "the wicked uncles," posed a threat to her daughter, the duchess reared Victoria according to "the Kensington system," by which she and Conroy systematically isolated Victoria from her contemporaries and her father's family. Conroy thus aimed to make the princess dependent on and easily led by himself.

"The Kensington system"

Strong-willed, and supported by Lehzen, Victoria survived the Kensington system; when she ascended the throne in 1837, she did so alone. Her mother's actions had estranged her from Victoria and taught the future queen caution in her friendships. Moreover, her retentive memory did not allow her to forgive readily.

Accession to the throne. In the early hours of June 20, 1837, Victoria received a call from the archbishop of Canterbury and the lord chamberlain and learned of the death of William IV, third son of George III. Later that morning the Privy Council was impressed by the graceful assurance of the new queen's demeanour. She was small, carried herself well, and had a delightful silvery voice, which she retained all her life. The accession of a young woman was romantically popular. But because of the existence in Hanover of the Salic law, which prevented succession by a woman, the crowns of Great Britain and Hanover became separated, the latter passing to William IV's eldest surviving brother, Ernest, the unpopular duke of Cumberland.

The queen, who had never before had a room to herself, exiled her mother to a distant set of apartments when they moved into Buckingham Palace. Conroy was pensioned off. Only Lehzen, of whom Victoria was still in awe, remained close to the queen. Even her beloved uncle Leopold was politely warned off discussions of English politics. "Alone" at last, she enjoyed her new-found freedom. "Victoria," wrote her cousin, Prince Albert of Saxe-Coburg-Gotha, who later married her,

> is said to be incredibly stubborn and her extreme obstinacy to be constantly at war with her good nature; she delights in Court ceremonies, etiquette and trivial formalities. . . . She is said not to take the slightest pleasure in nature and to enjoy sitting up at night and sleeping late into the day.

It was, in retrospect, "the least sensible and satisfactory time in her whole life"; but at the time it was exciting and enjoyable, the more so because of her romantic friendship with Lord Melbourne, the prime minister.

Melbourne was a crucial influence on Victoria, in many ways an unfortunate one. The urbane and sophisticated prime minister fostered the new queen's self-confidence and enthusiasm for her role; he also encouraged her to ignore or minimize social problems and to attribute all discontent and unrest to the activities of a small group

Lord Melbourne

of agitators. Moreover, because of Melbourne, Victoria became an ardent Whig.

Victoria's constitutionally dangerous political partisanship contributed to the first two crises of her reign, both of which broke in 1839. The Hastings affair began when Lady Flora Hastings, a maid of honour who was allied and connected to the Tories, was forced by Victoria to undergo a medical examination for suspected pregnancy. The gossip, when it was discovered that the queen had been mistaken, became the more damaging when later in the year Lady Flora died of a disease that had not been diagnosed by the examining physician. The enthusiasm of the populace over the coronation (June 28, 1838) swiftly dissipated.

Between the two phases of the Hastings case "the bedchamber crisis" intervened. When Melbourne resigned in May 1839, Sir Robert Peel, the Conservative leader, stipulated that the Whig ladies of the bedchamber should be removed. The queen imperiously refused, not without Melbourne's encouragement. "The Queen of England will not submit to such trickery," she said. Peel therefore declined to take office, which Melbourne rather weakly resumed. "I was very young then," wrote the queen long afterward, "and perhaps I should act differently if it was all to be done again."

The Albertine monarchy. Victoria's wedding to Prince Albert served as a stage for displays of political partisanship: very few Tories received invitations, and the Tories themselves rejected Victoria's request that Albert be granted rank and precedence second only to her own. Victoria responded violently, "Monsters! You Tories shall be punished. Revenge! Revenge!" Marriage to Albert, however, lessened the queen's enthusiasm for Melbourne and the Whigs. She admitted many years later regarding Melbourne that "Albert thinks I worked myself up to what really became rather foolish." Albert thus shifted Victoria's political sympathies; he also became the dominant figure and influence in her life. She quickly grew to depend on him for everything; soon she "didn't put on a gown or a bonnet if he didn't approve it." No more did Victoria rule alone.

Marriage to Albert

Attracted by Albert's good looks and encouraged by her uncle Leopold, Victoria proposed to her cousin on Oct. 15, 1839, just five days after he had arrived at Windsor on a visit to the English court. She described her impressions of him in the journal she kept throughout her life: "Albert really is quite charming, and so extremely handsome . . . a beautiful figure, broad in the shoulders and a fine waist; my heart is quite *going*." They were married on Feb. 10, 1840, the queen dressed entirely in articles of British manufacture.

Children quickly followed. Victoria, the princess royal (the "Vicky" of the *Letters*), was born in 1840; in 1858 she married the crown prince of Prussia and later became the mother of the emperor William II. The Prince of Wales (later Edward VII) was born in 1841. Then followed Princess Alice, afterward grand duchess of Hesse, 1843; Prince Alfred, afterward duke of Edinburgh and duke of Saxe-Coburg-Gotha, 1844; Princess Helena (Princess Christian of Schleswig-Holstein), 1846; Princess Louise (duchess of Argyll), 1848; Prince Arthur (duke of Connaught), 1850; Prince Leopold (duke of Albany), 1853; and Princess Beatrice (Princess Henry of Battenberg), 1857. The queen's first grandchild was born in 1859, and her first great-grandchild in 1879. There were 37 great-grandchildren alive at her death.

Victoria never lost her early passion for Albert: "Without him everything loses its interest." Despite conflicts produced by the queen's uncontrollable temper and recurrent fits of depression, which usually occurred during and after pregnancy, the couple had a happy marriage. Victoria, however, was never reconciled to the childbearing that accompanied her marital bliss—the "shadow-side of marriage," as she called it. Victoria explained to her eldest daughter in 1858:

> What you say of the pride of giving life to an immortal soul is very fine, dear, but I own I cannot enter into that; I think much more of our being like a cow or a dog at such moments; when our poor nature becomes so very animal and unecstatic.

At the beginning of their marriage the queen was insistent that her husband should have no share in the government of the country. Within six months, on Melbourne's repeated suggestion, the prince was allowed to start seeing the dispatches, then to be present when the queen saw her ministers. The concession became a routine, and during her first pregnancy the prince received a "key to the secret boxes." As one unwanted pregnancy followed another and as Victoria became increasingly dependent on her husband, Albert assumed an ever-larger political role. By 1845 Charles Greville, the observer of royal affairs, could write, "it is obvious that while she has the title, he is really discharging the functions of the Sovereign. He is the King to all intents and purposes." Victoria, once so enthusiastic about her role, came to conclude that "we women are not *made* for governing."

Relations with Peel

The prince came into his own to negotiate with Peel a compromise on the bedchamber question after the Melbourne government had been defeated in the general election of 1841. The queen's first interview with Peel went well, eased by Melbourne's advice to his successor:

> The Queen is not conceited—she is aware there are many things she cannot understand and she likes to have them explained to her elementarily—not at length and in detail but shortly and clearly.

If, as Lady Lyon once noted, "there was 'a vein of iron' which ran through the Queen's extraordinary character," the iron could bend; Victoria was able to revise her opinions and reevaluate her judgments. Peel's very real distress when in the summer of 1842 an attempt was made to assassinate the queen—together with the affinity between the prince and the new prime minister—soon converted the "cold odd man" of the queen's earlier comment into "a great statesman, a man who thinks but little of party and never of himself." Lord Aberdeen, the foreign secretary, also became a great favourite. "We felt so safe with them both," she told King Leopold.

The departure of the possessive Lehzen for Germany in 1842 signaled Albert's victory in the battle between the two for Victoria's loyalty and for power in the royal household. He became effectively the queen's private secretary—according to himself, "her permanent minister." As a result of Albert's diligence and refusal to accept the obstacles that ministers threw in his path, the management of the queen's properties was rationalized and her income thus increased.

A visible sign of the prince's power and influence was the building of the royal residences of Osborne, on the Isle of Wight, and Balmoral Castle in Scotland. Albert, who taught the once party-loving Victoria to despise London, played a central role in the acquisition of both properties as well as in designing the homes he and Victoria erected on them between 1845 and 1855.

Preference for Balmoral

Victoria described Osborne as "our island home" and retreated there frequently; it was, however, at Balmoral that she was happiest. The royal pair and their family were able to live there "with the greatest simplicity and ease," wrote Greville. The queen soon came to hold the Highlanders in more esteem than she held any other of her subjects. She liked the simpler life of the Highlands, as her published journal was to reveal: she came to make the most of the thin stream of Scottish blood in her veins; also, so long as the sermons were short enough, she came to prefer the Scottish form of religious service. "You know," she was to tell her prime minister, William Ewart Gladstone, "I am not much of an Episcopalian"; and she developed a comfort in the consolations of the Reverend Norman Macleod and also a delight in the plain speech of John Brown, the Highland servant who stalked with Albert and became her personal attendant.

The royal couple's withdrawal to Scotland and the Isle of Wight bore witness to a new sort of British monarchy. In their quest for privacy and intimacy Albert and Victoria adopted a way of life that mirrored that of their middle-class subjects, admittedly on a grander scale. Although Albert was interested in intellectual and scientific matters, Victoria's tastes were closer to those of most of her people. She enjoyed the novels of Charles Dickens and patronized the circus and waxwork exhibitions. Both Victoria and

Albert, however, differed from many in the middle class in their shared preference for nudes in painting and sculpture. Victoria was not the prude that many claimed her to be. She was also no Sabbatarian: "I am not at all an admirer or approver of our very dull Sunday."

Victoria's delight in mingling with the Scottish poor at Balmoral did little to raise the level of her social awareness. Although in 1846 she and Albert supported the repeal of the Corn Laws (protectionist legislation that kept the price of British grain artificially high) in order to relieve distress in famine-devastated Ireland, they remained much more interested in and involved with the building of Osborne and foreign policy than in the tragedy of Ireland. Victoria, moreover, gave her full support to the government's policy of repression of the Chartists (advocates of far-reaching political and social reform) and believed the workers in her realm to be contented and loyal. In 1848, rejoicing in the failure of the last great Chartist demonstration in London, the queen wrote:

> The loyalty of the people at large has been very striking and their indignation at their peace being interfered with by such worthless and wanton men—immense.

The consequences of continental revolutions led her to conclude:

> Revolutions are always bad for the country, and the cause of untold misery to the people. Obedience to the laws and to the Sovereign, is obedience to a higher Power, divinely instituted for the good of the *people,* not the Sovereign, who has equally duties and obligations.

Yet, revolution or no revolution, many of her people lived in "untold misery," a fact Victoria rarely confronted.

The Great Exhibition

For both the queen and the prince consort the highlight of their reign came in 1851, with the opening of the Great Exhibition. Albert poured himself into the task of organizing the international trade show that became a symbol of the Victorian Age. Housed in the architectural marvel of the Crystal Palace, a splendid, greenhouse-inspired glass building erected in Hyde Park, the Great Exhibition displayed Britain's wealth and technological achievements to a wondering world. To Victoria the success of the Great Exhibition provided further evidence of her husband's genius: "I *do* feel proud at the thought of what my beloved Albert's great mind has conceived." Profits from the Great Exhibition funded what became the South Kensington complex of colleges and museums.

Albert has been credited with teaching Victoria the importance of remaining above party. Certainly he saw the danger in the Whig partisanship she openly displayed before their marriage; more clearly than Victoria he realized the fine sense of balance required of a constitutional monarch. Albert's own actions, however, such as his much-criticized appearance in the gallery of the House of Commons during Peel's speech on the first day of the Corn Laws debates (and thus his open and partisan show of support for Peel), revealed his political sympathies. Gladstone noted in 1846 that

> the Prince is very strongly Conservative in his politics and his influence with the Q. is over-ruling; through him she has become so attached to Conservative ideas that she could hardly endure the idea of the opposite Party as her ministers.

Like the queen, Albert believed that the sovereign had an important and active role to play in British politics. The fluid political situation operating during the prince's lifetime made such an active role seem possible. After the repeal of the Corn Laws (1846) there was a period, not ending until the election of 1868, when politics tended to consist of a series of temporary alliances between splinter groups and no single group could guarantee its extended control over the House of Commons: the golden age of the private member, a condition rendering active political intervention by the crown not only possible but sometimes even necessary. There was a role for the cabinet maker, especially in helping to compose coalitions. Its significance must not, however, be overemphasized; although Victoria probably would not have admitted it, the queen's role, albeit "substantial," was always "secondary."

Foreign affairs

The tradition also persisted that the sovereign had a special part to play in foreign affairs and could conduct them alone with a secretary of state. Victoria and Albert had relatives throughout Europe and were to have more. Moreover, they visited and were visited by other monarchs. Albert was determined that this personal intelligence should not be disregarded and that the queen should never become (as his own mentor the Baron Stockmar had indicated) "a mandarin figure which has to nod its head in assent or shake it in denial as its Minister pleases." The result was a clash with Lord Palmerston, the foreign secretary, who could look back on a career of high office beginning before the royal couple was born. The prince distrusted Palmerston's character, disapproved of his methods, thought his policy shallow, and disagreed with his concept of the constitution.

Even after Victoria insisted to Palmerston in 1850, "having once given her sanction to a measure, that it be not arbitrarily altered or modified by the minister," the foreign secretary continued to follow policies disapproved of by both Albert and Victoria, such as his encouragement of nationalist movements that threatened to dismember the Austrian Empire. Finally, after Palmerston expressed his approval of the coup d'état of Louis Napoleon (later Napoleon III) in 1851 without consulting the queen, the prime minister, Lord John Russell, dismissed him. Within a few months the immensely popular Palmerston was back in office, however, as home secretary. He would serve twice as prime minister. After Albert's death Victoria's disapproval of Palmerston diminished; his conservative domestic policy and his insistence that Britain receive its due in world affairs accorded with her own later views.

Crimean War

On the eve of the Crimean War (1854–56) the royal pair encountered a wave of unpopularity, and Albert was suspected, without any foundation, of trying to influence the government in favour of the Russian cause. There was, however, a marked revival of royalist sentiment as the war wore on. The queen personally superintended the committees of ladies who organized relief for the wounded and eagerly seconded the efforts of Florence Nightingale: she visited crippled soldiers in the hospitals and instituted the Victoria Cross for gallantry.

With the death of Prince Albert on Dec. 14, 1861, the Albertine monarchy came to an end. Albert's influence on the queen was lasting. He had changed her personal habits and her political sympathies. From him she had received training in orderly ways of business, in hard work, in the expectation of royal intervention in ministry making at home, and in the establishment of a private (because royal) intelligence service abroad. The English monarchy had changed. As the historian G.M. Young said, "In place of a definite but brittle prerogative it had acquired an undefinable but potent influence."

Widowhood. After Albert's death Victoria descended into deep depression—"those paroxysms of despair and yearning and longing and of daily, nightly longing to die . . . for the first three years never left me." Even after climbing out of depression, she remained in mourning and in partial retirement. She balked at performing the ceremonial functions expected of the monarch and withdrew to Balmoral and Osborne four months out of every year, heedless of the inconvenience and strain this imposed on ministers. After an initial period of respect and sympathy for the queen's grief, the public grew increasingly impatient with its absent sovereign. No one, however, could budge the stubborn Victoria.

Although Victoria resisted carrying out her ceremonial duties, she remained determined to retain an effective political role in the period after Albert's death and to behave as he would have ordained. Her testing point was, then, her "dear one's" point of view; and this she had known at a particular and thereafter not necessarily relevant period in English political life. Her training and his influence were ill suited to the "swing of the pendulum" politics that better party organization and a wider electorate enjoined after the Reform Bill of 1867. And since she blamed her son and heir for Albert's death—the prince consort had come back ill from Cambridge, where he had gone to see the Prince of Wales regarding an indiscretion the young prince had committed in Ireland—she did not hesitate to vent her loneliness upon him or to refuse him all responsibility. "It quite irritates me to see him in the room," she

startled Lord Clarendon by saying. The breach was never really healed, and as time went on the queen was clearly envious of the popularity of the Prince and Princess of Wales. She liked to be, but she took little trouble to see that she was, popular.

It was despite, yet because of, Albert that Victoria succumbed to Benjamin Disraeli and thus made herself a partisan in the most famous political rivalry of the 19th century. Albert had thought Disraeli insufficiently a gentleman and remembered his bitter attacks on Peel over the repeal of the Corn Laws in 1846; the prince, on the other hand, had approved of Gladstone, Disraeli's political rival. Yet Disraeli was able to enter into the queen's grief, flatter her, restore her self-confidence, and make the lonely crown an easier burden. Behind all his calculated attacks on her affections there was a bond of mutual loneliness, a note of mystery and romanticism, and, besides, the return to good gossip. Disraeli, moreover, told the queen in 1868 that it would be "his delight and duty, to render the transaction of affairs as easy to your Majesty, as possible." Since the queen was only too ready to consider herself overworked, this approach was especially successful. Gladstone, on the other hand, would never acknowledge that she was, as she put it, "dead beat," perhaps because he never was himself; Disraeli, however, tired easily. The contrast between Disraeli's gay, often malicious, gossipy letters and Gladstone's 40 sides of foolscap is obvious. And there was no Albert to give her a neat précis. Gladstone, moreover, held the throne as an institution in such awe that it affected his relations with its essentially feminine occupant. His "feeling" for the crown, said Lady Ponsonby, was "always snubbed." The queen had no patience with Gladstone's moralistic (and, she believed, hypocritical) approach to politics and foreign affairs. His persistent and often tactless attempts to persuade her to resume her ceremonial duties especially enraged her.

Disraeli

Over the problem of Ireland their paths separated ever more widely. Whereas "to pacify Ireland" had become the "mission" of Gladstone's life, the queen (like the majority of her subjects) had little understanding of, or sympathy for, Irish grievances. She disliked disorder and regarded the suggestion of Irish Home Rule as sheer disloyalty. The proposal of an Irish "Balmoral" was repugnant to her, especially when it was suggested that the Prince of Wales might go in her place. To avoid the Irish Sea, she claimed to be a bad sailor; yet she was willing in her later years to cross the English Channel almost every year. In all, she made but four visits to Ireland, the last in 1900 being provoked by her appreciation of the gallantry of the Irish regiments in the South African War.

Gladstone's defeat in 1874

The news of Gladstone's defeat in 1874 delighted the queen. "What an important turn the elections have taken," she wrote.

> It shows that the country is not *Radical.* What a triumph, too, Mr. Disraeli has obtained and what a good sign this large Conservative majority is of the state of the country, which really required (as formerly) a strong Conservative party!

If, years before, Melbourne, almost despite himself, had made her a good little Whig, and if Albert had left her, in general, a Peelite, temperamental and subsequently doctrinal differences with Gladstone helped make it easy for Disraeli to turn Victoria into a stout supporter of the Conservative Party.

One of the bonds shared by Victoria and Disraeli was a romantic attachment to the East and the idea of empire. Although she supported Disraeli's reform of the franchise in 1867, Victoria had little interest in or sympathy with his program of social reform; she was, however, entranced by his imperialism and by his assertive foreign policy. She applauded his brilliant maneuvering, which led to the British purchase of slightly less than half of the shares in the Suez Canal in 1875 (a move that prevented the canal from falling entirely under French control), especially since he presented the canal as a personal gift to her: "It is just settled; you have it, Ma'am." The addition of "Empress of India" in 1876 to the royal title thrilled the queen even more. Victoria and Disraeli also agreed on their answer to the vexing "Eastern question"—what was to be done with the declining Turkish empire? Even the revelation of Turkish atrocities against rebelling Bulgarians failed to sway the sovereign and her prime minister from their position that Britain's best interests lay in supporting Turkey, the "Sick Man" of Europe. The fact that Gladstone took the opposing view, of course, strengthened their pro-Turkish sympathies. With the outbreak of a Russo-Turkish war in 1877, however, Disraeli found himself in the uncomfortable position of having to restrain his bellicose sovereign, who demanded that Britain enter the war against Russia. At the Congress of Berlin in 1878 Disraeli emerged triumphant: Russian influence in the Balkans was reduced, and Britain gained control of the strategically located island of Cyprus. The queen was ecstatic.

Victoria's delight in Disraeli's premiership made further conflict with Gladstone inevitable. When in September 1879 a dissolution of Parliament seemed imminent, the queen wrote to the Marchioness of Ely (who was, after the Duchess of Argyll, perhaps her most intimate friend):

> Dear Janie, . . . I hope and trust the Government will be able to go on after the Election, as change is so disagreeable and so bad for the country; but if it should *not,* I wish the *principal* people of the Opposition should *know* there *are certain* things which *I never can* consent to. . . .
>
> I never COULD take Mr. Gladstone . . . as my Minister again, for I never could have the slightest *particle* of confidence in Mr. Gladstone *after* his violent, mischievous, and dangerous conduct for the last three years.

Conservative defeat in 1880

After the blow fell with the Conservative Party's defeat in 1880, Victoria sent for Lord Hartington.

> Mr. Gladstone *she* could have nothing to do with, for she considers his whole conduct since '76 to have been one series of violent, passionate invective against and abuse of Lord Beaconsfield, and that *he caused* the Russian war.

Nevertheless, as Hartington pointed out, it was Gladstone whom she had to have. She made no secret of her hostility, she hoped he would retire, and she remained in correspondence with Lord Beaconsfield (as Disraeli had become). Gladstone, indeed, said that he himself "would never be surprised to see her turn the Government out, after the manner of her uncles." The queen abhorred Gladstone's lack of Disraelian vision of Britain's role in the world. Over the abandonment of Kandahar in Afghanistan, in 1881, for example, Sir Henry Ponsonby had never seen her so angry: "The Queen has never before been treated," she told him, "with such want of respect and consideration in the forty three and a half years she has worn her thorny crown."

Victoria convinced herself that Gladstone's government, dominated (she believed) by Radicals, threatened the stability of the nation:

> No one is more truly Liberal in her heart than the Queen, but she has always strongly deprecated the great tendency of the present Government to encourage instead of checking the stream of destructive democracy which has become so alarming. . . . She will not be a Sovereign of a Democratic Monarchy.

Nevertheless, Victoria did act as an important mediating influence between the two houses to bring about the compromise that resulted in the third parliamentary Reform Act in 1884.

Victoria never acclimatized herself to the effects of the new electorate on party organization. No longer was the monarchy normally necessary as cabinet maker; yet, the queen was reluctant to accept her more limited role. Thus, in 1886 she sought to avoid a third Gladstone ministry by attempting to form an anti-Radical coalition. Her attempt failed. Irish Home Rule, not the queen, would defeat the "People's William."

Last years. In the Salisbury administration (1895–1902), with which her long reign ended, Victoria was eventually to find not only the sort of ministry with which she felt comfortable but one which lent a last ray of colour to her closing years by its alliance, through Joseph Chamberlain, with the mounting imperialism that she had so greatly enjoyed in Disraeli's day when he had made her empress of India.

The South African War (1899–1902) dominated her final years. The sufferings of her soldiers in South Africa aroused the queen to a level of activity and public visibil-

ity that she had avoided for decades. With a demanding schedule of troop inspections, medal ceremonies, and visits to military hospitals, Victoria finally became the exemplar of a modern monarch.

Victoria absorbed a great deal of the time of her ministers, especially Gladstone's, but after 1868 it may be doubted whether, save in rare instances, it made a great deal of difference. She may have postponed an occasional evil day; she certainly hampered an occasional career. And sometimes that "continuous political experience," which Walter Bagehot remarked as a long-lived monarch's greatest asset, was invaluable: in stopping "red tapings," as the queen called them, or in breaking a logjam. Meanwhile—"a comparatively late growth"—she had gained the affection of her subjects. The sheer endurance of her reign in a time of swift change deepened her symbolic value and hence heightened her popularity. Lord Salisbury observed in the House of Lords (Jan. 25, 1901) after her death that:

> She had an extraordinary knowledge of what her people would think—extraordinary, because it could not come from any personal intercourse. I have said for years that I have always felt that when I knew what the Queen thought, I knew pretty certainly what views her subjects would take, and especially the middle class of her subjects.

The queen, as the Jubilees of 1887 and 1897 showed, was popular. Gone were the days when pamphlets were circulated asking what she did with her money. More and more fully with advancing years, she was able to satisfy the imagination of the middle class—and the poorer class—of her subjects.

She remained, nevertheless, either aloof from or in opposition to many of the important political, social, and intellectual currents of the later Victorian period. She never reconciled herself to the advance of democracy, and she thought the idea of female suffrage anathema. The sufferings of an individual worker could engage her sympathy; the working class, however, remained outside her field of vision. After Albert's death Victoria had little contact with intellectual and artistic subjects and so remained happily unaware of the unsettling new directions being explored in the world around her. Her reign was shaped by the new technology—without the railroad and the telegraph, her extended stays in Osborne and Balmoral would have been impossible—yet she never welcomed innovation.

Character and personality

Many of the movements of the day passed the aged queen by, many irritated her, but the stupendous hard work that Albert had taught her went on—the meticulous examination of the boxes, the regular signature of the papers. To the very end Victoria remained a passionate and strong-willed woman.

Those who were nearest to her came completely under her spell; yet all from the Prince of Wales down stood in considerable awe. A breach of the rules could still make a fearsome change in the kindly, managing great-grandmother in black silk dress and white cap. The eyes would begin to protrude, the mouth to go down at the corners. Those who suffered her displeasure never forgot it, nor did she. Yielding to nobody else's comfort and keeping every anniversary, she lived surrounded by mementos, photographs, miniatures, busts, and souvenirs in chilly rooms at the end of drafty corridors, down which one tiptoed past Indian attendants to the presence. Nobody knocked; a gentle scratching on the door was all that she permitted. Every night at Windsor Albert's clothes were laid out on the bed, every morning fresh water was put in the basin in his room. She slept with a photograph—over her head—taken of his head and shoulders as he lay dead.

Queen Victoria had fought a long rearguard action against the growth of "democratic monarchy"; yet, in some ways, she had done more than anyone else to create it. She had made the monarchy respectable and had thereby guaranteed its continuance—not as a political power but as a political institution. Her long reign had woven a legend, and, as her political power ebbed away, her political value grew. It lay, perhaps, more in what the electorate thought of her, indeed felt about her, than in what she ever was or certainly ever believed herself to be. Paradoxically enough, her principal contribution to the British monarchy and her political importance lay in regard to those "dignified" functions that she was accused of neglecting rather than to the "business" functions that, perhaps sometimes, she did not neglect enough.

Death

The queen died at Osborne on Jan. 22, 1901, after a short and painless illness. "We all feel a bit motherless today," wrote Henry James, "mysterious little Victoria is dead and fat vulgar Edward is King." She was buried beside Prince Albert in the mausoleum at Frogmore near Windsor. Young said,

> She had lived long enough. The idol of her people, she had come to press on the springs of government with something of the weight of an idol, and in the innermost circle of public life the prevailing sentiment was relief.

Her essential achievement was simple. By the length of her reign, the longest in English history, she had restored both dignity and popularity to a tarnished crown: an achievement of character, as well as of longevity. Historians may differ in their assessment of her political acumen, her political importance, or her role as a constitutional monarch. None will question her high sense of duty or the transparent honesty, the massive simplicity, of her royal character.

BIBLIOGRAPHY. The most important authority for the queen's life is *The Letters of Queen Victoria* in three series: for the years 1837–61, ed. by ARTHUR CHRISTOPHER BENSON and VISCOUNT ESHER (REGINALD B. ESHER), 3 vol. (1907); for 1862–85, ed. by GEORGE EARLE BUCKLE, 3 vol. (1926–28); and for 1886–1901, ed. by GEORGE EARLE BUCKLE, 3 vol. (1930–32). See also HECTOR BOLITHO (ed.), *Letters of Queen Victoria,* trans. from German (1938; U.K. title, *Further Letters of Queen Victoria,* 1938, reprinted 1976). Correspondence between the queen and her eldest daughter was edited by ROGER FULFORD: *Dearest Child* (1964, reissued 1977), covering the years 1858–61; *Dearest Mama* (1968, reissued 1977), covering 1861–64; *Your Dear Letter* (1971), covering 1865–71; *Darling Child* (1976), covering 1871–78; and *Beloved Mama* (1981), covering 1878–85. Her correspondence with Gladstone is collected in PHILIP GUEDALLA (ed.), *The Queen and Mr. Gladstone,* 2 vol. (1933); and with Palmerston in BRIAN CONNELL (ed.), *Regina vs. Palmerston* (1961), covering the years 1837–65. See also HOPE DYSON and CHARLES TENNYSON (eds.), *Dear and Honoured Lady: The Correspondence Between Queen Victoria and Alfred Tennyson* (1969, reissued 1971). *Leaves from the Journal of Our Life in the Highlands from 1848 to 1861* (1868, reprinted 1969), and *More Leaves from the Journal of a Life in the Highlands, from 1862 to 1882* (1884), both ed. by ARTHUR HELPS, were published in the queen's lifetime and played their part in eventually securing public affection. Selections from these two journals were published in DAVID DUFF (ed.), *Queen Victoria's Highland Journals,* new and rev. ed. (1980). See also CHRISTOPHER HIBBERT, *Queen Victoria in Her Letters and Journals: A Selection* (1984).

Important biographies are LYTTON STRACHEY, *Queen Victoria* (1921; reissued with the title *The Illustrated Queen Victoria,* 1987), which was written before the second and third series of her *Letters* had been published; ELIZABETH LONGFORD, *Victoria R.I.* (1964; U.S. title, *Queen Victoria: Born to Succeed,* 1965, reprinted 1974); CECIL WOODHAM-SMITH, *Queen Victoria, from Her Birth to the Death of the Prince Consort* (U.K. title, *Queen Victoria: Her Life and Times,* 1972); DOROTHY MARSHALL, *The Life and Times of Victoria* (1972); and STANLEY WEINTRAUB, *Victoria: An Intimate Biography* (U.K. title, *Victoria: Biography of a Queen,* 1987).

Additional sources on her private life include MARINA WARNER, *Queen Victoria's Sketchbook* (1979), a collection of the queen's artistic works with accompanying narrative text; TYLER WHITTLE (MICHAEL SIDNEY TYLER-WHITTLE), *Victoria and Albert at Home* (1980); RONALD W. CLARK, *Balmoral, Queen Victoria's Highland Home* (1981); ROBERT RHODES JAMES, *Albert, Prince Consort: A Biography* (1983; U.S. title, *Prince Albert: A Biography,* 1984); and DELIA MILLAR, *Queen Victoria's Life in the Scottish Highlands: Depicted by Her Watercolour Artists* (1985).

Studies of the queen's reign are FRANK HARDIE, *The Political Influence of Queen Victoria, 1861–1901,* 2nd ed. (1938, reissued 1963); HECTOR BOLITHO, *The Reign of Queen Victoria* (1948); THEO ARONSON, *Victoria and Disraeli: The Making of a Romantic Partnership* (1977, reprinted 1987); JEFFREY L. LANT, *Insubstantial Pageant: Ceremony and Confusion at Queen Victoria's Court* (1979); JOHN MAY, *Victoria Remembered: A Royal History, 1817–1861, Entirely Illustrated by Commemoratives* (1983); and BARRY ST.-JOHN NEVILL (ed.), *Life at the Court of Queen Victoria, 1861–1901* (1984).

(E.T.W./M.V.)

Vienna

The capital city of Austria, Vienna (German: Wien) is also one of the country's nine *Bundesländer* (federal states)—the smallest in area but the largest in population. Modern Vienna has undergone several historical incarnations. From 1558 to 1918 it was an imperial city—until 1806 the seat of the Holy Roman Empire and then the capital of the Austro-Hungarian Empire. In 1918 it became the capital of a truncated, landlocked central European country that emerged from World War I a republic by choice. From 1938 to 1945 Austria was a part of Hitler's "Greater" Germany, and Vienna became "Greater" Vienna, reflecting the Nazi revision of the city limits. In the decade following World War II Austria was occupied by British, French, American, and Soviet forces, and Vienna was divided into five zones, including an international zone, covering the Innere Stadt ("Inner City"). In 1955 the State Treaty, by which the country regained independence, was signed with the four occupying powers, and Vienna became once again the capital of a sovereign Austria.

Vienna is among the least spoiled of the great old western European capitals. Its central core, the Innere Stadt, is easily manageable by foot and public transportation. In a city renowned for its architecture, many of Vienna's urban prospects remain basically those devised over several centuries by imperial gardeners and architects. The skyline is still dominated by the spire of St. Stephen's Cathedral and by the giant Ferris wheel in the city's chief park, the Prater. The city suffered heavy damage in the last months of World War II, and much rebuilding was done after the war. Nevertheless, the character of Vienna as a whole remains much the same as in the years before 1914.

Viennese *Lebenskunst* ("art of living") has survived changing rulers and times. It is still possible to live in Vienna at almost the same pace and in much the same style as it was a century ago. The same music is played in the same rebuilt concert halls, and a theatrical or operatic success still stimulates lively conversation. One can drink the same sourish local wines in the taverns on the outskirts of town, consume the same mountains of whipped cream at Sacher's and Demel's, and sample the same infinite varieties of coffee in countless cafés. Thick woolen suits and overcoats in shades of green, gray, or brown loden cloth and colourful dirndl dresses are still the fashion. It is even possible for tourists, and for others on festive occasions, to ride in a traditional fiacre, the two-horse carriage driven by a bowler-hatted coachman.

Austria's capital has avoided many of the problems—financial crises, social unrest, urban decay—that afflict other European cities. Its people enjoy an enlightened health and welfare system, which originated in the reforms of Empress Maria Theresa and Emperor Joseph II in the 18th century. A city of green parks with ponds, cafés, and playing bands; opulent stores and elegant shopping streets; banks, bookshops, and crowded theatres; and boulevards for leisurely sauntering—Vienna is an invigorating distillation of human energy and imagination.

This article is divided into the following sections:

Physical and human geography 493
- The landscape 493
  - Site
  - Climate
  - Layout and architecture
- The people 495
  - Ethnicity and language
  - Religion
- The economy 495
  - Industry and trade
  - Transportation
- Administration and social conditions 495
  - Government
  - Health and welfare
  - Education
- Cultural life 496
  - Music and theatre
  - Museums and libraries
  - Coffeehouses and taverns

History 496
- The ancient city and medieval growth 496
- Development of imperial Vienna 497
- Evolution of the modern city 497

Bibliography 498

Physical and human geography

THE LANDSCAPE

Site. Vienna lies in the northeastern corner of Austria, between the foothills of the Alps and the Carpathians, where the Danube (German: Donau), Europe's second longest river, has cut its course through the mountains. The city is situated alongside the river, most of it on the right bank. The Vienna basin was a nodal point of ancient trade and military routes. It linked north and south along the "amber route" that ran southward from the Baltic and linked east and west along the Danube. Strategically, Vienna commands the surrounding regions, which include sections of Austria's border with Czechoslovakia and Hungary.

Climate. Lying east of the Alps, Vienna is protected from their climatic influences by a range of hills, the Vienna Woods (Wiener Wald). The city's weather comes both from the north, the winds bringing cool summers and warm winters, and from the southeast, bringing heat in summer and cold in winter. The result, despite some summer fog and heat and winter snow and ice, is a generally temperate and agreeable climate. Throughout the year the temperature averages above 50° F (10° C). The characteristic *Lüfterl* ("Vienna air"), a light breeze blowing from the northwest and west, provides relief on hot summer evenings. Rainfall is fairly low, averaging 26 inches (66 centimetres) per year, the greater part of it coming in summer downpours.

Layout and architecture. With an area of 160 square miles (415 square kilometres), Vienna reaches across the Danube on one side and climbs into the Vienna Woods on the other. There it includes the 1,585-foot (483-metre) Kahlen Mountain (Kahlenberg) and the 1,778-foot (542-metre) Hermanns Mountain (Hermannskogel), Vienna's highest point. The Vienna Woods slope to the river in four roughly semicircular terraces, with the Innere Stadt occupying the second lowest terrace. The city has a mean altitude of 1,804 feet (550 metres), but different sections vary considerably in height.

The Vienna Woods

A stretch of the Danube was straightened and confined in the 19th century to form the Danube Canal, a flood control canal, parallel to the main stream, that flows through the city. An island 13 miles long and 750 feet wide was thus created from former floodlands and was equipped as an all-sports park, adding to the city's already generous recreational space. The Lobau, a wooded section along the river, has, like the Vienna Woods, long been a protected greenbelt area. Since the 1970s the open spaces on the far side of the Danube have been exploited for apartment buildings and factories.

Administratively, Vienna is divided into 23 *Bezirke* (dis-

tricts). At the core is district I, the Innere Stadt, which contains most of the city's famous structures. Surrounding the heart of the city is the Ringstrasse, or Ring, a circular road lined with grand buildings, monuments, and parks. Beyond the Ring are the inner suburbs (districts II–IX). The many palaces, churches, embassies, and other buildings in this area are elegant, though generally less imposing than those in district I. Leopoldstadt (district II) was the area allotted in 1622 to the Jews, who lived there until 1938. In this district is the famous 3,200-acre (1,295-hectare) Prater, formerly the hunting and riding preserve of the aristocracy, but since 1766 a public park whose amenities include a stadium, fairgrounds, racetracks, and many restaurants. Beyond another ring road, the Gürtel, lie the outer suburbs (districts X–XX), which are largely residential. Also beyond the Gürtel is the vast Central Cemetery, where many great musical figures and other famous Viennese are buried. Districts XXI and XXII lie on the far side of the Danube; district XXIII is at the southern edge of the city.

Prominently situated in the centre of Vienna is St. Stephen's Cathedral (Stephansdom), one of the chief Gothic buildings of Europe. It incorporates remnants of the original 12th-century Romanesque structure, which was destroyed by fire. Reconstruction began in the early 14th century and continued for a century and a half. The northern tower, never completed, was topped off with a Renaissance dome between 1556 and 1587. The cathedral was again burned and partly destroyed in World War II but has since been restored. The 20-ton bell, made from captured Turkish cannons in 1711, was recast and rehung with much ceremony.

Other Gothic churches include the Church of the Augustinians, the Church of Maria am Gestade, and the Church of the Friars Minor (officially the Snow Madonna Italian National Church), all dating from the 14th century. Vienna's oldest church is St. Ruprecht's. Dating from the 13th century with parts from the 11th century, it is believed to have originally been erected in 740.

The Church of St. Peter, a Baroque structure thought to be standing on the site of a church founded by Charlemagne in 792, was built chiefly by the architect Johann Lucas von Hildebrandt in 1702–33. Other fine examples of Baroque art are the richly frescoed University Church (1627–31) and the Church of the Capuchins (1632), which contains the crypt of the Habsburg imperial family. The Church of the Scots (1155), together with a monastery for Scottish and Irish monks, was rebuilt in late Italian Renaissance style in 1638–48. The style of most of the finest secular buildings, such as the Harach and Kinsky palaces and the winter palace of Prince Eugene of Savoy, is Baroque, Vienna's leading architectural style in the 17th and 18th centuries.

© Paolo Koch—Rapho/Photo Researchers, Inc.

The Burgtheater on the Ring and, in the background, the spire of St. Stephen's Cathedral, which dominates the Vienna skyline.

The Hofburg

The vast complex of the Imperial Palace, the Hofburg (or Burg), lies along the Ringstrasse. It consists of a number of buildings of various periods and styles, enclosing several courtyards, the oldest part dating from the 13th century and the latest from the end of the 19th. The Hofburg abounds in magnificently appointed private and state apartments. It houses the imperial treasury of the Holy Roman and Austrian empires, the Austrian National Library, the Albertina and several other museums, and the Spanish Riding School. The state apartments in one wing of the Hofburg serve as the offices of Austria's president. Close by stands the Privy Court Chancery (1716–21), where the Congress of Vienna met after the Napoleonic Wars.

The other important buildings along the Ring are mainly mid-19th-century versions of earlier European styles. They include the Stock Exchange (Börse), in Neoclassical-Renaissance style, and the pseudo-Gothic Votive Church, built by Emperor Francis Joseph after he escaped an assassination attempt in 1853. Nearby is the University of Vienna, the oldest university in the German-speaking world, designed in the Italian Renaissance style. The university was founded in 1365, but its original buildings have disappeared.

Another landmark is the City Hall (Rathaus), in neo–Flemish-Gothic with Renaissance touches, and facing it is the Burgtheater, in a mixture of neo–Italian-High Renaissance with Baroque indulgences. The Neoclassical Parliament building lies adjacent to the Palace of Justice, built in 16th-century German Renaissance style. The neo-Renaissance Natural History Museum and the Kunsthistorisches Museum stand in front of an exhibition centre, formerly the royal stables. Across the Ring from the museums is the Hofburg's last extension, the Neue Hofburg, and eastward is the magnificent Vienna State Opera House, built in 1861–69. Purporting to be French early Renaissance, the State Opera is actually a conglomeration of imitative architectural styles, of pinnacles, arcades, colonnades, and heroic statuary, yet it somehow achieves a serene and noble harmony.

On the eastern side of the Innere Stadt lies the City Park, rich in monuments. The Innere Stadt and its immediate neighbourhood are still, unlike the older parts of most European cities, the fashionable quarter, containing the government offices, the principal hotels, embassies and legations, and many other fine buildings. The Schönbrunn Palace, the summer residence of the Habsburgs, with its splendid rooms decorated in Rococo style and its great formal park, lies to the southwest in the suburb of Hietzing.

The Belvedere

Another noble structure is the Belvedere, which is actually two Baroque palaces at either end of a terraced garden. It was built by Hildebrandt for the soldier and statesman Prince Eugene of Savoy. The Lower Belvedere (1714–16) was a summer garden palace, and the Upper (1721–24) was designed as a place of entertainment. Both now house museums of Austrian art. The Austrian State Treaty, which ended the four-power occupation of the country, was signed in the Upper Belvedere on May 15, 1955.

The Church of St. Charles, a vast structure dedicated to St. Charles Borromeo, was erected just outside the city walls in 1716–39. This Baroque edifice is fronted by a severely classical porch of columns in ancient Roman style, and before it stand spirally decorated twin columns carved with scenes from the saint's life. A few streets away from the Church of St. Charles is the Theater an der Wien, built between 1789 and 1801. Mozart conducted the first performance of *The Magic Flute* in 1791 in the theatre's wooden predecessor, and Beethoven's *Fidelio* had its premiere in the newly constructed theatre in 1805. All of the celebrated operetta composers of the 19th century presented works on its stage. In 1962 the municipality bought the dilapidated house, restored it, and now operates it as an orchestra hall.

Two monuments—built by Johann Bernhard Fischer

von Erlach and greatly esteemed by the Viennese—were thanksgiving offerings. One is the 69-foot Trinity Column, or Plague Column, on the fashionable shopping boulevard the Graben, which commemorates the cessation of the plagues that struck the city in 1679 and 1713; the other, in more sober Baroque style, is Joseph's Fountain, a votive column and fountain in the Hoher Market, donated by Emperor Leopold I for the safe return from battle of Joseph I, his firstborn son and heir.

THE PEOPLE

Before and during World War II a number of Vienna's citizens emigrated to the West to escape the Nazis. Following the war, Vienna's population decreased when part of Greater Vienna was reintegrated with the province of Lower Austria. There were also moves in population from eastern to western Austria connected with the German occupation from 1938 to 1945 and the presence of Soviet troops from 1945 to 1955. Altogether, Vienna's population decreased by about 15 percent between 1934 and 1951.

Vienna has a disproportionately large number of elderly, most of whom live alone in the older neighbourhoods. Characteristic of most major cities, however, Vienna's population is shifting outward to the periphery. Although Vienna has a low birth rate and a small average family size, new housing in the periphery helps to alleviate problems caused by the city's high percentage of pre-World War I residential buildings.

Ethnicity and language. The Viennese are the product of centuries of cross-fertilization between mountain and plain, between the Balkan strain from one direction and the Germanic from the other. As the names in Vienna's telephone directory indicate, the ancestors of one Viennese in three have come from Bohemia (now part of Czechoslovakia), one in five from Hungary, one in seven from Poland, and one in eight from what is now Yugoslavia. Along with those whose families migrated from Germany and other parts of Austria, this mixture makes up the melting pot of Vienna.

Vienna's ethnic melting pot

Wienerisch, the Viennese speech and accent, reveals social levels and origins. It also demonstrates that the people of Vienna have in their time been governed by Romans, Italians, Spanish, French, Magyars, and Slavs and have absorbed Turkish and Yiddish words into their German tongue in a way that renders the original unrecognizable. Their speech is in many ways closer to that of their neighbours to the south and east than to the German north, and one of its functions is to announce that "we are different." If the people have any leaning toward pomposity, it is balanced by a habit of self-mockery, as expressed in their saying, "The situation is hopeless but not serious." The famed *Gemütlichkeit* (untranslatable but akin to "coziness"), upon which the city's tourist trade thrives, is—like the nostalgia for wine, women, and song—part of a sentimental image of the Viennese.

Religion. Vienna is the seat of a Roman Catholic archbishop and a Protestant bishop. Two-thirds of the city's population are Roman Catholic and only a very small percentage Protestant. (Considerable numbers profess no religion.) The number of practicing Roman Catholics, however, is estimated to be only a small percentage of the population; like other modern capitals, Vienna is highly secular.

Before World War II the Viennese Jewish minority played a prominent role in the city, culturally and economically. It is estimated that two-thirds of all Jews emigrated to escape the Nazi occupation. Except for a small remnant that survived, either in hiding or in the concentration camps, the remainder of the Jewish Viennese were exterminated by the Nazis. They now make up less than 1 percent of the population.

THE ECONOMY

Although commerce and industry form the base of Vienna's economy, government and public administration on all levels is also a major employer in the Austrian capital. The federally owned theatres alone employ several thousand people. Tourism is also an important economic activity with some 2 million travelers visiting the city annually. Vienna provides approximately one-fourth of the jobs in the country and produces almost one-third of the gross national product. The steady reduction in numbers of active workers owing to the rising proportion of older people has necessitated the recruitment of foreign labour, mainly Yugoslavs and Turks, who work primarily in the service sector and in menial occupations.

Industry and trade. Vienna produces more than half of Austria's capital goods and almost half of its consumer goods. Leading industries include the manufacture of machinery (primarily electrical machinery and transportation equipment), electrical products, chemicals, and metal products. In the Vienna area oil processing, cement works, and brickmaking are important as well. There are more than 25,000 small manufacturing firms in Vienna, a majority of them employing fewer than five workers and most of the rest having fewer than 20. Special Viennese products are silk, velvet, linen, ceramics, jewelry, scientific and musical instruments, watches, cutlery, leather goods, furniture, paper, and carpets. The service industries in Vienna, including banking, account for half of Austria's total employment in this sector. The proportion of white-collar workers, public employees, and civil servants in the total labour force continues to grow.

The Vienna international trade fair that takes place twice a year in March and September is of special importance to the economy of Austria. The fair attracts exhibitors from both European and overseas countries and is attended by several hundred thousand visitors annually. Despite the political division in Europe, Vienna has remained an important centre of East–West trade. Several hundred American, West German, Japanese, and British firms, as well as many eastern European firms, use Vienna as a base for trading operations. Approximately 10 percent of Austria's exports go to the eastern European countries.

International trade fairs

Transportation. In the 19th century tens of thousands of immigrants from all parts of the Austro-Hungarian Empire arrived at Vienna's six major railway terminals. Today only two important stations are left, South Railway Station and West Railway Station. The city's busy international airport, Schwechat, is served by more than 30 airlines, and motorways radiate in all directions. Freight transport down the Danube to Yugoslavia, Romania, and the Black Sea, and via the Rhine–Main–Danube Canal to northwestern Europe, has grown in importance. Vienna's containerport has an annual turnover of about 1.6 million tons. After World War II Vienna chose to retain its tramway system instead of converting to buses. The old system, modernized and updated, continues to be an important low-cost form of public transportation. The extensive underground network has also been expanded. Consequently, within the city most people travel by public transportation or on foot.

ADMINISTRATION AND SOCIAL CONDITIONS

Government. Because of the dual character of Vienna as the capital city and a federal state of the republic, the municipal and the state administration are in the hands of the same elected representatives acting in institutionally separate capacities under the mayor and his two deputies. The mayor of the city also functions as governor of the state. The affairs of the 23 municipal districts are managed by appointed magistrates. Representatives to the city council are elected every five years by proportional representation. Since the end of World War I, the mayor's office and a majority of the council's 100 seats have been held by members of the Socialist Party. The mayor, his two deputies, and 14 councillors constitute the executive authority. Vienna sends 12 members to the Federal Council, the second chamber of the Austrian Parliament.

The government not only runs the city but also operates a major business, the Vienna Holding, a combination of state and private enterprise. Its firms include low-cost restaurants, a major publishing house, an insurance company, a cold-storage depot, shopping centres, cinemas, and the large, multifunctional Stadthalle ("City Hall"), with a seating capacity of 16,000, for sporting events, concerts, dances, exhibitions, and swimming. The old Theater an der Wien and the traditional Viennese porcelain factory,

which was closed in 1864, were rescued from extinction by this enterprise.

Health and welfare. Vienna's hospitals and medical training have been widely esteemed since the mid-18th century. Emperor Joseph II founded the General Hospital in 1784, and in the 19th century Viennese medicine led the world. Vienna claims several renowned medical scientists, among them Ignaz Philipp Semmelweis, discoverer of the cause of puerperal fever; Theodor Billroth, a pioneer in abdominal surgery; Karl Landsteiner, discoverer of the blood groups; and Sigmund Freud, the founder of psychoanalysis.

For many reasons the city's public health care system is regarded as one of the world's best. The number of doctors in proportion to the population is high; there are more than 40 general and special hospitals and numerous geriatric facilities. In addition to providing model health services, the municipal government has been among the world's pioneers in public welfare and social insurance. Vienna is also renowned for its clean drinking water, which comes from springs in the mountains around the city.

Universities

Education. Vienna has a much higher proportion of high-school and university graduates than the other Austrian provinces. Of the 12 universities in the country, five are located in Vienna: the University of Vienna, the University of Technology, the University of Agriculture, the University of Veterinary Medicine, and the University of Economics in Vienna. Other notable institutions include the Academy of Music and Dramatic Art, the Academy of Fine Arts, and the Academy of Applied Arts. There are also a Roman Catholic academy, several scientific societies, and many research institutes, as well as the venerable Austrian Academy of Sciences. Vienna's teaching of music, medicine, law, and the arts attracts many foreign students, who make up about 10 percent of the total student population.

CULTURAL LIFE

Music and theatre. Vienna is the undisputed cultural centre of Austria and one of the world capitals of music. Even the Salzburg and Bregenz festivals are dependent on Viennese orchestras, musicians, theatre directors, and actors. Operas, concerts, and theatrical performances have played a major part in Viennese life for centuries, and many world-famous composers lived and worked in the city. The famous Society of Friends of Music, founded in 1812, helps to ensure that Vienna remains a leading music centre.

The Vienna Boys' Choir, founded in 1498 (Haydn and Schubert were its most famous boy members), sings on Sunday mornings at the mass in the Hofburg Chapel. The Vienna Philharmonic Orchestra gives frequent Saturday afternoon and Sunday morning concerts and also performs during the week at the State Opera House. Altogether there are seven concert halls in Vienna. Among the highlights of the Viennese musical calendar are the annual gala performance of Johann Strauss's operetta *Die Fledermaus* on New Year's Eve and the New Year's concert of the Philharmonic, broadcast and televised throughout the world.

The two major opera houses, the State Opera and the People's Opera, and the two leading theatres, the Burgtheater and the Academy Theatre, are owned by the Austrian federal government, and their singers and actors enjoy respected civil servant status. The State Opera is one of the leading opera houses in the world, where Verdi and Wagner conducted and where Gustav Mahler was director. It opened in 1869 with a performance of Mozart's *Don Giovanni.* During World War II it was destroyed, and, after rebuilding, it reopened in 1955 with a performance of Beethoven's *Fidelio.* The Burgtheater, founded in 1776, is one of the most highly regarded German-language theatres in Europe. In addition to several large theatres, Vienna has numerous small theatres, which provide a home for more avant-garde works.

Art treasures

Museums and libraries. Vienna has a wide variety of museums and historic houses. Among them are the Albertina, with its immense collection of graphic arts, including engravings by Dürer and Rembrandt; the Kunsthistorisches Museum with the largest Bruegel collection outside The Netherlands; the Academy of Fine Arts, housing the superb Habsburg collection of old masters, especially rich in Flemish and Dutch paintings; the Imperial Treasury, with the imperial crown and the regalia of the Holy Roman emperors and the house of Habsburg; the museums of natural history, ethnology, military history, and technology; the Clock Museum; and the Museum of the City of Vienna, with its exhibits of Viennese history.

The Roman excavations in the Hoher Market, converted into an underground museum; the catacombs of St. Stephen's Cathedral; the Imperial Vault in the Church of the Capuchins, burial place of the Habsburg emperors; and the exhibits and imperial apartments in the Schönbrunn Palace offer a historical dimension to the city's art treasures. The houses in which Haydn, Mozart, Beethoven, Schubert, and Johann Strauss lived and worked are open to the public. The apartment that was Sigmund Freud's home and office for nearly 50 years is also a museum.

In addition to its museums and historical sites, Vienna is notable for its libraries, including the National Library, the University Library, the City Library, and the libraries of the Natural History Museum and the Academy of Sciences.

Coffeehouses and taverns. The coffeehouse has been a Viennese institution for three centuries. According to legend, the first such establishment opened with an inventory of Turkish coffee beans, part of the booty from the Siege of Vienna in 1683. There are a variety of coffees and an assortment of supplements such as cream or brandy to choose from. The Viennese have turned the coffeehouse into a sort of second living room, where they not only drink their beverage and consume pastries but also read periodicals, play cards, and chat with friends. There were once famous literary and theatrical cafés where artists and famous personalities held court; still flourishing is the Café Demel, a true custodian of the past.

Also peculiar to Vienna are the taverns in which is served the young, sour wine—*Heuriger*—of the previous year's local harvest. Some of the most famous taverns are in the outlying districts of Vienna, such as Grinzing, Nussdorf, and Sievering, and they are identified by evergreen branches hung over the entrance. The wine drinking is accompanied with music, usually played by a trio of instruments, including those such as a fiddle, accordion, guitar, or zither.

History

THE ANCIENT CITY AND MEDIEVAL GROWTH

Traces of human occupation of the site of Vienna have been found dating as far back as the Paleolithic period. The area was subsequently inhabited by the Illyrians and then the Celts. In 16–15 BC the Romans, under the future emperor Tiberius, occupied the foothills of the Alps, and in the next century the Celtic town of Vindobona (Celtic: "White Field"; later to become Vienna) became a strategic Roman garrison town. (The Roman camp is believed to have covered the area around the present Hoher Market.) Vindobona grew to about 15,000 inhabitants and became part of a widespread network of trade and communications. Emperor Marcus Aurelius is said to have died in Vindobona in 180 AD fighting off attacks by the Germanic tribes. The Romans were swept away in the turmoil of the 5th-century invasions, but enough of Vindobona remained to serve as the nucleus of the medieval city. The Bavarians occupied the area, and the people became Christianized. The city's name was recorded in 881 as Wenia and in 1030 as Wienis.

The Babenberg era

The dukes of Babenberg, a Frankish dynasty, were overlords of Vienna from 1156 to 1246. The city developed into an important trading centre, where crusaders on their way to the East bought provisions and equipment. In the 13th century, walls were built around the city, and Vienna remained largely confined within the walled area until the 1700s. The Babenbergs kept a brilliant court and encouraged artists like the famous minnesinger Walther von der Vogelweide.

In 1246 the last male of the Babenberg family died. In the ensuing struggle for domination, the king of Bohemia,

Otakar II, became overlord of what was to become Austria. Otakar established himself as a powerful central European prince, and by 1276 he was at war with the German king, Rudolf I of the Habsburg dynasty. When Otakar fell in battle in 1278, the Habsburgs took over his domain and retained it for more than 600 years. The capital city flourished, trading with Trieste, Venice, and Hungary; nevertheless, economic decline attended the numerous disputes over inheritance within the Habsburg family. In 1485, under siege by Matthias I Corvinus of Hungary, the city fathers surrendered in the hope of bettering their status. When Corvinus died five years later, Vienna reverted to the Habsburg emperor, Frederick III.

DEVELOPMENT OF IMPERIAL VIENNA

During the Renaissance, Vienna was a leader in science and fine arts, and the university (1365) was a centre of humanism. When Charles V became Holy Roman emperor in the 16th century he entrusted his Austrian territories to his brother, the future emperor Ferdinand I. Seeking to increase their liberties and economic position, the Lower Austrian Diet rebelled against their regent. Ferdinand responded by condemning the leaders of the insurrection to death, and in 1526 he issued an ordinance that stripped the city of almost all its rights. In the same year, he inherited the kingdoms of Bohemia and Hungary and, accordingly, the task of fighting the Turks, who commanded large parts of Hungary. Turkish forces besieged Vienna in 1529 but were successfully beaten off. When Ferdinand was crowned emperor in 1558, Vienna regained its political status and became the administrative seat of numerous kingdoms that the Habsburgs acquired by marriage.

The Reformation swept through Europe during the 16th century, arousing heated opposition from the Roman Catholic church. In an attempt to stem the controversy, the imperial Diet, in the Peace of Augsburg (1555), recognized the right of Lutheranism to exist but decreed that the regional princes were to determine which form of Christianity their subjects must follow. Because the Viennese were required to remain Roman Catholic, many of the great number who had become Protestant had to leave the city. It was during this period that new fortifications were built to replace the medieval city walls and the Hofburg was enlarged with the addition of new courts. The splendid secular buildings of the Baroque era proclaimed Vienna's stature as an imperial residence and one of the great world capitals.

Peace of Augsburg

In 1679 the bubonic plague struck the city, killing nearly a third of its population. Then, during the summer of 1683, Vienna suffered a second Turkish siege, this one led by the grand vizier Kara Mustafa. The Viennese defenders, together with imperial troops under Charles of Lorraine, held off the Turkish army, which was defeated with the help of relief forces led by John III Sobieski, king of Poland. Shortly thereafter Prince Eugene of Savoy succeeded in driving the Turks out of Hungary.

The Turkish siege

With the Turkish threat at an end, there followed an upsurge of building, particularly in the devastated suburbs. Between 1700 and 1730 a city of palaces and stately homes emerged. A second line of fortifications, the *Linienwall* ("straight rampart"), was built in 1704–06 to give the suburbs protection. In the densely built-up Innere Stadt old houses either had upper stories added or were demolished and replaced by Baroque structures. Hildebrandt, J.B. Fischer von Erlach, and Fischer von Erlach's son Joseph Emanuel were the great Viennese architects of the time, and their achievements are still evident in some of the city's best buildings.

During this period, immigrants arrived from other parts of the empire, and new factories heralded the city's transition from trade to manufacturing. The arts also received fresh energy, as instanced by Joseph Anton Stranitzky's newly created Viennese Impromptu Theatre, which opened with the character masque of Hanswurst.

The male line of the Habsburgs died out with Charles VI in 1740, but by the terms of the Pragmatic Sanction his daughter Maria Theresa gained the right of succession and reigned until 1780. She established compulsory primary-school attendance; separated the university from the church, bringing it under state control; and reorganized the economy, the army, and the judiciary. Her son and successor, Joseph II, was typical of the Enlightenment's absolute monarchs and continued in her reforming spirit. His Edict of Toleration guaranteed religious freedom to Protestants and Jews. He instituted many humanitarian measures, improved government and education, and supported the arts. Some of his actions, like the dissolution of the monasteries, brought him into conflict with the church. By the time Joseph died in 1790, there were 300 factories in Vienna, the population had increased to 235,000, and the built-up area had increased 10-fold since the Turkish siege. Gluck, Haydn, and Mozart had ushered in Vienna's first golden age of music; Beethoven and Schubert would carry it into the next century.

Musical pioneers

In 1804 Francis II declared himself emperor of Austria and in 1806 resigned his former imperial crown, thus bringing to an end the Holy Roman Empire, which had long been essentially a German monarchy. Napoleon's armies occupied Vienna in 1805 and again in 1809. Inflation and state bankruptcy followed the Napoleonic Wars. Politically, however, Vienna held a central position in the restoration of Europe at the Congress of Vienna (1814–15) under the leadership of the powerful statesman Prince Metternich.

By 1845 Vienna had 430,000 inhabitants. The aspirations and cultural interests of the middle class were growing, finding artistic expression primarily in the simple and commonplace forms of the Biedermeier style of decoration and furniture design. Joseph Lanner and the elder Johann Strauss enlivened the city with Viennese waltzes. The revolution of March 1848 in Vienna brought to an end Metternich's authoritarian rule. A second uprising, in October, was put down by the imperial army of Francis Joseph. The city continued to grow culturally as the Austrian (later the Austro-Hungarian) imperial capital.

EVOLUTION OF THE MODERN CITY

Vienna's inner ramparts were razed in 1857 and the city ditches filled in. They were replaced by the Ringstrasse, opened in 1865. The stately public buildings and parks along this great avenue emerged over the years. In other parts of the city old structures were demolished and new ones built. The drinking-water supply was improved with springwater; the Danube was regulated; and, later, gas and electric works were built. New regulations in 1859 established full freedom of trade. Vienna's economy grew rapidly, and with it the city's population.

In 1861 Vienna was granted city self-government through a freely elected city council, which acquired a liberal majority. The suburbs were brought under the city administration in 1890. Three years later the *Linienwall* was dismantled and a second ring road, the Gürtel, built in its place. Musical Vienna flourished under the composers Brahms, Anton Bruckner, Hugo Wolf, and Mahler. Operetta became established as a characteristic Viennese art form through the music of the younger Johann Strauss, Franz von Suppé, Franz Lehár, and Emmerich Kálmán.

At the turn of the century, with nearly 2 million inhabitants and an area of 105 square miles, Vienna began to spread to the left bank of the Danube. The capital city had become a fertile breeding ground for ideas that—for good or bad—were to shape the modern world. Among the thousands who flocked to Vienna seeking work or a vocation was the young Adolf Hitler. Having failed there as an art student, he adopted notions of Greater German nationalism and was influenced by both the rhetorical style and the virulent anti-Semitism of Vienna's mayor, Karl Lueger. In contrast, Theodor Herzl, the founder of Zionism, was living in the city at the same time. In Vienna as well were Sigmund Freud and Alfred Adler, at work developing their far-reaching psychiatric theories.

Artistic activity reached new heights in the early 20th century. In architecture, painting, and design, the Sezessionstil (German: Jugendstil) movement (Art Nouveau) provided means for young artists to rebel against pretentiousness in the Viennese art establishment, including the style of architecture on the Ring. Otto Wagner, Adolf Loos, and Josef Hoffmann were prominent architects and

Revolution in the arts

designers of the new school; among the painters were Gustav Klimt, Alfred Kubin, Oskar Kokoschka, and Egon Schiele. Viennese music continued to break new ground with the works of Arnold Schoenberg, Alban Berg, and Anton von Webern. Influential writers included Arthur Schnitzler, Hugo von Hofmannsthal, and Karl Kraus.

During World War I the city's population swelled to 2,239,000 with an influx of refugees. Francis Joseph died in 1916; his successor, Charles, was forced to abdicate at the end of the war. The Habsburg monarchy fell, and a German-Austrian republic was proclaimed in November 1918. In 1919 general suffrage produced a Social Democratic majority in the city council. The party introduced numerous reforms in housing, education, public welfare, and health care. The city became recognized internationally as a forerunner of the modern welfare state. In 1922 Vienna was made a federal province of Austria, the mayor of the city also serving as the governor of the province.

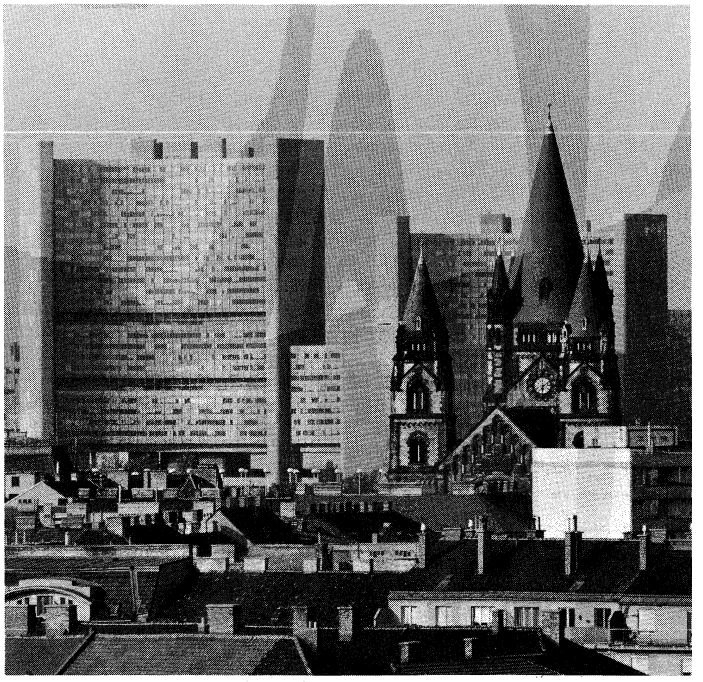
Economically and politically the young republic was on shaky ground. The Social Democrats (Marxist) and their Christian Social (conservative) opponents carried politics into the streets and trained their own armies. Mass political demonstrations were answered with savage repression, and in 1933 the Austrian chancellor, Engelbert Dollfuss, dissolved Parliament and set up an authoritarian regime. His government's decision to fire upon Viennese workers defending their apartment blocks (1934) was seen by many as a final betrayal of democracy. Disillusioned with party politics and drawn toward their fellow countryman Adolf Hitler in neighbouring Germany, many Austrians adopted the old idea that they were not only linguistically but also politically German and ought to unite with Germany. On March 12, 1938, the German army occupied Austria. A few days later, Hitler proclaimed the Anschluss; Greater Vienna, now including the province of Lower Austria, became a German province.

The Austrians soon became disenchanted with their new masters and, while fighting with Germany in World War II, developed a new sense of independence and nationhood. In April 1945 Vienna was taken by Soviet troops. In September the city, like the country as a whole, was divided into four zones of occupation by the Allied powers. Vienna emerged from the war with nearly a quarter of its buildings either partially or completely destroyed. The city's population had fallen to 1,322,000. The immense task of providing food and shelter, repairing the transportation network, and rebuilding the city began under the mayors Theodor Körner (1945–51) and Franz Jonas (1951–65), both of whom later became presidents of the republic. The Austrian State Treaty was signed in the Belvedere on May 15, 1955, leading to independence and the withdrawal of all Allied occupation troops.

Aftermath of World War II

Vienna's geographic position at the heart of Europe has been devalued three times in the 20th century: in 1918, when it became the top-heavy capital of a small republic; in 1938, when the takeover by Germany reduced it to a German province; and in 1945, when the Iron Curtain came down some 50 miles to the east. But to a large extent the city remains a focal point of East–West contacts. A neutral state since 1955, Austria has become important as a major international meeting place and conference centre. On the outskirts of Vienna, across the Danube, the modern buildings of the Vienna International Centre, or UNO-City, include the offices of the International Atomic Energy Agency, the United Nations Industrial Development Organization, and the Organization of Petroleum Exporting Countries. With the rise of the UNO-City, Vienna has become competitive with New York City and Geneva as a seat of world organizations.

© Sam C. Pierson, Jr.—Photo Researchers, Inc.



An old church and the modern buildings of UNO-City (Vienna International Centre) near Danube Park.

BIBLIOGRAPHY. General descriptions of Vienna include CHRISTA ESTERHÁZY, *Vienna* (1966); JOSEPH WECHSBERG, *Vienna My Vienna* (also published as *Sounds of Vienna,* 1968); and MARTIN HURLIMANN, *Vienna* (1970; originally published in German, 1968). The city's historical geography is presented in WIENER STADT- UND LANDESARCHIV, LUDWIG BOLTZMANN INSTITUT FÜR STADTGESCHICHTSFORSCHUNG, *Historische Atlas von Wien* (1981–), issued in parts. Photographs of Vienna with accompanying essays may be found in MARIA NEUSSER-HROMATKA, *Beautiful Vienna,* 8th ed. (1969; originally published in German, 1959); ANTON MACKU, *Vienna* (1957; originally published in German, 1956); INGE MORATH *et al., Bilder aus Wien: der liebe Augustin* (1981); ANNA GIUBERTONI, CLAUDIO MAGRIS, and TONI NICOLINI, *Austria* (1981), with Italian text; and GÜNTER DÜRIEGL, *Wien auf alten Photographies* (1981). HANS PEMMER and NINI LACKNER, *Der Prater: von den Anfängen bis zur Gegenwart,* 2nd ed. rev. by GÜNTER DÜRIEGL and LUDWIG SACKMAUER (1974), describes the Prater. Aspects of city life are discussed in E. BODZENTA, I. SPEISER, and K. THUM, *Wo sind Grossstädter daheim?: Studien über Bindungen an das Wohnviertel* (1981). LEOPOLD REDL and HANS WÖSENDORFER, *Die Donauinsel: ein Beispiel politischer Planung in Wien* (1980), analyzes economic conditions and municipal government.

Vienna's intellectual and cultural life is discussed by MARCEL BRION, *Daily Life in the Vienna of Mozart and Schubert* (1961; originally published in French, 1959); and by WILLIAM M. JOHNSTON, *Vienna, Vienna: The Golden Age, 1815–1914,* trans. from Italian (1981); CARL E. SCHORSKE, *Fin-de-Siècle Vienna: Politics and Culture* (1979), on the growth of modern art and thinking out of the political and social disintegration of turn-of-the-century Vienna; ROBERT WAISSENBERGER (ed.), *Vienna, 1890–1920* (1984, originally published in German, 1984); and MARK FRANCIS (ed.), *The Viennese Enlightenment* (1985). Viennese artists and musicians are portrayed in PETER VERGO, *Art in Vienna 1898–1918: Klimt, Kokoschka, Schiele, and Their Contemporaries* (1975, reissued 1981); KIRK VARNEDOE, *Vienna 1900: Art, Architecture, & Design* (1986); EGON GARTENBERG, *Vienna: Its Musical Heritage* (1968); and RICHARD RICKETT, *Music and Musicians in Vienna,* 2nd ed. (1981). ERNA LESKY, *The Vienna Medical School of the 19th Century* (1976; originally published in German, 1965), gives an account of Viennese medicine in its heyday.

Overviews of Vienna's history are provided by PETER CSENDES, *Geschichte Wiens* (1981); WALTER B. GOLDSTEIN, *1000 Jahre Wien und die Habsburger: eine europäische Legende* (1981), focusing especially on the house of Habsburg; INGE LEHNE and LONNIE JOHNSON, *Vienna—The Past in the Present: A Historical Survey* (1985); and PAUL HOFMANN, *The Viennese: Splendor, Twilight, and Exile* (1988), a cultural history. Important events and times are further examined in HISTORISCHEN MUSEUM DER STADT WIEN, *Die Türken vor Wien: Europa und die Entscheidung an der Donau, 1683* (1982); R. JOHN RATH, *The Viennese Revolution of 1848* (1957, reprinted 1977), a dramatic reconstruction of events; FREDERIC MORTON, *A Nervous Splendor: Vienna, 1888/1889* (1979); JOHN W. BOYER, *Political Radicalism in Late Imperial Vienna: Origins of the Christian Social Movement, 1848–1897* (1981); and ROBERT PICK, *The Last Days of Imperial Vienna* (1975). Studies of the Jewish community in Vienna include IVAR OXAAL, MICHAEL POLLAK, and GERHARD BOTZ (eds.), *Jews, Antisemitism, and Culture in Vienna* (1987), a collection of essays; and GEORGE E. BERKLEY, *Vienna and Its Jews: The Tragedy of Success* (1987).

(B.E./R.J.Hi.)

Virgil

Virgil was regarded by the Romans as their greatest poet, an estimation that subsequent generations have upheld; his fame rests chiefly upon his national epic poem, the *Aeneid,* which tells the story of Rome's legendary founder and proclaims the Roman mission to civilize the world under divine guidance. His reputation as a poet endures not only for the music and diction of his verse and for his skill in constructing an intricate work on the grand scale but because he embodied in his poetry aspects of experience and behaviour of permanent significance.

Publius Vergilius Maro (the spelling Virgil has become traditional in English, although the form Vergil also is current) was born of peasant stock at Andes, near Mantua, in Italy on October 15, 70 BC. His love of the Italian countryside and of the people who cultivated it colours all his poetry. He was educated at Cremona, at Milan, and finally at Rome, acquiring a thorough knowledge of Greek and Roman authors, especially of the poets, and receiving a detailed training in rhetoric and philosophy. It is known that one of his teachers was the Epicurean Siro, and the Epicurean philosophy is substantially reflected in his early poetry but gradually gives way to attitudes more akin to Stoicism.

Political background. During Virgil's youth, as the Roman Republic neared its end, the political and military situation in Italy was confused and often calamitous. The civil war between Marius and Sulla had been succeeded by conflict between Pompey and Julius Caesar for supreme power. When Virgil was 20, Caesar with his armies swooped south from Gaul, crossed the Rubicon, and began the series of civil wars that were not to end until Augustus' victory at Actium in 31 BC. Hatred and fear of civil war is powerfully expressed by both Virgil and his contemporary Horace. The key to a proper understanding of the Augustan Age and its poets lies, indeed, in a proper understanding of the turmoil that had preceded the Augustan peace.

Personal life

Virgil's life was devoted entirely to his poetry and to studies connected with it; his health was never robust, and he played no part in military or political life. It is said that he spoke once in the lawcourts without distinction and that his shy and retiring nature caused him to give up any ideas he might have had of taking part in the world of affairs. He never married, and the first half of his life was that of a scholar and near recluse. But, as his poetry won him fame, he gradually won the friendship of many important men in the Roman world. Gradually, also, he became a Roman as well as a provincial. (The area in which he had spent his youth, the area around the Po River known as the province of Cisalpine Gaul, was not finally incorporated into Italy until 42 BC. Thus Virgil came, as it were, to Rome from the outside. The enthusiasm of a provincial for Rome is seen in the first eclogue, one of his earliest poems, in which the shepherd Tityrus tells of his recent visit to the capital and his amazement at its splendours.)

Literary career. Some of Virgil's earliest poetry may have survived in a collection of poems attributed to him and known as the *Appendix Vergiliana,* but it is unlikely that many of these are genuine. His earliest certain work is the *Eclogues,* a collection of 10 pastoral poems composed between 42 and 37 BC. Some of them are escapist, literary excursions to the idyllic pastoral world of Arcadia based on the Greek poet Theocritus (fl. *c.* 280 BC) but more unreal and stylized. They convey in liquid song the idealized situations of an imaginary world in which shepherds sing in the sunshine of their simple joys and mute their sorrows (whether for unhappy love or untimely death) in a formalized pathos. But some bring the pastoral mode into touch with the real world, either directly or by means of allegory, and thus gave a new direction to the genre. The fifth eclogue, on the death of Daphnis, king of the shepherds, clearly has some relationship with the recent death of Julius Caesar; the 10th brings Gallus, a fellow poet who also held high office as a statesman, into the pastoral world; the first and ninth are lamentations over the expulsion of shepherds from their farms. (It was widely believed in antiquity that these poems expressed allegorically Virgil's own loss of his family farm when the veteran soldiers of Antony and Octavian—later the emperor Augustus—were resettled after the Battle of Philippi in 42 BC. It was thought that he subsequently recovered his property through the intervention of his powerful friends. However that may be, it is certain that the poems are based on Virgil's own experience, whether in connection with his own farm or with those of his friends; and they express, with a poignant pathos that has come to be regarded as specially Virgilian, the sorrow of the dispossessed.)

But one eclogue in particular stands out as having relevance to the contemporary situation, and this is the fourth (sometimes called the Messianic, because it was later regarded as prophetic of Christianity). It is an elevated poem, prophesying in sonorous and mystic terms the birth of a child who will bring back the Golden Age, banish sin, and restore peace. It was clearly written at a time when the clouds of civil war seemed to be lifting; it can be dated firmly to 41–40 BC, and it seems most likely that Virgil refers to an expected child of the triumvir Antony and his wife Octavia, sister of Octavian. But, though a specific occasion may be allocated to the poem, it goes beyond the particular and, in symbolic terms, presents a vision of world harmony, which was, to some extent, destined to be realized under Augustus.

Composition of the *Georgics*

One of the most disastrous effects of the civil wars—and one of which Virgil, as a countryman, would be most intensely aware—was the depopulation of rural Italy. The farmers had been obliged to go to the war, and their farms fell into neglect and ruin as a result. The *Georgics,* composed between 37 and 30 BC (the final period of the civil wars), is a superb plea for the restoration of the traditional agricultural life of Italy. In form it is didactic, but, as Seneca later said, it was written "not to instruct farmers but to delight readers." The practical instruction (about plowing, growing trees, tending cattle, and keeping bees) is presented with vivid insight into nature, and it is interspersed with highly wrought poetical digressions on such topics as the beauty of the Italian countryside (Book

By courtesy of the Musee Le Bardo, Tunis

Virgil (centre) holding a scroll with a quotation from the *Aeneid,* with the epic Muse (left) and the tragic Muse (right), Roman mosaic, 2nd–3rd century AD. In the Musée Le Bardo, Tunis.

II. line 136 ff.) and the joy of the farmer when all is gathered in (II.458 ff.).

The *Georgics* is dedicated (at the beginning of each book) to Maecenas, one of the chief of Augustus' ministers, who was also the leading patron of the arts. By this time Virgil was a member of what might be called the court circle, and his desire to see his beloved Italy restored to its former glories coincided with the national requirement of resettling the land and diminishing the pressure on the cities. It would be wrong to think of Virgil as writing political propaganda; but equally it would be wrong to regard his poetry as unconnected with the major currents of political and social needs of the time. Virgil was personally committed to the same ideals as the government.

In the year 31 BC, when Virgil was 38, Augustus (still known as Octavian) won the final battle of the civil wars at Actium against the forces of Antony and Cleopatra and from that time dates the Augustan Age. Virgil, like many of his contemporaries, felt a great sense of relief that the senseless civil strife was at last over and was deeply grateful to the man who had made it possible. Augustus was anxious to preserve the traditions of the republic and its constitutional forms, but he was in fact sole ruler of the Roman world. He used his power to establish a period of peace and stability and endeavoured to reawaken in the Romans a sense of national pride and a new enthusiasm for their ancestral religion and their traditional moral values, those of bravery, parsimony, duty, responsibility, and family devotion. Virgil, too, as a countryman at heart, felt a deep attachment to the simple virtues and religious traditions of the Italian people. All his life he had been preparing himself to write an epic poem (regarded then as the highest form of poetic achievement), and he now set out to embody his ideal Rome in the *Aeneid,* the story of the foundation of the first settlement in Italy, from which Rome was to spring, by an exiled Trojan prince after the destruction of Troy by the Greeks in the 12th century BC. The theme he chose gave him two great advantages: one was that its date and subject were very close to those of Homer's *Iliad* and *Odyssey,* so that he could remodel episodes and characters from his great Greek predecessor; and the other was that it could be brought into relationship with his contemporary Augustan world by presenting Aeneas as the prototype of the Roman way of life (the last of the Trojans and the first of the Romans). Moreover, by the use of prophecies and visions and devices such as the description of the pictures on Aeneas' shield or of the origins of contemporary customs and institutions, it could foreshadow the real events of Roman history. The poem, then, operates on a double time scale; it is heroic and yet Augustan.

Virgil's attachment to simple virtues and traditions

The enthusiasm that Virgil felt for the reborn Rome promised by Augustus' regime is often reflected in the poem. The sonorous and awe-inspiring prophecy by Jupiter (I.257 ff.), giving a picture of Rome's divinely inspired destiny, has a moving patriotic impact: "To these I set no bounds in space or time—I have given them rule without end" (278–279); and again, under Augustus, "Then shall the harsh generations be softened, and wars shall be laid aside" (291). The speech ends with a memorable image depicting the personified figure of Frenzy in chains, gnashing its bloodstained teeth in vain. At the end of the sixth book, Aeneas visits the underworld, and there pass before his eyes the figures of heroes from Roman history, waiting to be born. The ghost of his father (Anchises) describes them to him and ends by defining the Roman mission as one concerned with government and civilization (compared with the Greek achievement in art and literature and theoretical science). "Rule the people with your sway, spare the conquered, and war down the proud": this is the vision of Rome's destiny that the emperor Augustus and the poet Virgil had before them—that Rome was divinely appointed first to conquer the world in war and then to spread civilization and the rule of law among the peoples. As Horace told the Romans in one of his odes, "Because you are servants of the gods, you are masters on earth."

The vision of Rome that the *Aeneid* expresses is a noble one, but the real greatness of the poem is due to Virgil's awareness of the private, as well as the public, aspects of human life. The *Aeneid* is no panegyric; it sets the achievements and aspirations of the giant organization of Roman governmental rule in tension with the frustrated hopes and sufferings of individuals. The most memorable figure in the poem—and, it has been said, the only character to be created by a Roman poet that has passed into world literature—is Dido, Queen of Carthage, opponent of the Roman way of life. In a mere panegyric of Rome, she could have been presented in such a way that Aeneas' rejection of her would have been a victory to applaud; but, in fact, in the fourth book she wins so much sympathy that the reader wonders whether Rome should be bought at this price. Again, Turnus, who opposes Aeneas when he lands in Italy, resists the invader who has come to steal his bride. It is clear that Turnus is a less civilized character than Aeneas—but in his defeat Virgil allows him to win much sympathy. These are two examples of the tension against Roman optimism; in many other ways, too, Virgil throughout the poem explores the problems of suffering and the pathos of the human situation. Yet in the end, Aeneas endures and continues to his goal; his devotion to duty (*pietas*) prevails, and the Roman reader would feel that this should be. "So great a task it was to found the Roman nation" (I.33).

The greatness of the *Aeneid*

The *Aeneid* occupied Virgil for 11 years and, at his death, had not yet received its final revision. In 19 BC, planning to spend a further three years on his poem, he set out for Greece—doubtless to obtain local colour for the revision of those parts of the *Aeneid* set in Greek waters. On the voyage he caught a fever and returned to Italy but died soon after arrival at Brundisium. Whether the *Aeneid* would have undergone major changes cannot be guessed; the story goes that Virgil's dying wish was for his poem to be burned, but that this request was countermanded by the order of Augustus. As it stands, the poem is a major monument both to the national achievements and ideals of the Augustan Age of Rome and to the sensitive and lonely voice of the poet who knew the "tears in things" as well as the glory.

Influence and reputation. Virgil's poetry immediately became famous in Rome and was admired by the Romans for two main reasons—first, because he was regarded as their own national poet, spokesman of their ideals and achievements; second, because he seemed to have reached the ultimate of perfection in his art (his structure, diction, metre). For the latter reason, his poems were used as school textbooks, and the 1st-century Roman critic and teacher Quintilian recommended that the educational curriculum should be based on Virgil's works. A few years after his death, Virgil was being imitated and echoed by the younger poet Ovid, and this process continued throughout the Silver Age. The study of Virgil in the schools has lasted as long as Latin has been studied. By the 4th century a new reason for admiration was gaining ground: the store of wisdom and knowledge discovered by scholars in Virgil's poems—for which he was saluted not only as a poet but as a repository of information. This aspect figures largely in the writings of the writer and philosopher Macrobius (fl. *c.* AD 400), those of Virgil's commentator Servius of the late 4th and early 5th century, and those of many later writers. Allegorical interpretations began to gain ground and, under Christian influence, became especially widespread throughout the Middle Ages. The two main bases for Christian allegorization were the fourth eclogue, believed to be a prophecy of the birth of Christ, and the near-Christian values expressed in the *Aeneid,* especially in its hero, a man devoted to his divine mission. The culmination of this view is Virgil's place of honour in Dante's *Divine Comedy* as the poet's guide through Hell and Purgatory up to the very gates of Paradise.

Later imitations of Virgil

Virgil's influence on English literature has been enormous. He was Edmund Spenser's constant inspiration for the fanciful beauty of *The Faerie Queene.* The *Aeneid* was the model for John Milton's *Paradise Lost* not only in epic structure and machinery but also in style and diction. In the English Augustan age, John Dryden and countless others held that Virgil's poetry had reached the ultimate perfection of form and ethical content. There was some reaction against him in the Romantic period, but

the Victorians, such as Matthew Arnold and Alfred, Lord Tennyson, rediscovered in full measure that sensitivity and pathos that the Romantics had complained that Virgil lacked.

MAJOR WORKS

Eclogae (42–37 BC; *Eclogues*), comprising 10 poems also known as the *Bucolica; Georgica* (37–30 BC; *Georgics*), comprising 4 books of poems on farming and rural life in Italy; and *Aeneis* (30–19 BC; *Aeneid*), an epic poem in 12 books.

Many English translations of Virgil's works have been made, including verse translations by Gavin Douglas (1553), John Dryden (1697), C. Day Lewis (1940–63), and Rolfe Humphries (1951). Notable among prose translations are those by J.W. Mackail (2nd ed. rev., 1908–14), H.R. Fairclough ("Loeb Series," 1934–35); and W.F. Jackson Knight (1956). There is a French translation of all the works in the "Budé Series" (1926–56).

BIBLIOGRAPHY. Detailed bibliographical information about Virgil may be found in the lists in *L'Année philologique* (annual). In English there are bibliographical surveys by G.E. DUCKWORTH, *Classical World,* vol. 51 (1957–58) and vol. 57 (1963–64), reissued as *Recent Work on Vergil,* 2 vol. (1964); the summaries in *Vergilius* (annual), the journal of the Vergilian Society (U.S.); the reviews of important modern works in the *Proceedings of the Virgil Society* (annual), the journal of the Virgil Society (U.K.); and the survey by R.D. WILLIAMS, *Virgil* (1967).

Texts: The best texts of Virgil's work (including discussion of the manuscripts) are those of R.A.B. MYNORS, "Oxford Classical Texts" (1969), replacing Hirtzel's 1900 edition; R. SABBADINI (1930; reprinted with minor alterations by L. CASTIGLIONI, 1945–52); E. DE SAINT-DENIS (*Eclogues,* 1942; *Georgics,* 1956); and GOELZER-DURAND-BELLESSORT, *Aeneid,* 2 vol. (1925–36), in the "Budé Series."

Commentaries: The commentaries by SERVIUS have been edited by GEORG THILO and HERMANN HAGEN, *Servii Grammatici qvi fervntvr in Vergilii carmina commentarii,* 3 vol. (1881–1902); and by EDWARD K. RAND *et al.* (*Aeneid,* books 1 and 2, 1946; books 3 to 5, 1965); JAMES HENRY, *Aeneidea,* 4 vol. (1873–92, reprinted 1972); CONINGTON-NETTLESHIP-HAVERFIELD, 3 vol. (1858–98); THOMAS E. PAGE, *The Aeneid of Virgil* (1894–1900); and R.D. WILLIAMS, *The Aeneid of Virgil,* 2 vol. (1972–73).

Biography and criticism: For such biographical facts as are known about Virgil, see the introduction to JOHN W. MACKAIL, *Aeneid* (1930); and W.F. JACKSON KNIGHT, *Roman Vergil,* rev. ed. (1966, reissued 1971). GILBERT A. HIGHET, *Poets in a Landscape* (1957, reprinted 1979), gives a popular and readable account. The ancient sources are collected in *Vitae Vergilianae Antiquae* (1954, reissued 1966). An excellent critical evaluation of Virgil's *Aeneid* is W.A. CAMPS, *An Introduction to Virgil's Aeneid* (1969). Other important works are T.R. GLOVER, *Virgil,* 7th ed. (1942, reissued 1969); C.M. BOWRA, *From Virgil to Milton* (1945, reissued 1967); BROOKS OTIS, *Virgil: A Study in Civilised Poetry* (1963); KENNETH QUINN, *Virgil's Aeneid: A Critical Description* (1968); D.R. DUDLEY (ed.), *Virgil* (1969); and J. WILLIAM HUNT, *Forms of Glory* (1973), an introductory study of the *Aeneid.* On the *Eclogues,* see HERBERT J. ROSE, *The Eclogues of Vergil* (1942); and WILLIAM BERG, *Early Virgil* (1974). On the *Georgics,* see L.P. WILKINSON, *The Georgics of Virgil* (1969, reissued 1978); MICHAEL C.J. PUTNAM, *Virgil's Poem of the Earth* (1979); and GARY B. MILES, *Virgil's Georgics: A New Interpretation* (1980). On Virgil's influence, see G.A. HIGHET, *The Classical Tradition* (1949, reissued 1967); R.R. BOLGAR, *The Classical Heritage and Its Beneficiaries* (1954, reissued 1973); D. COMPARETTI, *Vergil in the Middle Ages* (1895); J.H. WHITFIELD, "Virgil into Dante"; and R.D. WILLIAMS, "Changing Attitudes to Virgil," chapters 4 and 5 in D.R. DUDLEY (cited above).

(R.D.W./Ed.)

Viruses

Virus is a Latin word meaning "slimy liquid" or "poison." The earliest indications of the biologic nature of viruses came from studies in 1892 by the Russian scientist Dmitry I. Ivanovsky and in 1898 by the Dutch scientist Martinus W. Beijerinck. Beijerinck first surmised that the virus under study was a new kind of infectious agent, which he designated *contagium vivum fluidum,* meaning that it was a live, reproducing organism that, nevertheless, differed from other organisms. Both of these investigators found that a disease of tobacco plants could be transmitted by an agent, later called tobacco mosaic virus, passing through a minute filter that would not allow the passage of bacteria. This virus and those subsequently isolated would not grow on an artificial medium and were not visible under the light microscope. In independent studies in 1915 by the British investigator Frederick W. Twort and in 1917 by the French-Canadian scientist Félix H. d'Hérelle, lesions in cultures of bacteria were discovered and attributed to an agent called bacteriophage (eater of bacteria), now known to be viruses that specifically infect bacteria.

The unique nature of these organisms meant that new methods and alternative models had to be developed to study and classify them. The study of viruses confined exclusively or largely to humans, however, posed the formidable problem of finding a susceptible animal host. In 1933 the British investigators Wilson Smith, Christopher H. Andrewes, and Patrick P. Laidlaw were able to transmit influenza to ferrets, and the influenza virus was subsequently adapted to mice. In 1941 the American scientist George K. Hirst found that influenza virus grown in tissues of the chicken embryo could be detected by its capacity to agglutinate (draw together) red blood cells.

A significant advance was made by the American scientists John Enders, Thomas Weller, and Frederick Robbins, who, in 1949, developed the technique of culturing cells on glass surfaces; cells could then be infected with the viruses that cause poliomyelitis (poliovirus) and other diseases. (Until this time the poliovirus could be grown only in the brains of chimpanzees or the spinal cords of monkeys.) Culturing cells on glass surfaces opened the way for the diagnosis of diseases caused by viruses identified by their effects on cells (cytopathogenic effect) and by the presence of antibodies to them in the blood. Cell culture then led to the development and production of vaccines (preparations used to elicit immunity against a disease), such as the poliovirus vaccine.

Scientists were soon able to detect the number of bacterial viruses in a culture vessel by measuring their ability to break apart (lyse) adjoining bacteria, which resulted in a clearing, or "plaque," in an area of bacteria (lawn) overlaid with an inert gelatinous substance, called agar. The American scientist Renato Dulbecco in 1952 applied this technique to measuring the number of animal viruses that could produce plaques in layers of adjoining animal cells overlaid with agar. In the 1940s the development of the electron microscope permitted individual virus particles to be seen for the first time, leading to the classification of viruses and giving insight into their structure.

Advancements that were made in chemistry, physics, and molecular biology since the 1960s have revolutionized the study of viruses. For example, electrophoresis on gel substrates gave a deeper understanding of the protein and nucleic-acid composition of viruses. More sophisticated immunologic procedures, including the use of monoclonal antibodies directed to specific antigenic sites on proteins, gave a better insight into the structure and function of viral proteins. The progress made in the physics of crystals that could be studied by X-ray diffraction provided the high resolution required to discover the basic structure of minute viruses. Applications of new knowledge about cell biology and biochemistry helped to determine how viruses use their host cells for synthesizing viral nucleic acids and proteins.

The revolution that took place in the field of molecular biology allowed the genetic information encoded in nucleic acids of viruses, which enables viruses to reproduce, synthesize unique proteins, and alter cellular functions, to be studied. In fact, the chemical and physical simplicity of viruses has made them an incisive experimental tool for probing the molecular events involved in certain life processes.

This article discusses the fundamental nature of viruses: what they are; how they cause infection; and how they may ultimately cause disease or bring about the death of their host cells. For more detailed treatment of specific viral diseases, see the article INFECTIOUS DISEASES.

For coverage of related topics in the *Macropædia* and *Micropædia,* see the *Propædia,* sections 313 and 423.

The article is divided into the following sections:

General features 502
  Definition 502
  Virus host range and distribution 503
  Size and shape 503
The cycle of infection 505
Viral DNA integration 506
  Lysogeny 506
  Malignant transformation 506
Disease 507
  Infectious patterns 508
  Latency 508
  Chronic and slowly progressive diseases 508
  Prevention 508
  Treatment 509
Evolution 509
Classification 509
  Distinguishing taxonomic features 509
  Annotated classification 510
Bibliography 511

GENERAL FEATURES

Definition. Viruses occupy a special taxonomic position: they are not plants, animals, or prokaryotic bacteria (single-cell organisms without defined nuclei), and they are generally placed in their own kingdom. In fact, viruses should not even be considered organisms, in the strictest sense, because they are not free-living (*i.e.,* they cannot reproduce and carry on metabolic processes without a host cell).

All viruses contain nucleic acid, either DNA (deoxyribonucleic acid) or RNA (ribonucleic acid), and protein. The nucleic acid encodes the genetic information unique for each virus. The infective, extracellular (outside the cell) form of a virus is called the virion. It contains at least one unique protein synthesized by specific genes (specific sequences of nucleotides that code for functions of heredity) in the nucleic acid of that virus. In virtually all viruses, at least one of these proteins forms a shell (called a capsid) around the nucleic acid. Certain viruses also have other proteins internal to the capsid; some of these proteins act as enzymes, often during the synthesis of viral nucleic acids. Viroids (meaning "viruslike") are disease-causing organisms that contain only nucleic acid and have no structural proteins.

Parasitic actions

Viruses are quintessential parasites; they depend on the host cell for almost all of their life-sustaining functions. Unlike true organisms, viruses cannot synthesize proteins because they lack ribosomes (cell organelles) for the translation of messenger RNA (mRNA; a complementary copy of the nucleic acid of the nucleus, which associates with ribosomes and directs protein synthesis) into proteins. Viruses must use the ribosomes of their host cells to translate viral mRNA into viral proteins.

Viruses are also energy parasites; unlike cells, they cannot generate or store energy in the form of adenosine triphosphate. The virus derives energy, and all other metabolic functions, from the host cell. The invading virus uses the nucleotides and amino acids of the host cell to synthesize its nucleic acids and proteins, respectively. Some viruses use the lipids and sugar chains of the host cell to form their membranes and glycoproteins.

The true infectious part of any virus is its nucleic acid, either DNA or RNA but never both. In many viruses, but not all, the nucleic acid alone, stripped of its protein coat (capsid), can infect (transfect) cells, although considerably less efficiently than can the intact virions.

Capsid functions

The virion capsid has three functions: (1) to protect the viral nucleic acid from digestion by certain enzymes (nucleases); (2) to furnish sites on its surface that recognize and attach (adsorb) the virion to receptors on the surface of the host cell; and (3), in some viruses, to provide proteins that form part of a specialized component which enables the virion to penetrate through the cell surface membrane or, in special cases, to inject the infectious nucleic acid into the interior of the host cell.

Virus host range and distribution. Logic originally dictated that viruses be identified on the basis of the host they infect. This is justified in many cases, but not in others, and the host range and distribution of viruses are only one criterion for their classification. It is still traditional to divide viruses into three categories: those that infect animals, plants, or bacteria.

Virtually all plant viruses are transmitted by insects or other vectors (organisms that transmit the virus from one host to another) that feed on plants. The hosts of animal viruses vary from protozoa (single-celled animal organisms) to humans. Many viruses infect either invertebrate animals or vertebrates, and some infect both. Certain viruses that cause serious diseases of animals and humans are carried by arthropods (insects and spiders). These vector-borne viruses, which cause human diseases such as yellow fever and encephalitis, multiply both in the invertebrate vector and the vertebrate host.

Certain viruses are limited in their host range to the various orders of vertebrates. Some viruses appear to be adapted for growth only in cold-blooded vertebrates (poikilotherms: fishes, reptiles, and amphibians), possibly because they cannot reproduce except at low temperatures. Other viruses are limited in their host range to warm-blooded vertebrates (birds and mammals).

Size and shape. The amount and arrangement of the proteins and nucleic acid of viruses determine their size and shape. The nucleic acid and proteins of each class of viruses assemble themselves into a structure called a nucleoprotein, or nucleocapsid. Some viruses have more than one layer of protein surrounding the nucleic acid; still others have a lipoprotein membrane (called an envelope), derived from the membrane of the host cell, that surrounds the nucleocapsid core. The protein and nucleic acid constituents have properties unique for each class of virus; when assembled, they determine the size and shape of the virus for that specific class.

Size ranges

Viruses vary in diameter from 20 nanometres (0.0000008 inch) to 250–400 nanometres. Only the largest and most complex viruses can be seen under the light microscope at the highest resolution. Any determination of the size of a virus must also take into account its shape, since different classes of viruses have distinctive shapes.

Shapes of viruses are predominantly of two kinds: rods, or filaments, so called because of the linear array of the nucleic acid and the protein subunits; and spheres, which are actually 20-sided (icosahedral) polygons. Most plant viruses are small and are either filaments or polygons, as

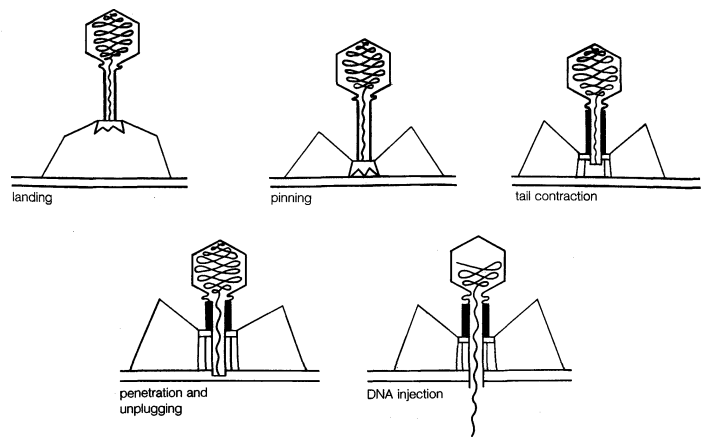


Figure 1: General structure of T4 bacteriophage and a model of its mode of attachment to, and injection of its DNA into, a bacterial cell.

From C.K. Mathews in H. Fraenkel-Conrat and R.R. Wagner (eds.), *Comprehensive Virology* (1977), vol. 7; Plenum Press

are many bacterial viruses. The larger and more complex bacteriophages, however, contain as their genetic information double-stranded DNA and combine both filamentous and polygonal shapes.

Figure 1 shows the classical T4 bacteriophage, composed of a polygonal head, which contains the DNA genome, and a special-function rod-shaped tail of long fibres. Structures such as these are unique to the bacteriophages.

The animal viruses exhibit extreme variation in size and shape (Figure 2). The smallest animal viruses are those belonging to the family Parvoviridae, which measure about 20 nanometres in diameter, and the family Picornaviridae, which measure about 30 nanometres in diameter. Viruses of these two families are polygonal (icosahedral) in shape and contain nucleic acids with limited genetic informa-

From F. Fenner *et al.*, *Veterinary Virology* (1987), Academic Press

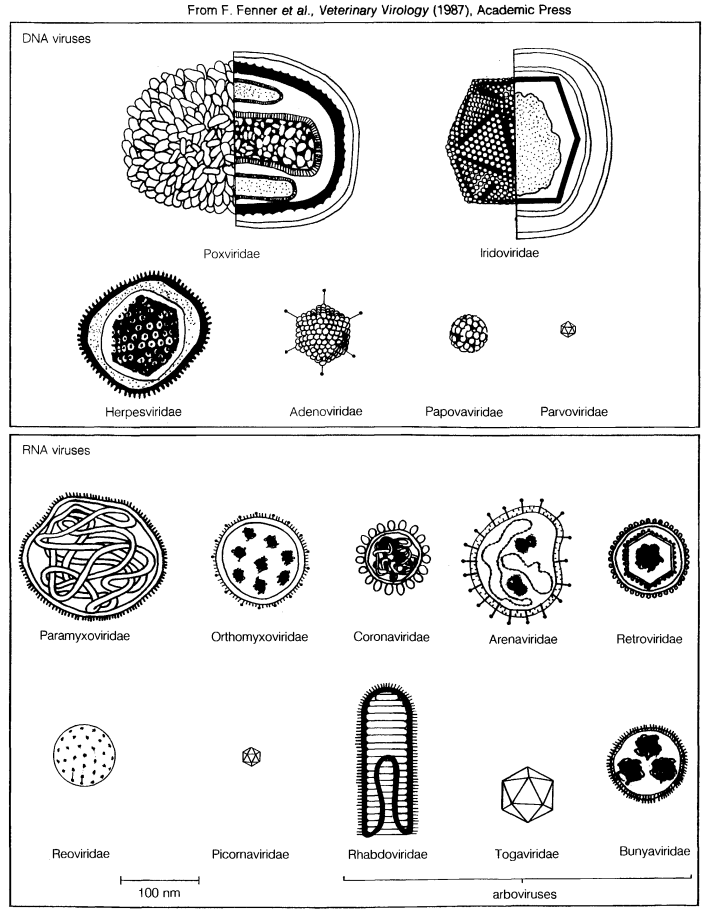


Figure 2: Shapes and relative sizes of animal viruses of the major families.

tion. Viruses of the family Poxviridae are about 250 to 400 nanometres in their longest dimension, and they are neither polygons nor filaments. Poxviruses are structurally more complex than, and resemble, simple bacteria. The only animal viruses that have rod-shaped (helical) nucleocapsids are those enclosed in a membrane (envelope); these viruses are found in the families Paramyxoviridae, Orthomyxoviridae, Coronaviridae, and Rhabdoviridae. Not all enveloped viruses contain helical nucleocapsids, however; those of the families Herpesviridae, Retroviridae, and Togaviridae have polygonal nucleocapsids. Most enveloped viruses appear to be spherical, although the rhabdoviruses are elongated cylinders, often bullet-shaped.

Criteria of classification

The criteria used for classifying viruses into families and genera are primarily based on three structural considerations: (1) the type and size of their nucleic acid; (2) the shape and size of the capsids; and (3) the presence of a lipid envelope, derived from the host cell, surrounding the viral nucleocapsid.

The nucleic acid. As is true in all forms of life, the nucleic acid of each virus encodes the genetic information for the synthesis of proteins, both as structural components and as enzymes. In almost all free-living organisms, the genetic information is in the form of double-stranded DNA arranged as a spiral lattice joined at the bases along the length of the molecule (a double-helix) as described in 1953 by James Watson and Francis Crick. Double-stranded DNA also carries the genetic information for some viruses, although probably in most the genetic information of the virion is RNA.

In extracellular virions, the DNA and RNA are called genomes, and they can be double- or single-stranded. In all double-stranded DNA viruses, the nucleic acid is a single large molecule. Most, if not all, of the double-stranded RNA viruses have segmented genomes; each segment usually is a single gene, which encodes the information for synthesizing a single protein. Those DNA viruses in which the genomic DNA is single-stranded are usually small, with limited genetic information. Some single-stranded DNA viruses are composed of two populations of virions, each encapsidating complementary single-stranded DNA of polarity opposite to that of the other.

Positive- and negative-stranded RNA

The virions of most plant viruses and many animal and bacterial viruses are composed of single-stranded RNA. In most of these viruses, the genomic RNA is termed a plus strand because the genomic RNA acts as mRNA for direct synthesis (translation) of viral protein. Several large families of animal viruses, and one that includes plant viruses (the Rhabdoviridae), however, contain genomic single-stranded RNA, termed a negative strand, which is complementary to mRNA. All of these negative-strand RNA viruses have an enzyme, called an RNA-dependent RNA polymerase (transcriptase), which must first catalyze the synthesis of complementary mRNA from the virion genomic RNA before viral protein synthesis can occur. These variations in the nucleic acids of viruses form one central criterion for classification of all viruses.

A distinctive large family of single-stranded RNA viruses is called Retroviridae; the RNA of these viruses is positive, but the viruses are equipped with an enzyme, called a reverse transcriptase, that copies the single-stranded RNA to form double-stranded DNA.

The protein capsid. The protein capsid provides the second major criterion for the classification of viruses. The capsid surrounds the virus and is composed of a finite number of protein subunits known as capsomers, which usually associate with, or are found close to, the virion nucleic acid.

Major viral classes

There are two major classes of viruses based on the protein capsid: (1) those in which a single (or segmented) linear nucleic acid molecule with two free ends is essentially completely extended or somewhat coiled (a helix), and (2) those in which the nucleic acid, which may or may not be a covalently closed circle, is wound tightly into a condensed configuration, like a ball of yarn. These two classes of virus assume in the first case a long, extended rod-like structure and in the second case a symmetrical polygonal structure. The helical nucleic acids of rod-shaped viruses are tightly enclosed in a sheath of protein molecules.

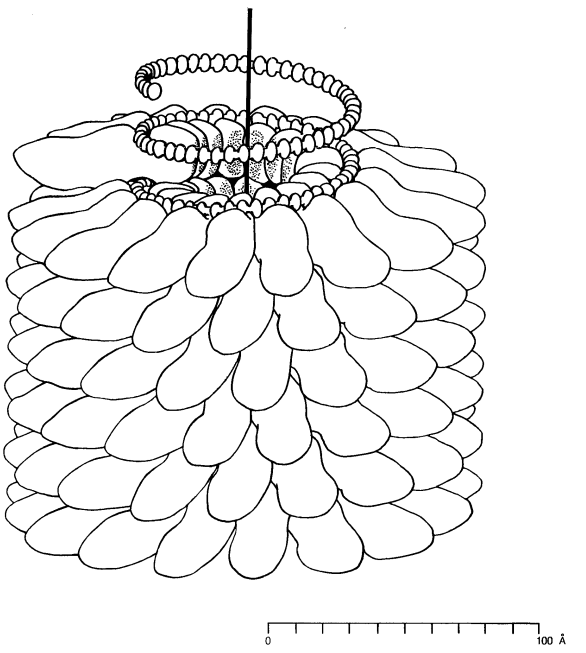


Figure 3: Schematic structure of the tobacco mosaic virus. The cut-away section shows the helical ribonucleic acid associated with protein molecules in a ratio of three nucleotides per protein molecule.
From A. Klug and D.L.D. Caspar, *Advances in Virus Research* (1960); Academic Press

By far the best-studied example of a helical rod-shaped virus is the tobacco mosaic virus, which was crystallized by Wendell Stanley in 1935. Figure 3 shows a model of the tobacco mosaic virus, the genome of which is single-stranded RNA encased by 2,130 molecules of a single protein; there are $16\frac{1}{3}$ protein molecules for each turn of the RNA helix in the ratio of three nucleotides for each protein molecule.

Under the right environmental conditions, a liquid suspension of viral RNA and protein molecules will assemble themselves into a perfectly formed and fully infectious virus. The length of the virus capsid is determined by the length of the nucleic acid molecule, which is the framework for the assembly of the capsid protein. The various helical viruses have different lengths and widths depending on the size of the nucleic acid as well as on the mass and shape of the protein molecule. Some of these helical viruses form rigid, and others flexuous, rods, depending on the properties of the assembled proteins.

Shapes of polygonal viruses

As shown in Figure 2, polygonal viruses vary greatly in size, from 20 nanometres to 150 nanometres in diameter, essentially proportional to the size of the nucleic acid molecule coiled up inside the virion. Most, if not all, of the polygonal viruses are 20-sided (an icosahedron); like a geodesic dome, they are formed by equilateral triangles, in this case 20. Each triangle is composed of protein subunits (capsomers), often in the form of hexons (multiples of six) that are the building blocks of the capsid. There are 12 vertices (the apical junctions of these 20 triangles), each comprising a penton (five subunits). These icosahedral virions have three axes of five-, three-, and twofold rotational symmetry (Figure 4). The number of subunits (capsomers) is a basis for taxonomic classification of these virus families. Certain icosahedral viruses, usually those

From W.K. Joklik, H.P. Willett, and D.B. Amos, *Zinsser Microbiology*, (1980), 17th ed., Appleton and Lange, Norwalk, Conn. (formerly Appleton-Century-Crofts, New York)

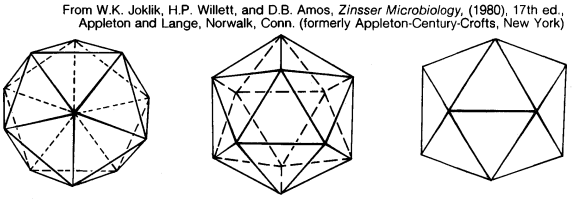


Figure 4: A virus icosahedron (20-sided structure) shown in the (left) fivefold, (centre) threefold, and (right) twofold axes of symmetry. Edges of the upper and lower surfaces are drawn in solid and broken lines, respectively.

that are more complex, contain internal proteins adhering to the nucleic acid that are not accessible at the surface of the virions.

Structure of viral envelopes

The lipoprotein envelope. Surrounding viruses of either helical or icosahedral symmetry are lipoprotein envelopes, unit membranes of two lipid layers interspersed with protein molecules (lipoprotein bilayer). These viral membranes are composed of phospholipids and neutral lipids (largely cholesterol) derived from the cell membrane during the process known as the budding of the virus from the outermost cell membrane. Virtually all the proteins of the cell membrane, however, are replaced by proteins of viral origin during the process of budding. Although all of the viral envelope lipids originate from the cell, their relative proportions vary from those in the cell membrane because the viral proteins select only certain lipids during budding.

Associated with the viral membrane are "integral" glycoproteins, which completely traverse the lipid bilayer, and "peripheral" matrix proteins, which line the inner surface. The glycoproteins contain regions of amino acids that, in the first step of viral infection, recognize host-cell receptors. Matrix proteins appear to function in the selection of regions of the cell membrane to be used for the viral membrane.

THE CYCLE OF INFECTION

Viruses can reproduce only within a host cell. The parental virus (virion) gives rise to numerous progeny, usually genetically and structurally identical to the parent virus. The actions of the virus depend both on its destructive tendencies toward a specific host cell and on environmental conditions. In the vegetative cycle of viral infection, multiplication of progeny viruses can be rapid. This cycle of infection often results in the death of the cell and the release of many virus progeny. Certain viruses, particularly bacteriophages, are called temperate because the infection does not immediately result in cell death. The viral genetic material (genome) remains dormant or is actually integrated into the genome of the host cell. Cells infected with temperate viruses are called lysogenic because the cells tend to be lysed (broken down) when they encounter some chemical or physical factor, such as ultraviolet light. In addition, many animal and plant viruses, the genetic information of which is not integrated into the host DNA, may lie dormant in tissues for long periods of time without causing much, if any, tissue damage. Viral infection does not always result in cell death or tissue injury; in fact, most viruses lie dormant in tissue without ever causing pathological effects, or they do so only under other, often environmental, provocations.

Although the reproductive pathways of different viruses vary considerably, there are certain basic principles and a particular series of events in the cycle of infection for most, if not all, viruses. The first step in the cycle of infection is that the invading parental virus (virion) must

Adsorption

attach to the surface of the host cell (adsorption). In the second step, the intact virion either penetrates the outer membrane and enters the cell's interior (cytoplasm), or it injects the genetic material of the virus into the interior of the cell while the protein capsid (and envelope, if present) remains at the cell surface. In the case of whole-virion penetration, a subsequent process (uncoating) liberates the genetic material of the virus from the capsid and envelope, if present. In either case, the viral genetic material cannot begin the process of protein synthesis until it emerges from the capsid or envelope.

Certain bacterial viruses, such as the T4 bacteriophage, have evolved an elaborate process of infection: following adsorption and firm attachment of the virus's tail to the bacterium surface by means of proteinacious "pins," the musclelike tail contracts and the tail plug penetrates the cell wall and underlying membrane and injects virus (phage) DNA into the cell (see Figure 1). Other bacteriophages penetrate the cell membrane by different means, such as injecting the nucleic acid through the male (sex) pili of the bacterium. In all bacterial viruses, penetration transmits the viral nucleic acid through a rigid bacterial cell wall.

Plant cells also have rigid cell walls, which plant viruses cannot ordinarily penetrate. Plant viruses, however, have not evolved their own systems for injecting nucleic acids into host cells, and so they are transmitted by the proboscis of insects that feed on plants. In the laboratory, plant viruses penetrate plant cells if the cell walls have been abraded with sand paper or if cell protoplasts (plasma membrane, cytoplasm, and nucleus) are devoid of walls.

Penetration of animal cells by viruses involves different processes because animal cells are not enclosed by walls but by a flexible lipoprotein bilayer membrane. Most animal viruses, whether or not they are encased in lipid envelopes, penetrate cells in an intact form by a process called endocytosis (Figure 5). The membrane invaginates and engulfs a virus particle adsorbed to a cell, presumably in an area of the membrane called a coated pit, which is lined by a special protein known as clathrin. As the coated pit invaginates through the membrane, it is pinched off in the cytoplasm to form a coated vesicle. The coated vesicle is fused with cytoplasmic endosomes (membrane-enclosed vesicles) and then with cell organelles called lysosomes, which are membrane-enclosed vesicles containing enzymes. In an acidic environment, the membrane of an enveloped virus fuses with the endosome membrane and the viral nucleocapsid is released into the cytoplasm. Nonenveloped viruses presumably undergo a similar process, by which the protein capsid is degraded, releasing the naked viral nucleic acid into the cytoplasm.

Endocytosis

K. Simons, H. Garoff, and A. Helenius, "How an Animal Virus Gets into and out of Its Host Cell," copyright © 1982, Scientific American, Inc., all rights reserved

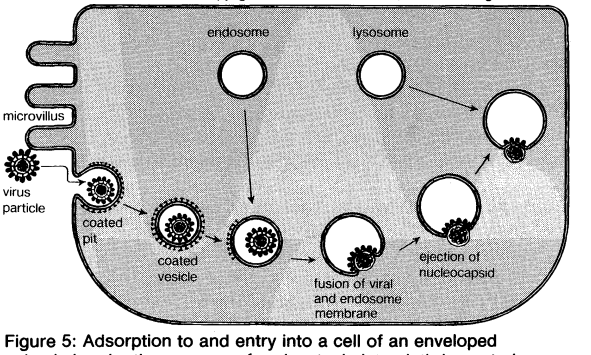


Figure 5: Adsorption to and entry into a cell of an enveloped animal virus by the process of endocytosis into clathrin-coated vesicles, which fuse with large vesicles (endosomes and lysosomes). The process triggered by the viral glycoprotein results in fusion and release of the viral nucleocapsid into the cytoplasm.

The order of the stages of viral replication that follows the uncoating of the genome varies for different virus classes. For many virus families the third step in the cycle of infection is transcription of the genome of the virus to produce viral mRNA, followed by the fourth step, translation of viral mRNA into proteins. For those viruses in which the genomic nucleic acid is an RNA that can serve as a messenger (*i.e.,* positive-strand RNA viruses), the third step is the translation of the RNA to form viral proteins; some of these newly synthesized viral proteins are enzymes that synthesize nucleic acids (polymerases), which carry out a fourth step, the transcription of more mRNA from the viral genome. For the more complicated DNA viruses, such as adenoviruses and herpesviruses, some regions of the genome synthesize "early" mRNAs, which are translated into polymerases that initiate the transcription of "late" regions of the DNA into mRNAs, which are then translated into structural proteins.

"Early" mRNAs

Regardless of how the third and fourth steps proceed, the fifth step in the cycle of infection is replication (reproduction of the parental genome to make progeny genomes). The sixth step is the assembly of the newly replicated progeny genomes with structural proteins to make fully formed progeny virions. The seventh, and last, step is the release of progeny virions by lysis of the host cell, by extrusion, or by budding, depending on the nature of the virus. In a host animal or cell culture this seven-step process may be repeated many times; the progeny virus particles (virions) released from the original site of infection are then transmitted to other sites or to other individuals.

VIRAL DNA INTEGRATION

Lysogeny. Many bacterial and animal viruses lie dormant in the infected cell, and their DNA may be integrated into the DNA of the host cell chromosome. Integrated DNAs of the virus replicate as the cell genome replicates; after cell division, the integrated viral DNA is duplicated and usually distributed equally to the two cells that result. The bacteria that carry the noninfective precursor phage (prophage) remain healthy and continue to grow until they are stimulated by some perturbing factor, such as ultraviolet light. The prophage DNA is then excised from the bacterial chromosome and the phage replicates, producing many progeny phage and lysing the host bacterial cell. This process, originally discovered in 1950 by the French microbiologist André Lwoff in temperate bacteriophages, is called lysogeny.

The classical example of a temperate bacteriophage is called lambda (λ) virus, which readily causes lysogeny in certain species of the bacterium *Escherichia coli.* The DNA of the λ bacteriophage is integrated into the DNA of the *E. coli* host chromosome at specific regions, called attachment sites.

Attachment sites

The integrated phage DNA, called a prophage, is the inherited, noninfectious form of the virus; it contains a gene that represses the lytic functions of the phage and thus assures that the host cell will continue to replicate the phage DNA and its own and that it will not be destroyed by the virus. Ultraviolet light, or other factors that stimulate the replication of DNA in the host cell, causes the formation of a recA protease, an enzyme that breaks apart the λ phage repressor and induces λ phage replication and, eventually, destruction of the host cell.

Excision of the prophage DNA from the host chromosomal DNA (as an initial step in the synthesis of an infective, lytic virus) sometimes results in the removal of some of the host cell DNA, which is packaged into defective bacteriophages; part of the bacteriophage DNA is removed and replaced at the other end by a gene of the host bacterium. Such a virus particle is called a transducing phage because when it infects a bacterial cell it can transmit the gene captured by λ phage DNA into the next bacterial cell it infects. Transduction by bacteriophages is an efficient means for transferring the genetic information of one bacterial cell to another.

Transducing phage

The λ bacteriophage and other similar phages are called specialized transducing phages because they transmit only those genes adjacent to the λ prophage attachment site. The prophages of other temperate bacteriophages can be integrated almost anywhere on the chromosomes of the bacterial host cell. These are called generalized transducing phages because of the randomness with which they capture and transmit genes from one bacterial cell to another. Transduction of this type is an efficient means of creating bacteria with new gene functions or of replacing those that have been lost by mutation.

Lysogenic conversion

This means of transferring genetic information, called lysogenic conversion, imparts genes with special functions to bacterial cells without such functions. This alteration in genetic functions of bacteria is common in nature. A bacteriophage released by one bacterial cell can infect another bacterial cell and impart this function by lysogenic conversion. This process is an important aspect of the epidemiology (incidence, distribution, and control) of infectious diseases. For example, a species of bacteria called *Corynebacterium diphtheriae* is the causative agent of diphtheria, but only when it contains the prophage of bacteriophage *β,* which codes for the toxin that is responsible for the disease. Similar examples of lysogenic conversion by bacteriophages that code for, or induce, bacterial toxins may be found in group A streptococci that contain a prophage that codes for the toxin causing scarlet fever.

Malignant transformation. A phenomenon analogous to bacterial cell lysogeny occurs in animal cells infected with certain viruses. These animal viruses do not generally cause disease immediately for certain animal cells. Instead, animal cells are persistently infected with such viruses, the DNA of which (provirus) is integrated into the chromosomal DNA of the host cell. In general, cells with integrated proviral DNA are converted into cancer cells, a phenomenon known as malignant transformation. As is the case with bacterial prophages, the transformed animal cell contains no infectious virus but only the integrated provirus DNA, which replicates along with the dividing cell's chromosomes. Therefore, following mitosis of the transformed cell, each new cell receives a copy of the proviral DNA. The hallmark of these transformed animal cells is that their growth is uncontrolled; unlike normal cells, their growth is not inhibited by contact with other cells, and they lose their capacity to adhere (anchor) to certain surfaces.

Among the animal viruses that cause malignant transformation by integration of proviral DNA are several families of DNA viruses and one large family of RNA viruses, the Retroviridae. Viruses of the family Papovaviridae were perhaps the first to be associated with malignancy (causing death or illness) in animals. Polyoma virus is widespread in mice; it can infect other rodents, and it can cause tumours in infected animals. Another virus of the family Papovaviridae is simian virus 40 (SV40), originally isolated from cells of the African green monkey, where it grows rapidly and kills the cells. Infection of rodent or human cells, however, results in an abortive infection (an incompatibility between the virus and the host cell) but sometimes induces malignancy (sarcomas or lymphomas) in the occasional cell that is transformed. Viruses related to polyoma virus and SV40 have been isolated from humans, one of which, the JC virus, appears to be the causative agent of a fatal neurological disease called progressive multifocal leukoencephalopathy. In general, however, the human papovaviruses are not clearly associated with disease.

Abortive infection

Other viruses of the family Papovaviridae include the papillomaviruses, which are also small polygonal viruses containing circular double-stranded DNA. The papillomaviruses are associated with usually benign (nonthreatening) but widespread tumours, called papillomas or polyps, occurring in human skin and the genital tract. Specific papillomaviruses have been identified in humans in common warts and in genital warts (condylomata acuminata).

Certain viruses of the family Adenoviridae (Figure 6), originally found in the tonsils and adenoids of humans, cause malignant transformation in certain cells. This phenomenon of cancer induction under laboratory conditions has been studied widely, but there is no evidence that the common adenoviruses cause cancers in humans. The common viruses of the family Herpesviridae, however, including the common herpes simplex viruses that cause cold sores and the venereal disease genital herpes, are suspected of being causative agents of cancer. Like the adenoviruses, the herpesviruses can cause malignant transformations, and their DNA is integrated into the host cell chromosome. A herpesvirus known as the Epstein-Barr virus causes a frequently fatal childhood cancer called Burkitt's lymphoma as well as the nonmalignant disease infectious mononucleosis. A different herpesvirus causes chicken pox (varicella); the same virus lies latent in the

Herpesviruses

From H. Persson and L. Philipson, *Current Topics in Microbiology and Immunology,* no. 97, (1982), Springer-Verlag

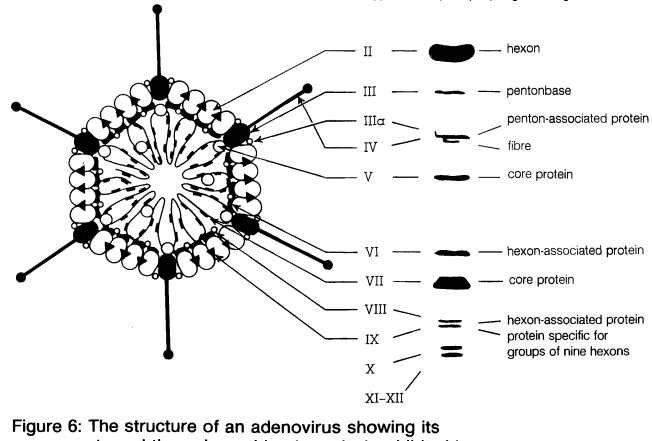


Figure 6: The structure of an adenovirus showing its components and the polypeptides (proteins) exhibited by staining after electrophoresis.

tissues for long periods of time (perhaps years or decades) and later undergoes recrudescence to cause the painful skin and neurological disease called herpes zoster, or shingles. In addition, there are herpesviruses that cause disease in animals, for example, the widespread and usually fatal disease in chickens called Marek's disease. The widespread distribution of viruses of the family Herpesviridae is evident from other diseases in monkeys and frogs.

The viruses of the family Retroviridae are perhaps the most widely distributed of the transforming viruses that infect eukaryotic cells ranging from yeast to humans. It was suggested early in the 20th century that viruses cause leukemias and lymphomas in birds. In 1911 the U.S. pathologist Peyton Rous first described a virus that causes sarcomas in chickens.

Retrovirus virions

The virions of retroviruses are spherical (or polygonal) and are surrounded by a lipid membrane containing a glycoprotein that recognizes and binds to cell receptors of a particular species (type-specific glycoproteins). Retrovirus genomes consist of two identical RNA molecules, each with 7,000 to 10,000 nucleotides. Associated with the virion RNA is an enzyme, an RNA-dependent DNA polymerase, also called a reverse transcriptase. Using the virion RNA as a template, the reverse transcriptase catalyzes the synthesis of a linear DNA molecule, complementary to the virion RNA. The new complementary strand of DNA also serves as a template for the reverse transcriptase, which makes a second anticomplementary DNA molecule, thus forming a double-stranded DNA. The genomic RNA of fully infectious bird retroviruses, those that can replicate autonomously, has four genes that code sequentially for group-specific antigens, the reverse transcriptase, the envelope glycoprotein, and the sarcoma-transforming protein. At each end of the genome are homologous flanking nucleotide sequences, known as long terminal repeats (LTR), which code for double-stranded DNA that can recognize host cell DNA sequences for integration of the proviral DNA into the host cell chromosome. Many retroviruses are defective and cannot replicate in cells without helper (nondefective) retroviruses. The helper retroviruses generally transform fibroblastic cells, resulting in malignant sarcomas, whereas the defective retroviruses transform blood-cell precursors, resulting in leukemias.

Many different retroviruses have been identified as causative agents of cancers in birds, rodents (particularly mice), domestic cats, monkeys, and, recently, humans. Certain lymphatic leukemias in humans are caused by human T-cell leukemia virus (HTLV); acquired immune deficiency syndrome (AIDS) is caused by a retrovirus called human immunodeficiency virus (HIV).

AIDS retrovirus

Retroviruses originated from genes in many different species of animals and even lower forms of life. Individual retroviruses are limited in their host range and do not readily cross species barriers. Virtually every retrovirus studied to date is analogous to the genes normally found in animals (including humans), known as proto-oncogenes (with the potential to change into cancer-causing genes). These proto-oncogenes have deoxynucleotide sequences closely, but not entirely, homologous (*i.e.*, of the same type and order) to the nucleotide sequences of a corresponding viral cancer-causing gene, called an oncogene. Integration of retrovirus DNA into cell chromosomes results in cancer, but the proto-oncogenes do not become cancer-causing genes unless triggered by another event. Cancers caused by chemical or physical carcinogens in the environment probably are often due to alterations in the sequences of proto-oncogenes that are converted to oncogenes. Some of the DNA tumour viruses, such as SV40 or adenoviruses, may induce malignant transformation when their DNA is integrated in proximity to the site of a proto-oncogene.

DISEASE

Benefits of viruses

Although viruses were originally discovered and characterized because of the diseases they cause, most viruses that infect bacteria, plants, and animals (including humans) do not cause disease. In fact, bacteriophages may be helpful in that they rapidly transfer genetic information from one bacterium to another, and viruses of plants and animals may convey genetic information among similar species, aiding the survival of their hosts in hostile environments. This could in the future be true for humans as well. Recombinant DNA biotechnology may allow genetic defects to be repaired by injecting afflicted persons with harmless viruses that carry and integrate functional genes to supplant defective ones. Such events may actually occur in nature in the transmission of "good" viruses from one person to another.

Of those viruses that cause disease, some cause short-term (acute) diseases and others recurring or long-term (chronic) diseases. Some viruses cause acute disease from which there is fairly rapid recovery, but the same virus may persist in the tissues, remaining dormant for long periods of time and then become active again, bringing about a serious disease decades later. Slowly progressive viruses have long incubation periods before the onset of disease. The DNA of certain viruses becomes integrated into the genome of the host cell, often resulting in malignant transformation of cells, which become cancers.

The nature of the disease caused by a virus is generally a genetic property of the virus as well as of the host cells. Almost all viruses, however, can remain dormant in the tissues of the host (latency). Viruses that cause acute disease are generally, but not always, those that rapidly harm or destroy cells (cytopathic effects) and have the capacity to shut off protein and/or nucleic acid synthesis within the host cell.

Human poliovirus

Human poliovirus and related picornaviruses that infect other animal species are examples of acute infectious agents that shut down protein synthesis in the host cell soon after infection; these picornaviruses also inhibit cellular RNA and DNA synthesis. Another virus that rapidly kills the infected cell is the negative-strand vesicular stomatitis virus (VSV) of the family Rhabdoviridae; viral RNA newly synthesized by infectious VSV rapidly shuts off cellular RNA synthesis and, to a somewhat lesser extent, cellular protein synthesis. In both poliovirus and VSV, the infected cell dies within hours of the inhibition of cellular RNA and protein synthesis. Other viruses that inhibit cellular macromolecule synthesis and produce acute infections include the poxviruses, reoviruses, togaviruses, adenoviruses, and herpesviruses; the latter two persist in host tissues for long periods of time and cause chronic infection as well. On the other hand, certain viruses, such as the influenza viruses, do not have the drastic effects on the synthesis of cellular macromolecules, such as proteins, but they still cause acute, usually self-limiting, infections.

Many, if not most, diseases resulting from viral infection of vertebrates are caused not by a direct effect of the virus but rather by a secondary immune response. Essentially all viral proteins are recognized by vertebrate animals as immunologically foreign, and the immune systems of these animals mount an immune response. There are two kinds of immune responses: humoral, in which B lymphocytes make antibodies (proteins that recognize and bind foreign molecules) to the viral protein, and cellular, in which T lymphocytes kill a cell infected with a virus that alters surface histocompatibility antigens (antigens that recognize and initiate immune responses against noncompatible cells). (In an immune response, an antigen stimulates the production of antibodies or binds to antibodies; an antibody is synthesized as a result of antigen stimulation.)

Immune responses

The antibody synthesized as a result of the immune response against a specific viral antigen usually benefits the infected host because that antibody can neutralize the infectivity of the specific virus in the blood and tissues of the infected host. Viruses inside the cell are not accessible to the antibody because it cannot cross the cell membrane barrier.

In contrast to antibodies, T cells do not significantly neutralize the infectivity of a virus. A killer T cell that recognizes a virus-infected cell because of the viral antigen on its surface can kill that cell and abort the infection because a virus will not grow within a dead cell. If the virus-infected cells are not essential for host functions, the killer T cell can prevent the spread of the infecting virus to other cells and distant tissues. Not infrequently, the virus-specific T lymphocyte kills vital cells such as nerve cells (neurons), muscle cells, and liver cells, all of which

carry out important functions. In addition, the death of cells results in an inflammatory response, which can also damage vital tissues.

In both cases, the immune response of the cell to a viral infection can cause disease. In general, diseases caused by chronic viral infections, but also occasionally by subacute (between acute and chronic) viral infections, are caused by cellular immune responses that damage the virus-infected tissue.

Local and systemic infection

Infectious patterns. Acute viral infections are of two types—local and systemic—both usually resulting from a direct effect of the invading virus on host tissue cells. Acute local infections generally occur at the site of viral infection. For example, acute respiratory infections include (1) the common cold, in which the rhinovirus infects only the nasal (nose) mucosa; (2) influenza, in which the virus is found in both nasal and bronchial mucosa, where severe damage can result in death; (3) grippelike illnesses caused by adenoviruses localized in lymphoid tissue of the throat (although infection can also occur in the intestine and the eye or be spread to the heart); and (4) severe respiratory infections of infants and children, caused by parainfluenza viruses or respiratory syncytial viruses, which cause severe croup that may be life-threatening. Examples of acute infections localized to the intestine include those that result in enteritis (bowel infection), which may be accompanied by diarrhea; these are often caused by rotoviruses and coronaviruses.

Many viruses transmitted by the respiratory route (from sneezes and coughs, for example) begin their cycle of infection in the upper respiratory tract (nose and throat) and then enter the bloodstream, where they are spread to distant tissues. Examples of such diseases are measles, mumps, and chicken pox, in which the growth of the specific virus in the mucosal cells (those that line body cavities) of the throat during the first few days of infection usually results in mild fever and achiness, called a prodromal period of the illness. During the next few days the virus enters the draining lymph nodes and then the bloodstream, where it is spread throughout the tissues of the body, resulting in fever and rash (in the case of measles and chicken pox) and inflammation of the parotid glands and, less frequently, the testes, ovaries, and joints (in the case of mumps). Varicella (chicken pox) virus rarely causes pneumonia, but all of these viruses can cause meningitis and, rarely, encephalitis. A similar pattern of infection formerly occurred with smallpox, a disease that was more frequently fatal.

A large number of viruses of the digestive tract (enteroviruses), among them poliovirus, Coxsackie viruses, and echoviruses (enteric cytopathic human orphan), also cause a two-phase illness. Enteroviruses grow initially in the intestinal tract and are transmitted by mouth through water, food, and other materials contaminated with feces. The viruses are resistant to the acid normally found in the stomach and thus reach the intestinal tract, where they multiply in living mucosal cells. This initial period of viral invasion and growth in the intestine causes either an initial mild febrile illness or is asymptomatic. Over the next few days these enteroviruses are spread from the intestinal mucosa to the draining lymph nodes, from which they invade the bloodstream, resulting in a condition known as viremia. From the bloodstream the viruses are widely spread to all tissues but, in most cases, no symptomatic disease occurs. Poliovirus in less than 1 percent of cases affects the spinal cord and/or brain, resulting in paralysis or death. Different types of Coxsackie viruses and echoviruses can cause acute illnesses, usually nonfatal, such as meningitis, carditis, pleurisy, severe muscle pains, or rashes.

Route of enteroviruses

Latency. Inapparent infections (those that do not cause specific signs and symptoms) often result after exposure to picornaviruses, influenza viruses, rhinoviruses, herpesviruses, and adenoviruses, but less frequently to measles and chicken pox viruses. In cases of inapparent infection, long-lasting immunity develops, but only to the strain of virus that has the same antigenic composition as the original infecting virus.

Certain of these viruses persist in the tissues of the host after the initial infection despite the presence of a circulating antibody in the blood and tissues. Such viruses probably reside inside cells, where they are protected from antibodies that cannot penetrate the cell membrane. Among persistent viruses are adenoviruses, measles virus, and, in particular, many kinds of herpesviruses. The genetic information of herpesviruses and adenoviruses can be integrated into the genome of the host cell, but it is believed that these viruses frequently, and the measles virus invariably, reside in cells in the form of extrachromosomal genes (genes not integrated in chromosomes). These dormant viruses can be activated by many factors, such as trauma, another infection, emotional stress, menstruation, excessive exposure to sunlight, and various illnesses.

The phenomenon of latency and reactivation is particularly common among viruses of the family Herpesviridae, which cause chronic or recurrent diseases: (1) herpes simplex virus type 1, which causes recurrent cold sores; (2) herpes simplex virus type 2 in genital tissue, which causes repeated herpetic infections of the vagina or penis; (3) cytomegalovirus, which usually produces an inapparent infection activated by simultaneously occurring disease to cause severe liver, lung, or nervous-system disease; and (4) varicella virus, which is the causative agent of chicken pox but which can be activated decades later to produce herpes zoster (shingles). A rare, but invariably fatal, disease of the nervous system is subacute sclerosing panencephalitis (SSPE), which is a progressive, degenerative condition caused by measles virus (a paramyxovirus), lying dormant in brain cells for many years and then reactivated, usually in early adult life. There is no simple explanation why latent viruses, such as those in the family Herpesviridae, that are present in the tissues of most humans can be activated to cause disease in some people but not in others.

Chronic and slowly progressive diseases. Although some viruses multiply slowly, this is not always the explanation for the chronicity or the slow progression of the diseases caused by these viruses. Hepatitis, for example, is a subacute and/or chronic disease, with a long latent period, that is caused by at least three viruses with different properties. Hepatitis A is caused by a picornavirus usually transmitted by the fecal–oral route in a manner similar to that of poliovirus. Hepatitis B is caused by a small DNA virus that contains its own DNA polymerase and is transmitted by blood transfusion, other blood products, or nonsterile hypodermic needles. A disease described as non-A, non-B hepatitis is caused by a third virus, known only as the delta agent. Non-A, non-B hepatitis has become the most common form of hepatitis transmitted by transfusion or blood products.

Many other agents that appear to cause chronic and slowly progressive diseases, particularly those affecting the nervous system, have been tentatively identified. A fatal neurological disorder of sheep, called scrapie, has an incubation period of years and may be caused by a protein free of nucleic acid, which is postulated to activate latent genes that cause the disease (a prion). Similar, rather obscure, agents have been identified for two uncommon, fatal disorders of the nervous system, called Creutzfeldt-Jakob disease and kuru; the latter is found among cannibalistic natives in New Guinea.

Neurological diseases

Prevention. The spread of many viral diseases can be prevented by hygienic practices such as efficient sanitation facilities, effective waste disposal, clean water, and personal cleanliness. Active immunization by vaccines (antigen-containing preparations that elicit the synthesis of antibodies and thus immunity) has been useful in preventing common epidemics caused by acutely infectious viruses.

The best example of such a preventable disease is smallpox, caused by a disease-producing virus that at one time was found worldwide. In 1798 the English physician Edward Jenner discovered that the milder cowpox virus could serve as a live vaccine (an antigenic preparation consisting of viruses whose disease-producing capacity has been weakened) for preventing smallpox. The program of vaccination that resulted from Jenner's discovery is one of the greatest success stories in the annals of medicine; in 1980 the World Health Organization declared that the disease had been eliminated.

Vaccination

A different prospect is presented by rabies, an invariably

fatal viral disease mentioned in ancient Greek literature and in the Bible. Transmitted by the bite of dogs and other domestic and wild animals, the rabies virus is more difficult to eradicate because it is present in wild animals throughout the world, except in certain island countries such as Great Britain and Australia. Influenza, another worldwide virus, is, of the three major immunologic types, only one of which (type A) is responsible for large epidemics. The worldwide epidemic (pandemic) of influenza at the end of World War I is estimated to have caused 20,000,000 deaths, mostly of adolescents and young adults. Because of virus mutations that produce minor antigenic changes every few years and major antigenic shifts around every 10 years, influenza viruses have the capacity to resist inactivation by antibodies acquired by previous infection or vaccination. Development of effective vaccines to combat influenza is a difficult task, although existing vaccines are partially effective and are recommended for people at high risk (*i.e.,* the elderly and those with recurring disease of the respiratory or circulatory systems).

Vaccines are most successful when directed against those viruses that do not mutate and that infect only humans. In addition to smallpox, a successful vaccine program has been carried out against poliomyelitis. Polioviruses exist in only three antigenic types, each of which has not changed significantly for decades. The vaccines available are the "killed" (Salk) vaccine, composed of inactivated virus of the three types, and the "live" (Sabin) vaccine, composed of genetically attenuated viruses of the three types. These vaccines, which were introduced in the 1950s, have lowered the incidence in developed countries of paralysis due to poliomyelitis. The disease still occurs in developing countries and recurs in some developed countries where vaccination programs have not been enforced. Some cases of poliomyelitis occur, particularly in the United States, from the Sabin vaccine strain of type 3 poliovirus, which is genetically unstable and occasionally reverts to the virulent form.

Poliomyelitis

Vaccination can prevent diseases caused by strictly human viruses that exist in only one antigenic and stable type. Measles has been prevented in developed countries with routine vaccination. Measles, however, may still be the major cause of death in childhood in developing countries. Vaccination for mumps and chicken pox promises to be successful because the causative viruses of these diseases show little tendency to vary antigenically and are confined to humans. On the other hand, development of vaccines for the common cold caused by rhinoviruses, similar to polioviruses, will be a formidable, if not impossible task, because there are at least 100 antigenic types of the rhinovirus.

Treatment. Unlike bacteria, viruses mimic the metabolic functions of their host cells. Antibiotics and other antimicrobial agents are, therefore, ineffective against viruses because the chemical compounds that inhibit the multiplication of viruses generally also slow the functions of, and are toxic to, the host cell. The only chemotherapeutic agents (drugs used to control or eliminate disease) that exert a selectively greater effect on viral replication than they do on cell replication are certain man-made compounds, such as ribavirin, acyclovir, or azidothymidine (AZT). These drugs have been partially effective in improving, if not curing, viral diseases without causing major toxic side effects. AZT has been used with some success in prolonging the lives of patients with acquired immune deficiency syndrome (AIDS).

Interferons

Natural products of cells, called interferons, may have potential antiviral and anticancer properties. Interferons are proteins normally synthesized by the cells of vertebrates, including humans, either intrinsically and without stimulation, or in response to certain viral infections, chemicals, or immune reactions. In general, the multiplication of viruses is inhibited by interferons, some to a much greater extent than others. Interferons are generally species-specific; *i.e.,* they are effective in inhibiting viral infection only in cells of the same species that naturally synthesize the interferon.

There are three classes of interferons: α-interferons, produced by blood leukocytes; β-interferons, produced by tissue cells and fibroblasts; and γ-interferons (also called immune interferons or interleukins), produced by immune reactions in blood lymphocytes. Until recently, interferons were difficult to produce commercially because cells and tissues synthesize only small amounts. Recombinant DNA technology, however, has permitted cloning of interferon genes in bacteria or in other cells, thus producing large amounts that can be readily purified.

There has been some success in using interferons to treat viral diseases, such as colds caused by rhinoviruses; infections caused by herpesvirus; and benign tumours and warts caused by papillomaviruses. Local administration at the sites of viral infection affords the best results, although injections of large amounts of interferons can be harmful, probably because they tend to inhibit protein synthesis in the host cell.

EVOLUTION

Owing to their simplicity, viruses were at first considered to be the primordial life forms. This concept is almost certainly incorrect because viruses are completely devoid of the machinery for life processes; therefore, they could not have survived in the absence of cells. Viruses probably evolved from cells rather than cells from viruses. It seems likely that all viruses trace their origins to cellular genes and can be considered as pieces of rogue nucleic acids. Although it is easier to imagine the cellular origin of DNA viruses than RNA viruses, the RNA viruses conceivably could have had their origins from cellular RNA transcripts made from cellular DNA. In fact, the discovery that many cells contain reverse transcriptase capable of converting RNA to DNA seems to suggest that conversion of RNA to DNA and DNA to RNA is not rare.

Origins of viruses

Other possible progenitors of viruses are the plasmids (small circular DNA molecules independent of chromosomes), which are more readily transferred from cell to cell than are chromosomes. Theoretically, plasmids could have acquired capsid genes, which coded for proteins to coat the plasmid DNA, converting it into a virus.

In brief, it is likely that viruses originated from the degradation of cellular nucleic acids, which acquired the property of being readily transferable from cell to cell during the process of evolution.

CLASSIFICATION

Distinguishing taxonomic features. Viruses are classified on the basis of their nucleic acid content, their size, the shape of their protein coat (capsid), and the presence of a surrounding lipoprotein membrane (envelope).

The primary taxonomic division is into two classes based on nucleic acid content: DNA viruses or RNA viruses. The DNA viruses are subdivided into those that contain either double-stranded or single-stranded DNA. The RNA viruses are also divided into those that contain double-stranded or single-stranded RNA. Further subdivision of the RNA viruses is based on whether the RNA genome is segmented or not. If the viruses contain single-stranded RNA as their genetic information, they are divided into positive-strand viruses if the RNA is of messenger sense (directly translatable into proteins) or negative-strand viruses if the RNA must be transcribed by a polymerase into mRNA.

All viruses falling into one of these nucleic acid classifications are further subdivided on the basis of whether the nucleocapsid (protein coat) assumes a rodlike or a polygonal (usually icosahedral) shape. The icosahedral viruses are further subdivided into families based on the number of protein subunits (capsomers) making up the capsids. Finally, all viruses fall into two classes depending on whether or not the nucleocapsid is surrounded by a lipid envelope.

Subdivisions of viruses

Some virologists adhere to a division of viruses into those that infect bacteria, plants, or animals; these classifications have some validity, particularly for the unique bacterial viruses with tails, but there is otherwise so much overlap that taxonomy based on hosts seems unworkable. Classification based on diseases caused by viruses is also not tenable because closely related viruses frequently do not cause the same disease. Eventually, it is likely that the classification of viruses will be based on the nucleotide

sequences of their nucleic acids rather than on structural components, as is now the case.

The basic taxonomic group is called a family, designated by the suffix -viridae. The major taxonomic disagreement among virologists is whether to segregate viruses within a family into a specific genus and further subdivide them into species names. Most virologists believe that a binomial nomenclature, as used for bacteria, into italicized and Latinized genera and species is unwarranted. For this reason, no Latin names are used in the classification of viruses presented here. Moreover, only the well-characterized viruses of animals are presented.

Annotated classification.

DNA viruses

FAMILY POXVIRIDAE

Large viruses of complex structure with dimensions of 400 nm × 200 nm, the genome of which is linear double-stranded DNA. Virions contain at least 40 proteins and lipid, as well as internal structures called lateral bodies. The 2 subfamilies are called Chordopoxvirinae, which infect vertebrates and are closely related antigenically, and Entomopoxvirinae, which infect arthropods. The Chordopoxvirinae are composed of groups called orthopoxviruses (vaccinia), parapoxviruses, avipoxviruses of birds, and many others that infect sheep, rabbits, and swine.

FAMILY ADENOVIRIDAE

Nonenveloped virions of icosahedral symmetry, about 80 nm in diameter, and capsids containing 252 capsomers with 12 vertices to which are attached glycoprotein fibres 10–30 nm in length with knobs at the ends. The genome is linear double-stranded DNA. Classified in 2 subgroups: Mastadenoviruses, which infect mammals; and Aviadenoviruses, which infect birds. Common acute respiratory and gastrointestinal pathogens of humans, and some types cause malignant transformation of cultured cells and animals.

FAMILY HERPESVIRIDAE

Icosahedral virion with capsid about 105 nm in diameter and 162 capsomers surrounded by a floppy envelope containing glycoprotein spikes. Genome composed of linear double-stranded DNA. There are 3 known subfamilies: Alphaherpesvirinae, consisting of human herpes simplex viruses types 1 and 2, bovine mamillitis virus, SA 8 virus and monkey B virus, pseudorabies virus, equine herpesvirus, and varicella-zoster virus; Betaherpesvirinae, composed of species of cytomegaloviruses; and Gammaherpesvirinae, composed of genera familiarly called Epstein-Barr virus, baboon herpesvirus, chimpanzee herpesvirus, Marek's disease virus of chickens, turkey herpesvirus, herpesvirus saimiri, and herpesvirus ateles.

FAMILY IRIDOVIRIDAE

Large enveloped or nonenveloped icosahedral virions measuring 130–300 nm in diameter and containing linear double-stranded DNA. Genera include iridovirus and chloridirovirus, which infect insects, African swine fever virus, and lymphocystic disease virus.

FAMILY PAPOVAVIRIDAE

Icosahedral, nonenveloped virions with 72 capsomers comprising 2 genera: the polyomaviruses (SV40 and polyoma virus), measuring 45 nm in diameter; and the papillomaviruses, measuring 55 nm in diameter. Virions of both genera contain covalently linked circular DNA. The polyomaviruses produce malignant transformation of infected cells, whereas the papillomaviruses, which do not grow in cell culture, usually cause warts and benign papillomas.

FAMILY PARVOVIRIDAE

Small icosahedral, nonenveloped virions with 32 capsomers measuring 18–26 nm in diameter that contain single-stranded DNA. Viruses of this family infect vertebrates and insects, the latter sometimes designated densoviruses. The vertebrate viruses fall into 2 classes: those that replicate autonomously, and those that replicate only in the presence of helper adenoviruses or herpesviruses, designated adenoassociated viruses (AAV) or dependoviruses.

RNA viruses

FAMILY PICORNAVIRIDAE

Small icosahedral, nonenveloped virions 27–30 nm in diameter, composed of 60 capsomers and containing nonsegmented single-stranded, plus-strand RNA. There are 4 recognized genera designated enteroviruses (polioviruses, Coxsackie viruses, echoviruses), cardioviruses, rhinoviruses (common cold viruses), and aphthoviruses (foot-and-mouth disease virus of cattle).

FAMILY CALICIVIRIDAE

Icosahedral, nonenveloped virions about 38 nm in diameter, composed of 32 capsomers and 180 molecules of a single capsid protein. The genome consists of single plus-strand RNA. The prototype virus of this family is the vesicular exanthema of swine virus.

FAMILY TOGAVIRIDAE

Small enveloped virions spherical in shape with icosahedral nucleocapsid of 32 capsomers about 30 nm in diameter. The genome is single plus-strand RNA. There are 3 recognized genera: one transmitted by arthropods (exclusively mosquitoes) and designated alphaviruses (prototypes Sindbis virus, eastern and western equine encephalitis viruses), and the other two nonarthropod-borne, designated rubivirus (the causative agent of German measles) and pestivirus (mucosal disease viruses such as bovine virus—or diarrhea virus—and hog cholera virus).

FAMILY FLAVIVIRIDAE

Viruses of this newly designated family are enveloped and spherical in shape, with a genome consisting of nonsegmented single plus-strand RNA. These viruses are transmitted by either insects or arachnids and cause severe diseases such as yellow fever, dengue, tick-borne encephalitis, and Japanese B encephalitis.

FAMILY CORONAVIRIDAE

Enveloped virions 60–220 nm in diameter with helical nucleocapsid containing single plus-strand RNA genome. Club-shaped glycoprotein spikes in envelope give crownlike (coronal) appearance. Viruses of this family are important agents of gastrointestinal disease in humans, poultry, and bovines.

FAMILY ORTHOMYXOVIRIDAE

Enveloped virions about 100 nm in diameter with helical nucleocapsid containing 8 segments of negative-strand RNA and endogenous RNA polymerase. The lipoprotein envelope contains 2 glycoproteins, designated hemagglutinin (major antigen) and neuraminidase. The only viruses in this family are influenza viruses of 3 distinct antigenic types (A, B, and C).

FAMILY PARAMYXOVIRIDAE

Enveloped virions varying in size from 150 to 300 nm in diameter with a helical nucleocapsid containing a single negative-strand nonsegmented RNA and an endogenous RNA polymerase. The lipoprotein envelope contains 2 rods of glycoprotein spikes designated hemagglutinin-neuraminidase (HN) and fusion factor (F). The major genus is paramyxovirus and is composed of human parainfluenza viruses and mumps virus, as well as Newcastle disease virus of poultry. The genus morbillivirus contains the agents that cause measles in humans, distemper in dogs and cats, and rinderpest in cattle. The third genus, pneumovirus, causes the serious respiratory synctial virus disease in human infants.

FAMILY RHABDOVIRIDAE

Enveloped virions, usually bullet-shaped, 180 to 300 nm × 65 nm containing a helical nucleocapsid with single minus-strand RNA and an endogenous RNA polymerase. The lipoprotein envelope contains a single glycoprotein, which is the type-specific antigen. Viruses of this family are widely infectious for plants and for animals varying from insects to humans. Two animal genera are designated vesiculovirus, which includes the virus that causes vesicular stomatitis in cattle, swine, and equines; and lyssavirus, which includes the causative agent of rabies.

FAMILY ARENAVIRIDAE

Enveloped virions 100–200 nm in diameter, with a helical nucleocapsid in 2 segments containing minus-strand RNA as well as small amounts of ribosomal RNA. There are 4 genera, with viruses widely distributed in animals and causing serious diseases transmitted by insects.

FAMILY BUNYAVIRIDAE

Enveloped virions about 95 nm in diameter, with 3 segmented helical nucleocapsids containing single-stranded RNA of negative sense. Many viruses grouped in 4 genera: bunyavirus, phlebovirus, nairovirus, and uukuvirus.

FAMILY RETROVIRIDAE

Enveloped virions about 90 nm in diameter, with 2 identical copies of single plus-strand RNA in nondefective virions and a reverse transcriptase, which promotes synthesis of double-stranded DNA from the viral RNA template. A hallmark of the virion RNA templates is long terminal repeat (LTR) nucleotide sequences, which serve for integration of the DNA in chromosomes of the host cell. Retroviridae cause cancers in many species of animals, including humans, and are probably derived from normal cell nucleotide sequences called proto-oncogenes.

FAMILY REOVIRIDAE

Nonenveloped icosahedral virions with outer and inner protein shells 60–80 nm in diameter and containing double-

stranded RNA in 10 to 12 segments. Viruses in this family infect many species of plants and animals. The animal Reoviridae are divided into 4 genera, designated orthoreoviruses, orbiviruses (widely distributed in insects and vertebrates, including bluetongue virus disease of sheep), rotaviruses (widespread causative agents of gastroenteritis in mammals, including humans), and cypovirus (prototype causes cytoplasmic polyhedrosis disease in insects).

BIBLIOGRAPHY. BERNARD N. FIELDS (ed.), *Virology* (1985), a text on the structure, biologic properties, replication, and immunology of virtually all human viruses of medical importance, including descriptions of the diseases and their epidemiology; DAVID O. WHITE and FRANK FENNER, *Medical Virology,* 3rd ed. (1986), a text intended for medical students and other health professionals; HEINZ FRAENKEL-CONRAT and ROBERT R. WAGNER (eds.), *Comprehensive Virology,* 19 vol. (1974–84), a detailed critical analysis of the biology, chemistry, and physical properties of viruses; C.H. ANDREWES, *The Natural History of Viruses* (1967), a personal account by one of the pioneers in the field; A.P. WATERSON and LISE WILKINSON, *An Introduction to the History of Virology* (1978), a brief but incisive history of the science of virology; and SHERWOOD CASJENS (ed.), *Virus Structure and Assembly* (1985), an illustrated series of essays by some of the major contributors to the understanding of the physical principles that determine the structure and assembly of viruses. See also N.J. DIMMOCK and S.B. PRIMROSE, *Introduction to Modern Virology,* 3rd ed. (1986). Useful journal articles include JAMES M. HOGLE, MARIE CHOW, and DAVID J. FILMAN, "The Structure of Poliovirus," *Scientific American,* 256(3):42–49 (March 1987), an illustrated account of the crystallographic structure of poliovirus; and two articles from *Advances in Virus Research:* G. STREISSLE, A. PAESSENS, and H. OEDIGER, "New Antiviral Compounds," 30:83–138 (1985), an account of potentially useful chemicals for the treatment of viral diseases; and ERLING NORRBY, "Toward New Viral Vaccines for Man," 32:1–34 (1987), an analysis of principles for creating effective vaccines for preventing viral infection. On viral classification, see R.E.F. MATTHEWS (ed.), *A Critical Appraisal of Viral Taxonomy* (1983), opinions on logical means for classifying viruses; and HEINZ FRAENKEL-CONRAT, *The Viruses: Catalogue, Characterization, and Classification* (1985), a dictionary of names, with classifications and brief descriptions of all viruses, listed alphabetically under the categories of animal viruses, plant viruses, and bacteriophages of prokaryotes and adhering to the principles for classification laid down by the International Committee on Taxonomy of Viruses.

(R.R.W.)

Volcanism

Volcanoes figure prominently in the mythology of many peoples who have learned to live with eruptions, but science was late in recognizing the important role of volcanism in the evolution of the Earth. One major 18th-century school of thought held that molten rock and volcanoes were simply accidents caused by burning coal seams. Geologists today agree that volcanism is a profound process resulting from the thermal evolution of planetary bodies. Heat does not easily escape from large bodies by conduction or radiation. Instead, partial melting and buoyant rise of magma are major contributors to the process of heat flux from the Earth's interior. Volcanoes are the surface manifestation of this thermal process, which has its roots deep inside the Earth and which hurls its ashes high into the atmosphere.

The term volcano can either mean the vent from which magma erupts to the surface, or it can refer to the landform created by the solidified lava and fragmental volcanic debris that accumulate near the vent. One could say, for example, that large lava flows are erupted from Kilauea Volcano in Hawaii, the word volcano here signifying a vent. By contrast, one could say that Kilauea is a gently sloping volcano of modest size as Hawaiian volcanoes go, the reference in this case being to a landform.

Broadly defined, all igneous rocks are the result of volcanism. If igneous rocks solidify from magmas that have not reached the surface, they are called intrusive igneous rocks, and this process in a more restricted sense is termed plutonism. Igneous rocks that cool and solidify at the Earth's surface are known as extrusive igneous rocks, and these are unequivocally products of volcanism.

Volcanoes (and their products) are not the realm of any single scientific discipline. Rather, they require study by many scientists from several specialties: geophysicists and geochemists to probe the deep roots of volcanoes; geologists to decipher prehistoric volcanic activity; biologists to learn how life becomes established and evolves on barren volcanic islands; and meteorologists to determine the effects of volcanic dust and gases on the atmosphere, weather, and climate.

This article discusses volcanism both by topics and by examples. Under the topic of types of volcanic activity, the 1980 eruption of Mt. St. Helens in southwestern Washington is used as an example of an explosive eruption that involves ejection of pyroclastic fragments. The 1984 eruption of Mauna Loa Volcano in Hawaii serves to illustrate a "quiet" eruption involving the effusion of lava. The widely varying composition of volcanic gases from fumaroles is explained by studies at Kilauea Volcano, and Yellowstone National Park in the western United States provides classic examples of hot springs and geysers.

Volcanic landforms evolve from the cumulative effects of both constructive and destructive eruptions. Mt. Fuji exemplifies a stratovolcano and Mauna Loa typifies a shield volcano. Iceland provides fine examples of volcanic plateaus, and since it sits astride the Mid-Atlantic Ridge, it is also a logical starting place for a discussion of submarine volcanic structures.

Volcanoes are closely associated with tectonic activity. Most of them occur on either the overriding or the diverging margins of the enormous lithospheric plates that make up the Earth's surface. The volcanoes of Japan provide an excellent example of the former, while those of the Mid-Atlantic Ridge the latter. Intraplate volcanoes such as those of the Hawaiian chain provide important evidence as to the direction and rate of plate motion.

Volcanoes affect humankind in many ways. Their destructiveness is awesome, but the risk involved can be reduced by assessing volcanic hazards and forecasting volcanic eruptions. Volcanism provides fertile soils, valuable mineral deposits, and geothermal energy. Over geologic time volcanoes recycle the Earth's hydrosphere and atmosphere, and explosive eruptions can affect climate.

For coverage of related topics in the *Macropædia* and *Micropædia,* see the *Propædia,* sections 212 and 231.

This article is divided into the following sections:

Types of volcanic activity 512
  Volcanoes 512
    Types of eruptions
    Some notable examples
  Other manifestations of volcanic activity 515
    Volcanic gases
    Hydrothermal features
Volcanic landforms 516
  Determinants of size and shape 516
  Major types of volcanic landforms 517
Volcanism and tectonic activity 519
  Volcanoes related to plate boundaries 519
  Intraplate volcanism 520
Impact of volcanic activity on humankind and the environment 521
  Hazards from various types of volcanic eruptions 522
  Effects of volcanic eruptions on soil and climate 522
  Geothermal energy 523
Bibliography 523

Types of volcanic activity

VOLCANOES

Basis for classification

Volcanoes can be classified by their eruptive habits and are generally arranged by progressive increases in the explosiveness of their eruptions. The type of volcanic eruption also plays an important role in the evolution of a volcanic landform, thus forming a significant link between eruptive habit and volcanic structure.

Less explosive eruptions involve the effusive outpouring of basaltic magma that is relatively low in viscosity and in gas content. (The term lava is applied to molten magma after it has erupted to the surface.) More explosive eruptions generally involve magma that is more viscous and has a higher gas content. Such magma is shattered into pyroclastic fragments by explosive boiling during an eruption.

Types of eruptions. In classification schemes based on character of eruption, volcanic activity and volcanic areas are commonly divided into six major types in order of increasing degree of explosiveness: (1) Icelandic, (2) Hawaiian, (3) Strombolian, (4) Vulcanian, (5) Pelean, and (6) Plinian.

Fissure eruptions

The Icelandic type is characterized by effusions of molten basaltic lava that flow from long parallel fissures. Such outpourings often build lava plateaus.

The Hawaiian type is similar to the Icelandic variety. In this case, however, fluid lava flows from summit and radial fissures to form shield volcanoes.

The Strombolian type involves moderate bursts of expanding gases that eject clots of incandescent lava in cyclical or nearly continuous small eruptions. Because of such small intermittent outbursts, Stromboli Volcano (off the west coast of Italy) has been called the "lighthouse of the Mediterranean."

The Vulcanian type, named for the island of Vulcano that neighbours Stromboli, generally involves moderate explosions of gas laden with volcanic ash (*i.e.,* fine volcanic particles). This mixture forms dark, turbulent erup-

tion clouds that rapidly ascend and expand in convoluted shapes.

Explosive eruptions

The Pelean type is associated with explosive outbursts that generate dense mixtures of hot volcanic fragments and gas. It is named for the destructive eruption of Mt. Pelée in Martinique in 1902. The fluidized slurries produced by Pelean-type eruptions are heavier than air but are of low viscosity and pour down valleys and slopes at velocities exceeding 100 kilometres per hour (60 miles per hour). Variously called *nuées ardentes,* glowing avalanches, ash flows, or pyroclastic flows, such hot fluidized flows of volcanic gases and fragments are extremely destructive.

The Plinian type is an intensely violent kind of volcanic eruption exemplified by the outburst of Mt. Vesuvius in AD 79 that killed the famous Roman scholar Pliny the Elder, from whom the name is taken. In this type, gases boiling out of gas-rich magma generate enormous, nearly continuous jetting blasts that rip apart and core out the magma column. The uprushing gases and volcanic fragments appear like a gigantic rocket blast directed vertically upward. Plinian eruption clouds can rise into the stratosphere and are sometimes sustained for several hours.

Figure 1 features schematic diagrams of these eruption types. There are, however, many gradations among—and exceptions to—such idealized types, and it is not unusual for an eruption sequence to involve more than one type of activity. For example, the eruptions of Mt. St. Helens from 1980 to 1986, in sequence, consisted of small Vulcanian-type explosions, large Pelean and Plinian explosions, and extrusions of viscous lava into a lava dome that caps the vent. The different types of volcanic activity can best be described by comparing specific eruptions.

From A. Holmes, *Principles of Physical Geology;* reproduced with permission of Van Nostrand Reinhold (UK) Ltd, Wokingham, Berkshire, UK

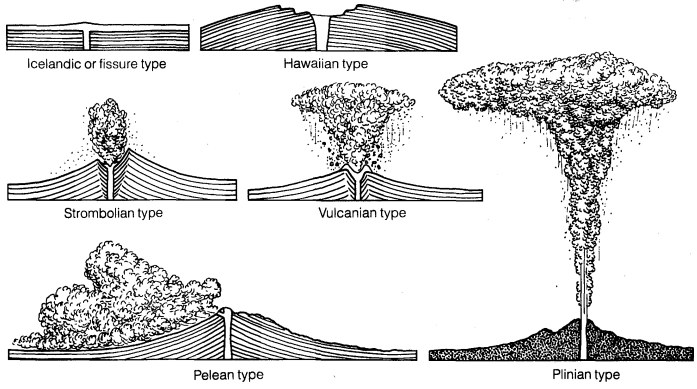


Figure 1: The major types of volcanic eruptions.

Some notable examples. *Mt. St. Helens, Washington, 1980–86.* Prior to 1980 Mt. St. Helens was a steep, conical volcanic peak rising to 2,950 metres (9,680 feet). Its summit was capped by snowfields and small glaciers. It had erupted during at least 20 intervals over the past 4,500 years, with the last active period between 1831 and 1857. Dwight Crandell and Donal Mullineaux of the U.S. Geological Survey studied the volcanic deposits in the vicinity of Mt. St. Helens and noted in their 1978 report that the volcano would probably erupt violently and intermittently, just as it had in the recent geologic past, "within the next 100 years, and perhaps even before the end of this century." Two years later these prophetic words became reality.

Pre-eruption earthquake swarm

On March 20, 1980, an earthquake swarm began near Mt. St. Helens. By March 25 the swarm increased to its maximum energy release with 47 earthquakes of magnitude 3 or greater on the Richter scale within a 12-hour period. Seismographs located the source of the earthquake swarm at shallow depth beneath the north flank of the mountain (Figure 2).

An eruption was expected, and it soon began. The first small explosions on March 27 excavated a crater about 70 metres wide on the ice-covered summit. Several such small Vulcanian-type eruptions occurred daily. They ejected dark, turbulent clouds of volcanic ash that blackened the downwind snowfields. No new magma appeared

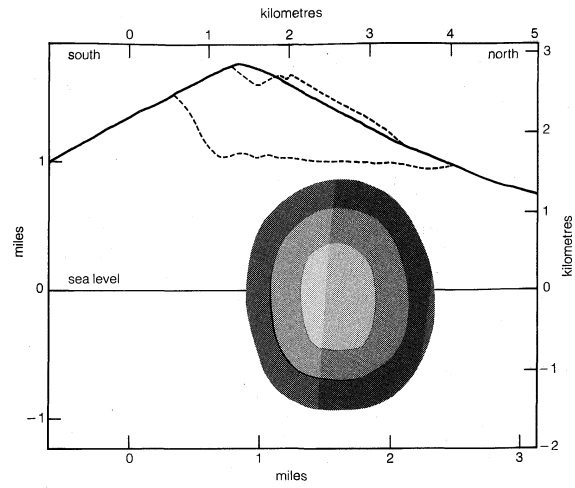


Figure 2: *Topographic profiles of Mt. St. Helens and location of the pre-eruption earthquake swarm.*
The solid line represents the profile before March 20, 1980; the upper line depicts the profile of May 17, 1980, just before the major eruption; and the lower line indicates the profile immediately after the eruption of May 18. The shaded area represents the region in which the earthquake swarm of March 20–May 18 occurred. The darker the shading, the higher the density of earthquake locations.

From R. Decker and B. Decker, "The Eruptions of Mount St. Helens," copyright © 1981 by Scientific American, Inc., all rights reserved; data from the U.S. Geological Survey and the Department of Geophysics of the University of Washington

in the ash, however; only old volcanic debris was thrown out. These small explosive eruptions apparently resulted from groundwater in the volcanic cone being heated above its subsurface boiling point and flashing into steam. The crater was enlarged by the small explosive eruptions to an oval basin 500 by 300 metres wide and 200 metres deep.

At the peak of seismic energy release on March 25, large east–west cracks formed across the summit of Mt. St. Helens. Soon after this, people familiar with the mountain noticed changes in its shape and in ice fracture patterns on the high north face of the volcano. Geodetic surveys and aerial photographs in April and early May 1980 confirmed that a growing bulge nearly two kilometres (1.2 miles) in diameter on the high north slope was moving upward and outward by as much as one to two metres per day.

Most of the scientists studying the mountain at this time agreed that an intrusion of magma was being injected at a shallow depth beneath the north side of the volcano. Whether this intrusion would erupt at the surface—and if so, when—were unanswered questions.

Massive landslide of rock and ice

On the morning of May 18, Mt. St. Helens was shaken by an earthquake of about 5 Richter magnitude, and ice falls occurred within the crater. The entire north side of the summit then began to move as one huge mass. This giant landslide, composed of more than two cubic kilometres (0.48 cubic mile) of rock and ice, was immediately followed by an enormous explosion of expanding steam and volcanic gases. Fluidized by the expanding gases, the avalanche reached velocities as high as 250 kilometres per hour. One lobe swept into Spirit Lake, but the main body of the debris avalanche roared 21 kilometres down the valley of the Toutle River, forming a hummocky deposit one to two kilometres wide and up to 150 metres deep.

As the great avalanche suddenly released the pressure inside the volcanic cone, superheated groundwater flashed into steam, and gases dissolved in the magma boiled out with explosive violence. These expanding gases formed a lateral blast of hot, dense, debris-filled clouds that surged northward along the ground at high speed. The blast devastated 550 square kilometres (210 square miles) of mountain terrain to the northeast, north, and northwest of Mt. St. Helens. The ground-hugging black clouds of steam and hot rock fragments rolled over four major ridges and valleys as far as 28 kilometres from the volcano's summit.

The destruction was awesome. For the first few kilometres everything was obliterated. Great conifer trees up to three metres in diameter were uprooted and blown away. Beyond this was a blowdown zone where virgin Douglas

firs were snapped off like matchsticks. The U.S. Forest Service estimated that 10,000,000 trees were felled by the blast. At the limits of the blast cloud, trees were still standing but their needles were scorched beyond recovery. About 55 to 60 people were killed by the avalanche and blast.

Plinian cloud

As the ground-hugging blast clouds raced northward, a Plinian eruption began to jet upward from the exposed-magma column. This vertical eruption cloud mushroomed to more than 20 kilometres in altitude and was sustained most of the day by the escaping gases as they cored out the gas-rich magma. Northeast and east winds blew ash from the Plinian cloud in a wide swath over Washington, Idaho, and Montana. At Ritzville, Wash., 330 kilometres east-northeast of Mt. St. Helens, volcanic ash that appeared "like gray talcum powder" was seven centimetres (nearly three inches) thick. The weight of the measured ashfall equaled 0.15 cubic kilometre of magma, and the fine ash and gases dispersed into the stratosphere were estimated to equal another 0.1 cubic kilometre of magma.

During the afternoon of May 18, pyroclastic flows of fine ash and pumice blocks began rushing down the north slope of Mt. St. Helens through the deep breach in the north face of the mountain. These fluidized emulsions of hot rock fragments mixed with hot volcanic gases were denser than the ascending ash cloud, and they issued from the crater beneath the uprushing cloud. In effect, both Plinian- and Pelean-type eruptions occurred at the same time.

R.W. Decker

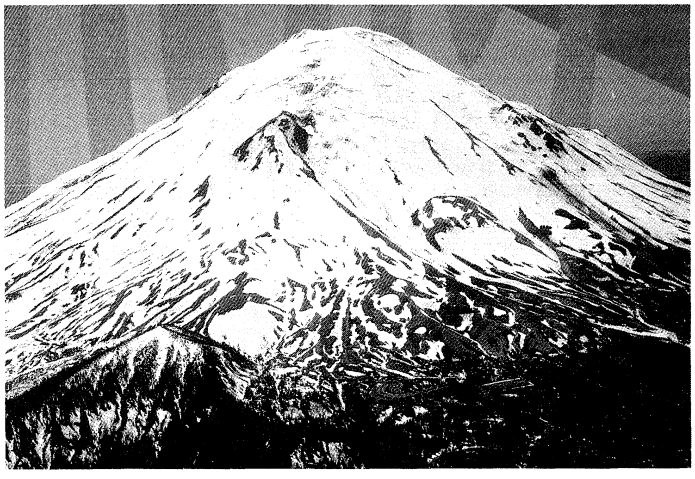


Figure 3: The north face of Mt. St. Helens in June 1970. The summit elevation was 2,950 metres.

Floods and mudflows consisting of slurries of volcanic ash and fine rock particles mixed with water poured down the rivers draining Mt. St. Helens. The largest mudflow occurred in the north fork of the Toutle River where it exceeded historic flood levels by nine metres.

The May 18 eruption displaced 2.7 cubic kilometres of volcanic rock, including 0.5 cubic kilometre of new magma, and destroyed the summit of Mt. St. Helens. Before the eruption the peak was 2,950 metres high (Figure 3). This was replaced by a horseshoe-shaped crater two kilometres across and roughly 750 metres deep, whose rim has an elevation of 2,400–2,550 metres—400 metres below the old summit elevation. Overall, Mt. St. Helens lost about 1,100 metres of elevation from the old peak to the new crater floor.

Formation of a lava dome

Following the great eruption of May 18, Mt. St. Helens produced five smaller explosive eruptions during the summer and fall of 1980. After 1980 a dome of lava grew intermittently in the crater. This lava is so viscous that it does not flow away; instead it piles up into a steep mound or dome over the vent. By 1986 the dome measured about 250 metres high and 1,000 metres wide, and it contained some 75,000,000 cubic metres of dacitic lava (see below *Determinants of size and shape;* see also Figure 4).

Explosive eruptions of fragmental volcanic rocks vary greatly in scale. Measured by the volume of new magma involved, the May 18 eruption of Mt. St. Helens was only a moderate eruption. The explosive eruption of Tambora Volcano in Indonesia in 1815 is estimated to have expelled from 50 to 100 cubic kilometres of magma, and the 1883 eruption of Krakatoa (spelled Krakatau in Bahasa Indonesia), another Indonesian volcano, about 18 cubic kilometres. Fortunately, as the eruptive magnitude increases, the frequency of eruptions greatly decreases. Civilizations have never been tested by a cataclysm on the scale of the eruption at Yellowstone about 2,000,000 years ago; that eruption involved nearly 3,000 cubic kilometres of explosively boiling magma.

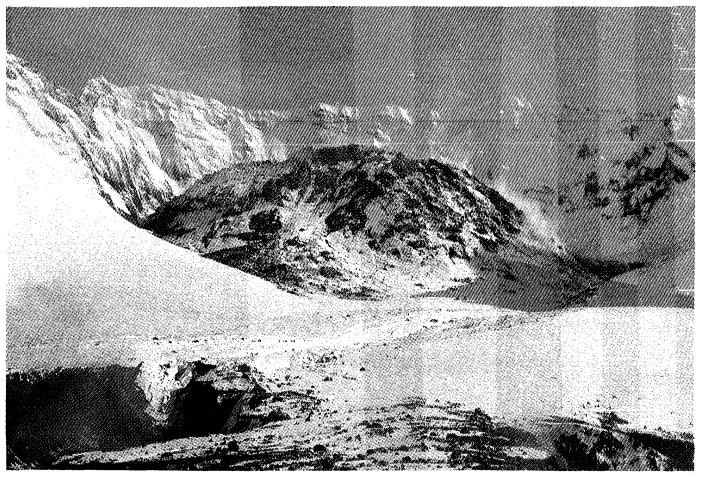


Figure 4: The lava dome of Mt. St. Helens on May 16, 1984, at which time it measured 850 metres wide and 220 metres high.
By courtesy of the U.S. Geological Survey; photograph, Lyn Topinka

Mauna Loa, Hawaii, 1984. The May 18, 1980, eruption of Mt. St. Helens is a classic example of explosive volcanism. The 1984 eruption of the great shield volcano Mauna Loa, by contrast, is illustrative of effusive volcanism.

Every three to four years, on average, Mauna Loa erupts with fountains and streams of incandescent lava. The summit of the volcano is 4,170 metres above sea level and more than 9,000 metres above the seafloor that surrounds the Hawaiian Ridge. Its volume above the seafloor, estimated to be about 40,000 cubic kilometres, qualifies it as the world's largest volcano.

Following a year of increased seismicity, Mauna Loa began erupting at 1:25 AM on March 25, 1984. The outbreak began along a fissure that split the long axis of the summit caldera, a cliff-bounded basin with a three-by-five-kilometre diameter that had been formed by prehistoric subsidence. Lava fountains along the fissure formed a curtain

Katia Krafft

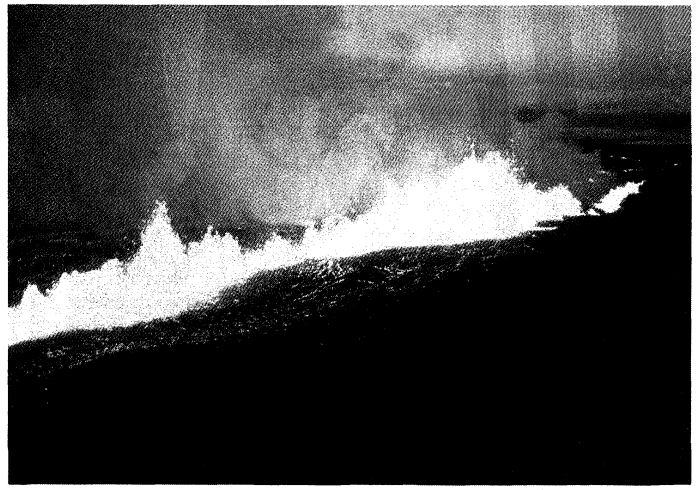


Figure 5: Curtain of fire formed by line of lava fountains along a fissure at an elevation of 3,800 metres on the northeast rift zone of Mauna Loa Volcano, Hawaii. The lava fountains are about 30–50 metres high.

Fiery lava fountains

of fire that illuminated the clouds and volcanic fume into a red glow backlighting the black profile of the volcano's huge but gently sloping summit. Lava from the summit fissure ponded in the caldera, and the first observers in the air reported that much of the caldera floor was covered by a lake of orange-red molten rock, which quickly cooled to a black crust with zigzag, still incandescent fractures.

At dawn the summit fissure began to propagate down the northeast rift zone, and a new line of lava fountains formed at an elevation of 3,800 metres (Figure 5). Two hours later the fracture broke an additional six kilometres down the northeast rift, forming another curtain of fire about two kilometres long and 50 metres high at an elevation of 3,450 metres. As new vents opened at lower elevations, the higher vents stopped erupting.

The vents at 3,450 metres continued to erupt throughout the early afternoon, sending a small lava flow down the high southeast flank of Mauna Loa. At about 4:00 PM the lava fountains dwindled, and a swarm of new earthquakes indicated that the fissure was breaking even farther down the rift. The new and final vents opened at 4:40 PM at an elevation of 2,900 metres, some seven kilometres down the ridge from the 3,450-metre vents.

The output of lava from these vents was vigorous. Although the fountains were only about 20 metres high, the volume of lava amounted to approximately 500,000 cubic metres per hour (about 17,600,000 cubic feet per hour). In 24 hours the river of lava flowed 12 kilometres northeast toward the city of Hilo. The temperature of the erupting lava was 1,140° C (2,084° F) and its viscosity about 10^3 poise (roughly the same as liquid honey at 20° C).

The fissure did not break any farther down the northeast rift zone of Mauna Loa, and the vents at 2,900 metres remained in steady eruption for the next 10 days. Even though the eruption rate remained high, the advance of the front of the lava flow slowed—six kilometres the second day, four kilometres the third day, and three kilometres the fourth day. This progressive slowing of the lava-flow front was caused by several factors: slow widening of flows at higher elevations; thickening of flows at higher elevations by overplating (*i.e.,* accumulation of new layers on top of layers only a few hours or days old); thickening and widening of flows at lower elevations where the slope of the land is more gradual; and branching of the flows upstream into new lobes that robbed the lower flows of their lava supply. A household analogue of a Hawaiian lava flow in miniature is the slow and erratic advance of molten wax as it adds new lobes to a pile of candle drippings.

By April 5 the output from the vents at 2,900 metres began to wane, and by April 15 the eruption was over. The longest flows traveled 27 kilometres, stopping at an elevation of 900 metres—10 kilometres from the outskirts of Hilo. The total volume of the eruption was 220,000,000 cubic metres, and new lava flows covered 48 square kilometres. No one was hurt, and the only significant damage was the cutting of power lines and the blocking of a few jeep roads.

Controlling factors in eruptions

Why are some volcanic eruptions so explosive while others are so spectacular but relatively harmless? The answer involves at least three factors: the amount of gas dissolved in the magma, the viscosity of the magma, and the rate of decompression of the magma as it rises toward the surface. Volcanoes related to converging plate margins (see below) generally have a high gas content and their magma is very viscous. This is an explosive combination because the gases cannot easily boil out; rather, they remain pent up until they reach the pressure at which they blow the viscous magma into fragments. The rate at which pressure is reduced also controls the explosiveness. If magma moves slowly toward the surface its dissolved gases are released slowly. At Mt. St. Helens the avalanche that initiated the eruption caused very rapid decompression of the shallow intrusion, resulting in violent and explosive boiling of the trapped gases (Figure 6).

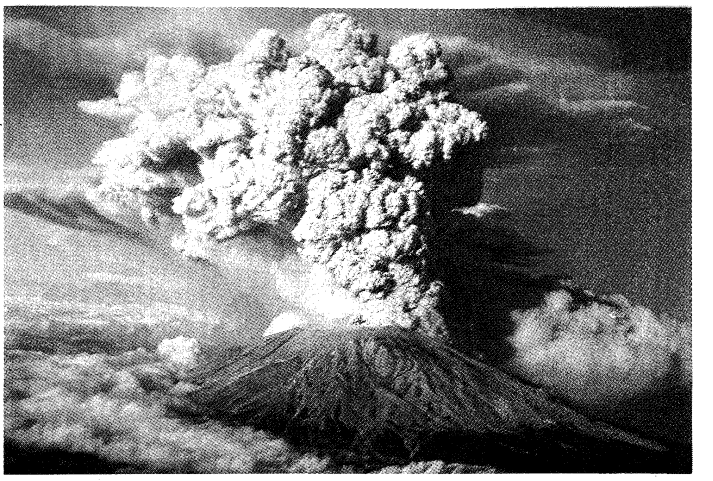


Figure 6: Explosive eruption of Mt. St. Helens on July 22, 1980, as viewed from the south.
Katia Krafft

OTHER MANIFESTATIONS OF VOLCANIC ACTIVITY

Volcanic gases. The most common volcanic gases are carbon dioxide, water, sulfur dioxide, and hydrogen sulfide. Small quantities of other volatile elements and compounds also are present; *e.g.,* hydrogen, helium, nitrogen, hydrogen chloride, hydrogen fluoride, and mercury. The specific gaseous compounds released from magma depend on temperature, pressure, and overall composition of the volatile elements present. The amount of available oxygen is of critical importance. When oxygen is lacking, methane, hydrogen, and hydrogen sulfide are stable, but when hot volcanic gases mix with atmospheric gases, carbon dioxide, water, and sulfur dioxide are stable.

Some volcanic gases are less soluble in magma than others and separate at higher pressures. Studies at Kilauea Volcano indicate that carbon dioxide begins to separate from its parent magma at depths of about 40 kilometres, whereas most of the sulfur gases and water are not released until the magma has nearly reached the surface. Fumaroles (volcanic gas vents) near Halemaumau Crater at Kilauea's summit are rich in carbon dioxide that leaks from the magma chamber located at three to four kilometres beneath the surface. Fumaroles on the rift zones of Kilauea, however, are richer in water and sulfur because much of the carbon dioxide leaks away at the summit before the magma is intruded laterally into the rift zones (Figure 7).

Hydrothermal features. Hot springs and geysers also are manifestations of volcanic activity. They result from the

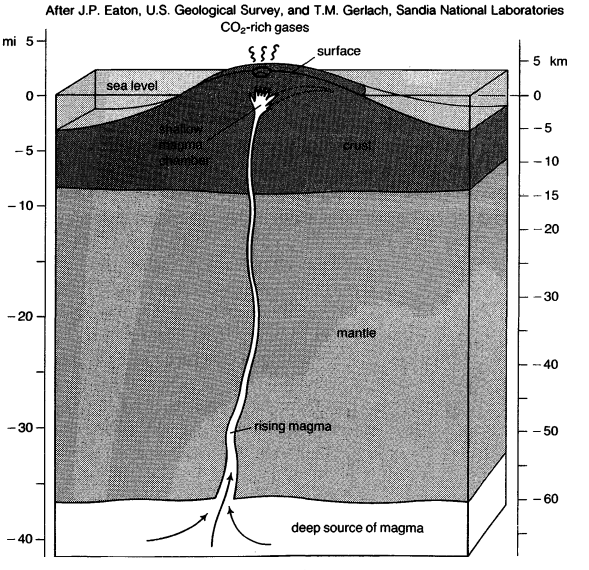


Figure 7: *Cross section of an active Hawaiian volcano.* The vertical scale is exaggerated by a factor of three. The dotted line extending to the right from the shallow magma chamber is a rift zone that receives intermittent lateral injections of magma from the chamber. Magma rising nearly continuously from the deep source is stored in the chamber. Gases rich in CO_2 escape through fumaroles above the magma chamber.

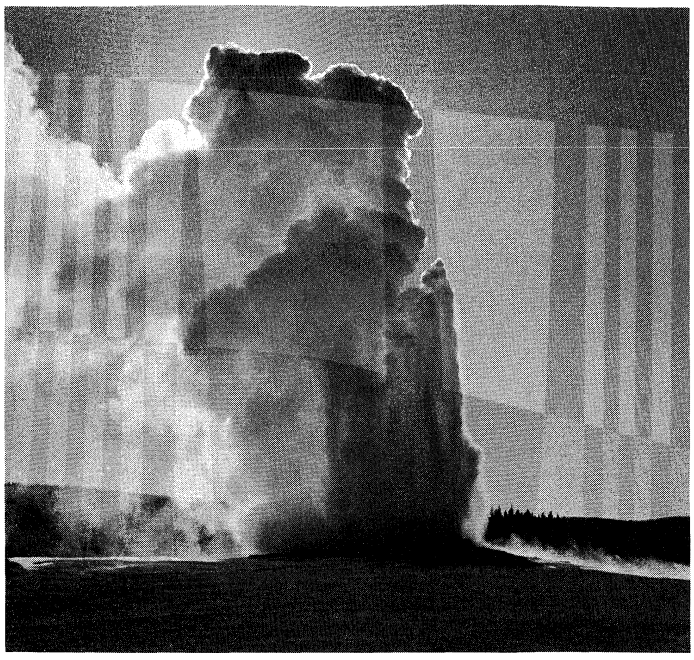


Figure 8: Old Faithful Geyser, Yellowstone National Park, Wyoming. Intermittent spouts of steam and hot water to heights of 35–50 metres occur on the average about every 60 minutes throughout the year.
R.W. Decker

Hot springs and geysers

interaction of groundwater with magma or with solidified but still hot igneous rocks at shallow depths.

Yellowstone National Park is one of the most famous areas of hot springs and geysers in the world. The total heat flux from these thermal features is estimated to be 300 megawatts. The last great eruption at Yellowstone occurred about 600,000 years ago when some 1,000 cubic kilometres of rhyolite magma were ejected in huge ash flows and resulted in the formation of a caldera approximately 45 by 75 kilometres in size. (A caldera is a large circular or oval depression caused by surface collapse into the subsurface region from which magma had been removed.) Yellowstone Lake now occupies part of this giant caldera. Since that last great outburst, about 1,200 cubic kilometres of rhyolite lava flows and domes have erupted in numerous smaller events. The cooling roots of such past eruptions, or possibly new intrusions of magma at shallow depth, are the apparent sources of heat for the Yellowstone hot springs and geysers.

Geysers are actually hot springs that intermittently spout a column of hot water and steam into the air. This action is caused by the water in deep conduits beneath a geyser approaching or reaching the boiling point. At 100 metres below the surface, the boiling point of water increases to approximately 150° C because of the increased pressure of the overlying hot water. As bubbles of steam or dissolved gas begin to form and hot water spills from the vent of a geyser, the pressure is lowered on the water column below. Water at depth then momentarily exceeds its boiling point and more water flashes into steam. This chain reaction continues until the geyser exhausts its supply of boiling water.

After a geyser stops spouting, the conduits at depth refill with groundwater and reheating to near the boiling point begins again. In geysers such as Yellowstone's Old Faithful, the spouting and recharge period is quite regular. This famous geyser has gushed to heights of 35–50 metres about every 40–80 minutes for more than 100 years (Figure 8). Other geysers have much more erratic recharge times.

Volcanic landforms

DETERMINANTS OF SIZE AND SHAPE

The common mental image of a volcano is that of a steep symmetrical cone sweeping upward in a concave curve to a sharp summit peak. Mt. Fuji in Japan is the archetypal volcanic landform, but only a few volcanoes attain this ideal shape.

Fundamental variables

The shape and size of a volcano are controlled by several factors. These include (1) the volume of volcanic products, (2) the composition of volcanic products, (3) the sequence of eruption (*e.g.*, many small eruptions with short repose times between eruptions, or a few large eruptions with long repose times), (4) the variety of volcanic eruption types, (5) the geometry of the vent, and (6) the environment into which the volcanic products are erupted.

Volumes released in any one eruption can vary enormously from a few cubic metres of magma to as much as 3,000 cubic kilometres. Small volumes tend to build up mounds close to the vent, whereas very large eruptions create plateaus of lava or ash flows. Eruptions of large ash-flow plateaus almost always form a caldera several kilometres in diameter over the eruption site.

The chemical composition of magma affects its physical properties, which in turn have a major influence on the landform built by a volcanic eruption. Four common volcanic rock types are shown in Table 1: basalt, andesite, dacite, and rhyolite. As the silica content increases, these rock types generally become more viscous; and as the magmatic gas content increases, they become more explosive. Effusive eruptions form lava plateaus or gently sloping shield volcanoes; moderately explosive eruptions form stratovolcanoes; and giant explosive eruptions form ash-flow plateaus. Naturally, since many other factors are involved in determining volcanic landforms, there are exceptions to these rules.

A series of small eruptions usually build up their products near their vents, whereas a large-volume eruption tends to disperse its products over a greater distance. Lava flows that erupt at high-volume rates usually travel greater distances from their vents. Other physical properties are, however, important in determining the character of lava flows. Hot basaltic lava forms pahoehoe flows (smooth to ropy surfaces), which, other conditions being equal, will flow farther than the cooler aa flows of the same composition (rough, broken surfaces; Figure 9).

Relation of volcanic landform to eruptive habit

If a volcano has consistent eruption habits, its landform will reflect that character. The shape of the huge but gently sloping shield volcano Mauna Loa, for example, indicates a long record of eruption of fluid lava flows from the summit and two persistent rift zones. It is these rift zones that give Mauna Loa its elongate shape.

Barbara Decker

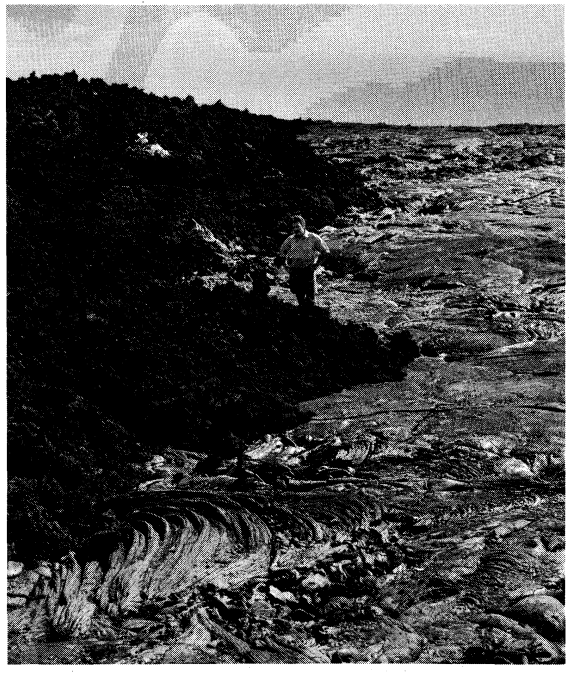


Figure 9: Aa lava flow (left) overlying pahoehoe lava flow on Kilauea Volcano, Hawaii. These basaltic lava flows were erupted in 1973 and were still largely barren of vegetation a decade later.

Table 1: Common Types of Volcanic Rock

name	silica (SiO_2) content (percent)	major minerals	colour	approximate density with no voids (g/cm^3)
Basalt	45–53	Ca feldspar, pyroxene, olivine	dark gray	3.0
Andesite	53–62	CaNa feldspar, pyroxene, amphibole	medium gray	2.9
Dacite	62–70	Na feldspar, amphibole, biotite, quartz	light gray* to tan	2.8
Rhyolite	70–78	K, Na feldspars, quartz, biotite	light gray* to pink	2.7

*Obsidian glass can be dark gray to black.

The beautiful symmetrical shape of the stratovolcano Mt. Fuji indicates a long record of moderately explosive eruptions from its summit, the eruptions generally consisting of explosions of fragments followed by thick viscous lava flows. Such alternating layers of ash and lava make up the strata denoted by the name stratovolcano. They are not layers in a blanketlike sense, however, but are overlapping lobes or tongues of ash and lava. For this reason, many geologists refer to stratovolcanoes as composite volcanoes.

In contrast to simple shield and stratovolcanoes, many volcanoes change their eruptive habits—both in eruption type and in the location of their vents—over time. This results in a mixture of volcanic landforms called a complex volcano.

The geometry of the vent or vents also exerts profound control on volcanic landforms. In Iceland the volcanic vents often are long fissures parallel to the rift zone. Renewed eruptions generally occur from new parallel fractures offset by a few hundred to thousands of metres from the earlier fissures. This distribution of vents and voluminous eruptions of fluid basaltic lava usually build up a thick lava plateau rather than a single volcanic edifice. Hekla Volcano in Iceland is transitional between plateau-building fissure eruptions and the single-major-vent eruptions that produce a symmetrical stratovolcano. Hekla erupts from a fissure that is parallel to the Mid-Atlantic Ridge and about 20 kilometres in length. Eruptions recur from this fracture—many from its central part. The result is a volcano shaped like an overturned canoe. Viewed along the fissure, Hekla looks like a stratovolcano; perpendicular to the fissure, it appears as an elongate ridge. Multiple point-source vents that erupt only once or a few times at most form a volcanic field of dozens of small cinder cones instead of a single large volcanic edifice.

Finally, but not of least importance, is the environment into which the volcanic products are erupted. Submarine volcanoes are surprisingly similar to their counterparts on land, but their slopes are generally steeper because water cools the lavas more rapidly. Deep submarine volcanism tends to be less explosive or nonexplosive, since the pressure of the water prevents explosive boiling.

Subglacial volcanism

Subglacial volcanism is drastically different in form from subaerial volcanism. This is particularly apparent in Iceland where glaciers covered the entire island 15,000 years ago and large ice caps still cover extensive areas today. Fissure eruptions beneath the ice form steep ridges of broken lava fragments rather than lava-flow plateaus, while subglacial eruptions from point-source vents that erupt repeatedly form table mountains. Table mountain volcanoes have steep sides of pillow lavas that formed in deep lakes or water-filled caverns within the ice. Pillow lavas are sacklike structures that form from underwater flows of basaltic lava. These structures are capped by several tens of metres of broken lava fragments from explosive shallow-water eruptions. The broken lava fragments in turn are overlain by shield-building lava flows erupted above the glacial surface.

MAJOR TYPES OF VOLCANIC LANDFORMS

Each of the more than 1,300 potentially active volcanoes or volcanic areas listed in the Smithsonian Institution's *Volcanoes of the World* (1981) has a distinct form, but most can be generalized into nine categories (Table 2). Some of these categories were already described, but they are treated below in order of their numerical importance.

Stratovolcanoes. These are steep volcanic cones built by both pyroclastic and lava-flow eruptions. They are composed of volcanic rock types that vary from basalt to rhyolite, but their composition is generally andesite. Stratovolcanoes may erupt many thousands of times over life spans of millions of years. A typical eruption would begin with ash explosions and end with extrusion of thick viscous lava flows. The cone-shaped form begins gradually and becomes steeper (up to 35°) toward the summit, which generally contains a crater.

Table 2: Number of Potentially Active Volcanoes of Various Landform Types*

type	number
Stratovolcanoes	577
Shield volcanoes	107
Submarine volcanoes	107
Calderas	81
Complex and compound volcanoes	76
Cinder cones	71
Volcanic fields	49
Fissure vents and crater rows	45
Lava domes	31

*Modified from *Volcanoes of the World* by T. Simkin *et al.* (1981); Smithsonian Institution.

Shield volcanoes. Structures of this type are large, dome-shaped mountains built of lava flows (the name derives from their similarity in shape to a warrior's shield lying face up on the ground). They are usually composed of basalt. Small shield volcanoes may form rapidly from almost continuous eruptions, but the larger shields are formed by hundreds of thousands of effusive eruptions of fluid lavas from their summits and rift zones over a span of about 1,000,000 years. The slopes of shield volcanoes are gentle and seldom exceed 6°. The summits are nearly flat but generally indented by cliff-walled craters or calderas.

Cone-shaped seamounts and guyots

Submarine volcanoes. These structures occur in various forms, but many are cone-shaped seamounts. Some ancient island volcanoes were eroded flat or covered with a coral cap at sea level before they sank. These flat-topped seamounts are called guyots. Most of the known active submarine volcanoes occur at shallow depths beneath the sea. They are known because their explosive eruptions can be detected and located by hydrophones. Active submarine volcanoes at depths of a few thousand metres are probably common, particularly along the mid-ocean ridges, but the water pressure at these depths prevents explosive boiling and the eruptions are undetectable. One exception to this is the submarine volcano Loihi, a seamount whose summit caldera occurs one kilometre below sea level, 30 kilometres southeast of the island of Hawaii. Although eruptions of this youngest volcano of the Hawaiian chain have not been directly observed, swarms of earthquakes at shallow depth beneath the summit of Loihi in 1971–72 and 1975 were recorded by seismographs. This is indirect evidence, but not proof, of submarine eruptions at Loihi Seamount.

Calderas. Large circular or oval depressions more than one kilometre in diameter, most calderas have formed by inward collapse. Many have steep cliffs bounding the depression, and some are filled with lakes. The terms crater and caldera are often used synonymously, but calderas are

larger than craters. A crater can occur inside a caldera, but not the other way around. Calderas are often associated with large eruptions (10 cubic kilometres or more) of dacitic or rhyolitic magma that form pyroclastic plateaus.

Calderas also occur on shield volcanoes and are thought to form when large rift eruptions or intrusions remove tremendous quantities of magma from the shallow magma chambers beneath the summit. The collapse and refilling of calderas on active Hawaiian volcanoes probably recur many times during a volcano's lifetime.

Whether a volcano is designated a caldera or shield or stratovolcano with a caldera depends on the principal landform feature. For example, Crater Lake in Oregon in the northwestern United States is designated a caldera in Table 2, but Kilauea Volcano in Hawaii is designated a shield volcano even though it has a large summit caldera.

Complex volcanoes. Such structures are mixed landforms. In most cases, they occur because of changes either in eruptive habit or in location of the principal vent area. A stratovolcano may form a large explosion crater that later becomes filled by a lava dome, or several new cones and craters may develop on a caldera's rim. One stratovolcano cone may overlap another and have multiple summits. The Three Sisters volcanic complex in Oregon is an example of a complex volcano with three summits.

Structures of loose pyroclastic fragments

Cinder cones. These are relatively small volcanic landforms: steep, angle-of-repose (about 30°) cones of loose pyroclastic fragments, most of which are in the cinder-size range (Table 3). Generally, the crater from which the cinder-sized fragments were ejected is located in the centre of the cone. In areas with strong prevailing winds, however, the crater may be upwind of the cone. The rock type involved in cinder cone volcanoes is generally basalt or basaltic andesite, and the eruption type is Vulcanian or high-lava-fountain Hawaiian.

Table 3: Volcanic Products

form	name	characteristics (including dimensions)
Gas	fume	
Liquid	lavas	
	aa	rough, blocky surface
	pahoehoe	smooth to ropy surface
Solid	airfall fragments	
	dust	< 1/16 mm
	ash	1/16–2 mm
	cinders	2–64 mm
	blocks	> 64 mm solid
	bombs	> 64 mm plastic
	pyroclastic flows	flows fluidized by hot gases
	mudflows	flows fluidized by rainfall, melting ice and snow, or ejected crater lakes

Some cinder cones such as Parícutin in Mexico form during a single eruption. Parícutin is approximately 410 metres high and one kilometre wide; it formed during nearly continuous eruptions from 1943 to 1952. Cinder cones also form at some vents on shield volcanoes, but they are not included in Table 2 because they are not separate, individual volcanoes. Certain cinder cones have multiple eruptions, but if activity continues for thousands to tens of thousands of years from the same vent, they probably develop into stratovolcanoes or complex volcanoes.

Volcanic fields. Such areas have many geologically young cinder cones or other features that have not been individually identified as separate volcanoes. If the conduits through which magma ascends to the surface are scattered over a broad area, many short-lived volcanoes are formed rather than a major volcano with repeated eruptions. The area in which Parícutin formed is a volcanic field with dozens of prehistoric—but geologically young—cinder cones and lava flows. The most likely place for the birth of a new volcano is in a known volcanic field.

Association with dikes

Fissure vents. These features constitute the surface trace of dikes (underground fractures filled with magma). Most dikes measure about one-half to two metres in width and several kilometres in length. The dikes that feed fissure vents reach the surface from depths of a few kilometres.

Fissure vents are common in Iceland and along the radial rift zones of Hawaiian volcanoes. Curtain-of-fire eruptions from fissure vents produce low ramparts of basaltic spatter on both sides of the fissure. More isolated lava fountains along the fissure produce crater rows of small spatter and cinder cones. The fragments that form a spatter cone are hot and plastic enough to weld together, while the fragments that form a loose cinder cone have cooled enough during their rise and fall not to stick together.

The largest effusive eruption of lava in recorded history occurred in 1783 in Iceland from the Laki fissure. This vent produced high lava fountains, a 25-kilometre-long crater row, and 565 square kilometres of basaltic lava flows with a volume of approximately 12 cubic kilometres.

Lava domes. Landforms of this sort consist of steep domal mounds of lava so viscous that the lava piles up over its vent without flowing away. The rock types that form lava domes are generally andesites, dacites, or rhyolites. Somehow these viscous lavas have lost much of their gas content in prior eruptions or by slow rise to the surface. Even so, it is not unusual for an actively growing lava dome to have an explosive eruption that disrupts all or part of the dome. Many lava domes grow by internal intrusion of lava that causes swelling and oversteepening of the dome. Rockslides build up an apron of talus blocks around the lower sides of the dome. Lava domes can form mounds several hundred metres high and from several hundred to more than 1,000 metres in diameter. Thick lava flows sometimes move short distances from the dome and distort its generally circular or oval shape. The dome currently forming in the explosion crater at Mt. St. Helens is a good example of a lava dome.

Other volcanic structures and features. Less common volcanic forms have also been described, and sometimes other names have been used for those cited above. Such inconsistency of nomenclature stems from the fact that no one volcanologist has examined all the world's volcanoes, and the descriptions that exist have been made by many people with diverse backgrounds. Other types and terms that may be encountered include volcanic cones, simply a descriptive term pertaining to shape with no implication of size, rock type, or genesis; explosion crater, a large circular, elongate, or horseshoe-shaped excavation with ejected debris on its rim or flanks; hydrothermal regions, fumarole fields, and solfatara fields, areas of hot springs and gas vents where magma or still-hot igneous rocks at shallow depth are leaking gases or interacting with the groundwater system; maar, a low-relief crater often filled with water and surrounded by a rim of ejected material that was probably formed by explosive interaction of magma and groundwater; tuff rings and tuff cones (Figure 10), landforms that resemble maars but that have higher rims and are not filled with water (a tuff is a compacted pyroclastic deposit that often includes some sediments); pumice cone, a structure similar to a cinder cone but made up of volcanic glass fragments so filled with gas-bubble holes that it resembles a sponge and is very lightweight; pyroclastic cone, a cone-shaped volcano composed almost

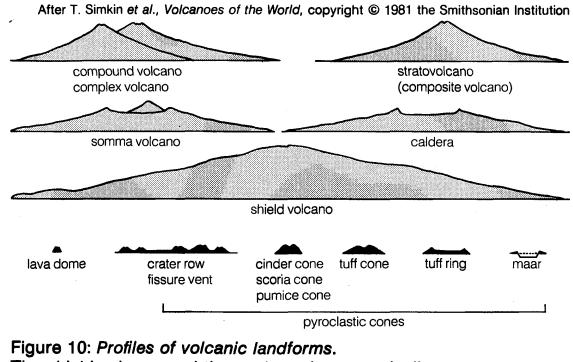


Figure 10: *Profiles of volcanic landforms.* The shield volcano and those above it are vertically exaggerated by a factor of two. The landforms below the shield volcano are vertically exaggerated by a factor of four. Relative sizes are approximate because of large variations in size within each type.

entirely of pyroclastic material; scoria cone, a synonym for cinder cone (cinders or scoriae are vesicular pyroclastic fragments usually heavier than pumice); somma volcano, a caldera partially filled by a new central cone, named for Mt. Somma in Italy; and subglacial volcano, a volcanic form produced by eruptions beneath a glacier or beneath the surface of a lake within a glacier.

Taking a more distant view of volcanic landforms from space, one can see that individual volcanoes form linear to arcuate belts across the Earth's surface. It is now clear that these volcanic chains are closely related to global tectonic activity.

Volcanism and tectonic activity

Distribution of volcanoes

Active volcanoes are not scattered over the Earth randomly; instead, most occur in belts, especially in the island arcs and mountain ranges bordering the Pacific Ocean (Figure 11). The concept of seafloor spreading and, more broadly, the theory of plate tectonics offer a logical explanation for the location of most volcanoes.

VOLCANOES RELATED TO PLATE BOUNDARIES

Figure 12 shows the Earth's topography as well as the locations of larger earthquakes that occurred from 1960 to 1980. This map clearly indicates the boundaries of the 12 major tectonic plates. For example, the Pacific Plate is bounded by the earthquake zones of New Zealand, New Guinea, the Mariana Islands, Japan, Kamchatka, the Aleutian Islands, western North America, the East Pacific Rise, and the Pacific-Antarctic Ridge.

The Earth's plates, which move horizontally with respect to one another at a rate of a few centimetres per year, form three basic types of boundaries: convergent, divergent, and side-slipping. Japan and the Aleutian Islands are examples of convergent boundaries where the Pacific Plate is pushing beneath the adjacent continental plates—a process known as subduction. The San Andreas Fault system in California exemplifies a side-slipping boundary where the Pacific Plate is moving northwest relative to the North American Plate—a process called strike-slip, or transform, faulting. The East Pacific Rise is representative of a divergent boundary where the Pacific Plate and the Nazca Plate (west of South America) are pulling apart—a process known as rifting.

Volcanoes occur along both subduction and rift zones but are generally absent along strike-slip plate margins. Of the 1,144 volcanoes listed in Table 2, 80 percent occur along subduction zones and 15 percent along rift zones. These percentages are somewhat misleading, however, because most of the Earth's rift zones are about two to three kilometres below sea level. At those inaccessible depths active submarine volcanoes have yet to be observed, though a few hydrothermal areas have been found along submarine rift zones with deep-diving submarines.

Iceland, a rare above-water section of the Mid-Atlantic Ridge, has 70 volcanoes that have erupted during the past 10,000 years. If this is a typical number for a short section of the world rift system, it suggests that there may be several thousand potentially active volcanoes along the largely submarine world rift system that follows the crest of the mid-ocean ridges.

Subduction-related volcanoes

Most subduction-related volcanoes are explosive and build stratovolcanoes, while rift volcanoes tend to be more effusive and build shield volcanoes, though there are exceptions to both these generalities. Subduction-related volcanoes erupt basalt, andesite, dacite, and rhyolite, with andesite the predominant rock type. Rift-related volcanoes, especially on the ocean floor, erupt mainly basalt.

Figure 13 is a schematic cross section showing conceptual models of how subduction and rift volcanoes may form. As an oceanic plate (2) is subducted beneath a continental plate (3), seafloor sediments rich in water and carbon dioxide are carried beneath the overriding plate. These compounds may act as fluxes and reduce the melting temperature of magma. Frictional heating also may be important in the formation of subduction-zone magmas, but since the subducted plate is colder than the mantle rocks it is penetrating, this would counteract any frictional heating. Although the process is not clearly understood, magma apparently forms and rises by buoyancy from a depth of 100 to 200 kilometres. Subduction-zone volca-

After T. Simkin *et al., Volcanoes of the World*, copyright ©1981 the Smithsonian Institution

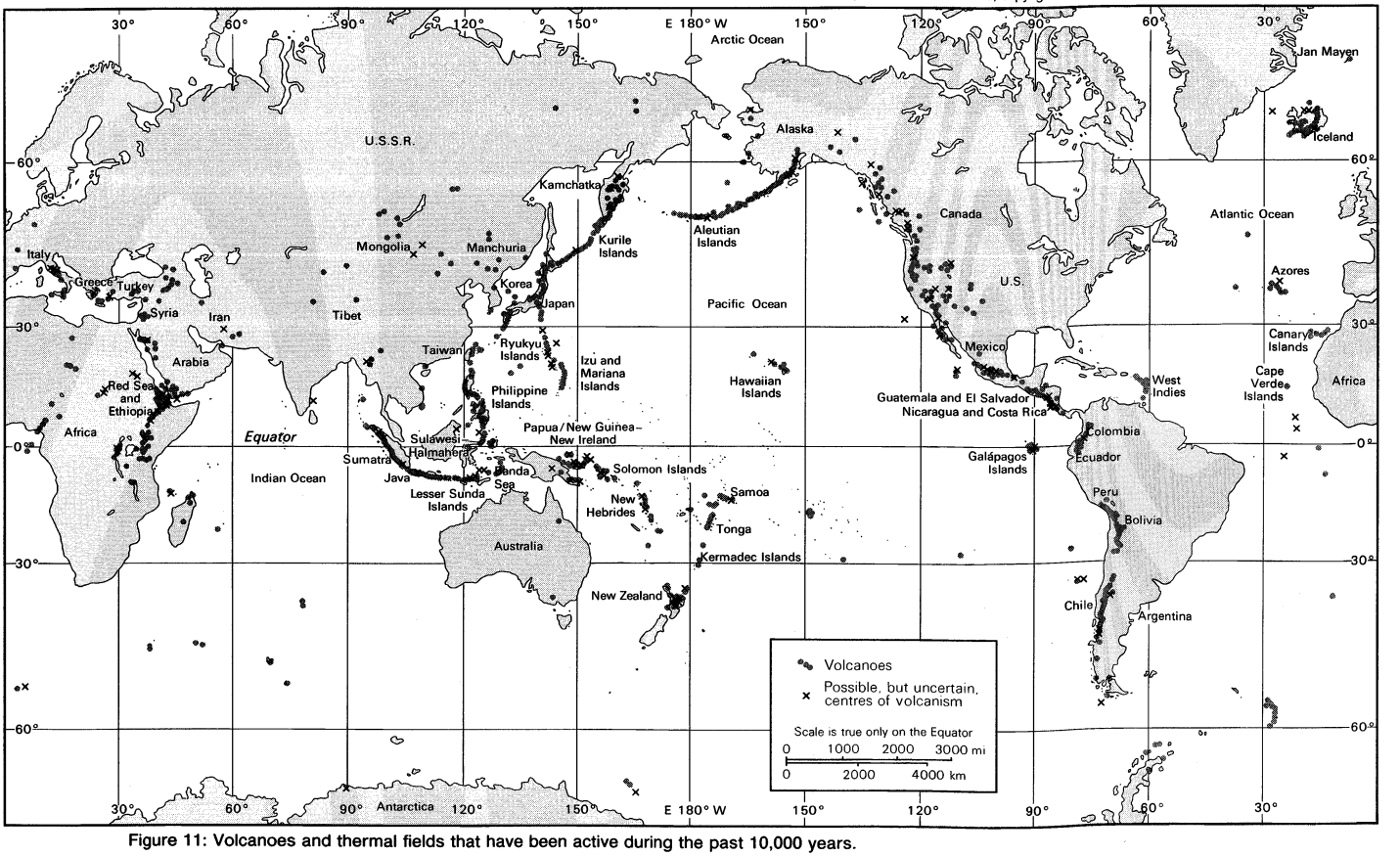


Figure 11: Volcanoes and thermal fields that have been active during the past 10,000 years.

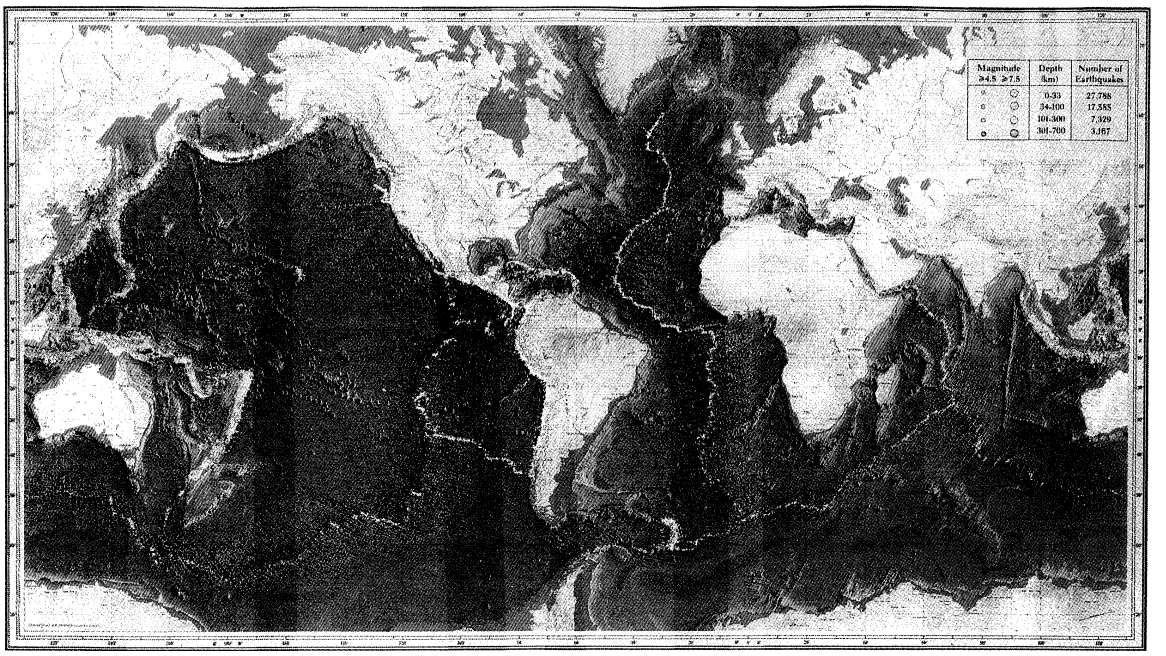


Figure 12: *Seismicity and topography of the Earth.*
The epicentres of the larger earthquakes that occurred from 1960 to 1980 are shown as red dots in this topographic map of the continents and ocean floor. The dots sharply delineate the boundaries of the principal tectonic plates of the Earth's surface.

Seismicity of the Earth by A.F. Espinosa, W. Rinehart, and M. Tharp, copyright 1982 by Rebecca M. Espinosa and Marie Tharp, based on the World Ocean Floor map by B.C. Heezen and M. Tharp, copyright 1977 by Marie Tharp; reproduced with permission, all rights reserved

noes occur on the overriding plate and are offset inland from the actual plate boundary along the ocean trench.

The rising subduction-zone magma is probably basaltic in composition and is formed by the partial melting of mantle rocks. As the rising magma moves slowly up through the continental crust of the overriding plate, however, two things may occur to increase significantly the silica content of the magma. Crystallization of olivine and pyroxene minerals from the basalt can leave the residual melt enriched in silica and depleted in magnesium, iron, and calcium. This process is called fractional crystallization. Also, basaltic magmas have enough excess heat to partially melt the continental host rocks through which they are ascending. Because continental rocks are generally higher in silica, potassium, and sodium than are oceanic rocks, this process of assimilation and mixing can also play an important role in producing the wide range of compositions that occur in rocks from subduction volcanoes.

The additional gas content of many magma batches at subduction volcanoes (which, coupled with their often high-viscosity magma, makes them dangerously explosive) also may be explained by more than one process. Additional water and carbon dioxide may come from both subducted seafloor sediments and assimilated crustal rocks. Furthermore, any fractional crystallization tends to concentrate volatile elements in the residual melt. If volcanic gases form separate fluid phases within batches of ascending magma (carbon dioxide is most likely to do this), these fluid phases may ascend more rapidly than the overall magma body and be concentrated in the upper portion. Explosive boiling of these volcanic gases at atmospheric pressure is the apparent reason for the highly explosive nature of many subduction volcanoes.

After L. Sykes *et al.*, in F. Press and R. Siever, *Earth*, second edition; copyright © 1974, 1978 W.H. Freeman and Company, used by permission

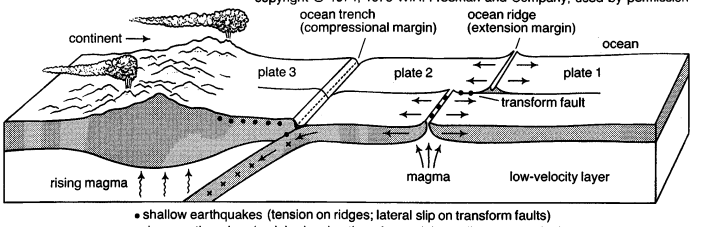


Figure 13: Cross section of plate boundaries formed by ocean ridges (rift zones), strike-slip (transform) faults, and subduction zones.

Rift volcanoes, as shown in Figure 13, form when magma rises into the gap between diverging plates (1 and 2). They thus occur at or near actual plate boundaries. Measurements in Iceland suggest that the separation of plates is a continuous process but that the fracturing is intermittent. An analogy would be a rubber band that is slowly stretched until it snaps. Earthquake swarms and volcanic eruptions occur when the stretching exceeds the strength of the near-surface rocks, which then fracture along steeply dipping cracks parallel to the rift. Basaltic magma rising along these fractures causes Icelandic fissure eruptions. Rift volcanoes

Rift volcanoes in continental locations such as East Africa are more complex. Assimilation of continental crust apparently gives them some of the characteristics more generally associated with subduction volcanoes; *e.g.*, having a wider range of rock types and explosive habits.

To return to the example of the Pacific Plate in Figure 12, the volcanoes on the western and northern margin (New Zealand, New Guinea, Mariana Islands, Japan, Kamchatka, and the Aleutian Islands) are all subduction volcanoes. The rift volcanoes, except for a few near Baja California, are hidden along the submarine crest of the East Pacific Rise and the Pacific-Antarctic Ridge at depths of two to three kilometres below sea level. The Cascade volcanoes in the northwestern United States and the volcanoes in Mexico and Central America are subduction volcanoes related to the small Juan de Fuca and Cocos plates on the east side of the Pacific Plate. Similarly, the volcanoes of the Andes are related to the subduction of the Nazca Plate beneath the South American Plate.

INTRAPLATE VOLCANISM

A red dot representing an earthquake location in the north central Pacific in Figure 12 marks the volcanic island of Hawaii, far from any plate margin. How can it be explained in terms of plate tectonics?

The 5 percent of known world volcanoes that are not closely related to plate margins are generally regarded as hot-spot volcanoes. Hawaiian volcanoes are the best examples of this type, and though they occur near the centre Hot-spot volcanoes

of the northern portion of the Pacific Plate, they show a dramatic relationship to the past movements of that plate. Kilauea and Mauna Loa volcanoes on the island of Hawaii at the southeast end of the Hawaiian volcanic chain are two of the most active volcanoes in the world. In addition, the five volcanoes that form the island of Hawaii are all less than 1,000,000 years old.

Northwestward along the Hawaiian chain each island is progressively older. The extinct volcano that formed the island of Kauai is about 5,000,000 years old. Assuming that the hot spot that generates Hawaiian volcanoes is relatively fixed in place—and there is good evidence to support this assumption—then the movement of the Pacific Plate during the past several million years has been northwestward at a rate of about 10 centimetres per year. Figure 12 shows a major submarine continuation of the Hawaiian Ridge to the northwest of the Hawaiian Islands, and then a dogleg bend into the Emperor Seamounts, which comprise an entirely submarine ridge continuing northward to the edge of the Pacific Plate.

Ages of rocks obtained by dredging and drilling the Emperor Seamounts indicate that the Hawaiian-Emperor Ridge is a progressively older volcanic chain formed by volcanism at the Hawaiian hot spot. This major centre of volcanism in the Pacific has been active for at least 80,000,000 years, and the Pacific Plate has moved over it, first northward and later northwestward, at a rate of approximately eight to 10 centimetres per year. The bend between the Emperor Seamounts and the Hawaiian Ridge occurred about 40,000,000 years ago and indicates a significant shift in the direction of movement of the Pacific Plate.

It is not known how a volcanic hot spot maintains its position for millions of years while a plate passes over it. One concept suggests that a hot spot is a deep-mantle plume (Figure 14), which is caused by very slow convection of highly viscous mantle material. As hot but solid mantle rock moves upward, partial melting may occur from the lowering of its pressure-dependent melting temperature. Detailed seismic sounding of the mantle over the next decade or two should help to resolve this problem of the mechanism of hot spots.

Deep-mantle plumes

Small and isolated intraplate volcanoes may simply be the result of deep fractures within the plates that allow pockets of partial melt in the low-velocity layer below to leak to the surface. The low-velocity layer begins about 50 to 150 kilometres below the surface and extends to a depth of roughly 300 kilometres; it is so named because earthquake waves travel more slowly in this hot, low-strength layer than in the overlying rigid plates (Figure 13). Some studies indicate that there may be a few percent partial melt within the low-velocity layer. Once a sufficient volume of magma forms in the subsurface, it tends to rise from its own buoyancy. Any fracture system at the plate margins or within the plates will facilitate this process.

After G.B. Dalrymple, E.A. Silver, and E.D. Jackson, "Origin of the Hawaiian Islands," *American Scientist* (May–June 1973)

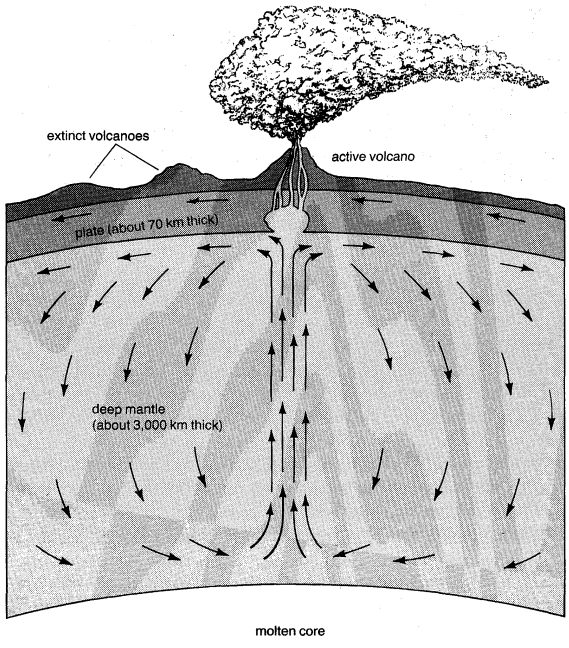


Figure 14: Hot-spot volcanoes may be formed by deep mantle plumes generated by slow convection currents.

Impact of volcanic activity on humankind and the environment

Since the late 1700s volcanoes have caused more than 250,000 deaths. Most of these occurred during four disastrous eruptions.

Catastrophic eruptions of modern times

The largest eruption of the four occurred at Tambora Volcano on Sumbawa Island, Indonesia, on April 10–11, 1815. Fifty cubic kilometres of magma were expelled in Plinian ash clouds and pyroclastic flows. Ash greater than one centimetre thick fell on more than 500,000 square kilometres of Indonesia and the Java Sea. Before the eruption Tambora was a 4,300-metre-high stratovolcano; following the eruption 1,400 metres of the summit cone were missing and in its place was a collapsed caldera measuring six by seven kilometres wide and one kilometre deep. About 10,000 people were killed by the explosive eruption and the tsunamis caused by massive pyroclastic flows entering the sea. Agricultural losses from the thick ash deposits resulted in famine and disease, leading to an additional 82,000 deaths.

The second largest eruption of the 19th century also occurred in Indonesia. Krakatoa (or Krakatau), a compound volcano on a small uninhabited island between Sumatra and Java, erupted explosively on Aug. 26–27, 1883. The eruption was similar to the Tambora outburst but smaller, involving only about 18 cubic kilometres of magma erupted in Plinian ash clouds and pyroclastic flows. Krakatoa was a much smaller volcano than Tambora, and when the eruption had emptied its magma chamber, it collapsed to form a caldera that was partly below sea level. Twenty-three square kilometres of the island of Krakatoa disappeared, and where a volcanic peak 450 metres high once stood was water as deep as 275 metres. The largest explosion on the morning of August 27 produced an ash cloud that was reported to have reached 80 kilometres high, and the detonation was heard in Australia, 4,800 kilometres away. A tsunami, over 30 metres high, followed the explosion and apparent caldera collapse, killing about 36,000 people on the adjacent shores of Java and Sumatra.

On May 8, 1902, there occurred a violent eruption of Mt. Pelée, a stratovolcano on the island of Martinique in the Caribbean Sea. Although less than one cubic kilometre of magma was erupted, much of it formed a high-velocity glowing avalanche of hot gases and pyroclastic fragments that swept down a steep valley to the port of St. Pierre. Within minutes the town and virtually all of its inhabitants (some 29,000 people) were incinerated.

The second worst volcanic disaster of the 20th century occurred on Nov. 13, 1985, when a relatively small eruption of Nevado del Ruiz, a stratovolcano in Colombia, killed 25,000 people. This 5,400-metre-high volcano has a glacial ice cap, and when a brief explosive eruption dumped several million cubic metres of hot pyroclastic fragments on the ice cap surrounding the summit crater, a sudden surge of meltwater sent massive mudflows down canyons on both the east and west sides of the Andean volcano. Much of the town of Armero, built on a low plain beside the Lagunilla (also spelled Lagunillas) River 50 kilometres east of and nearly five kilometres below the summit of Nevado del Ruiz, was buried by the mudflows. Twenty-two thousand of its inhabitants were killed.

About 70 percent of the people who died from volcanic eruptions in the past 200 years perished in those four outbursts. The remaining 30 percent were killed in many other less devastating eruptions. As world population increases, however, so does the risk of greater loss of life from volcanic eruptions. This was made all too clear by the tragedy at Armero. In 1845 a mudflow from Nevado del Ruiz killed approximately 1,000 people on farms near the site where the town of Armero was later built. In the

1985 mudflow, which was smaller in volume than the 1845 mudflow, more than 20 times as many people were killed.

Property damage from volcanic eruptions is difficult to estimate because of differing value systems and changes in land use. One study estimates an average of $100,000,000 per year in property damage worldwide from volcanic eruptions. As with casualties, a few eruptions cause staggering damage, while most are much less destructive. The Mt. St. Helens eruption in 1980 did more than $1,000,000,000 worth of damage, mainly to the timber industry.

HAZARDS FROM VARIOUS TYPES OF VOLCANIC ERUPTIONS

The list of hazards associated with volcanic eruptions is long and varied: explosions, toxic gas clouds, ash falls, pyroclastic flows, avalanches, tsunamis, mudflows, and lava flows, as well as secondary effects such as fluoride poisoning, starvation, and disease.

The sudden depressurization of a shallow hydrothermal system or gas-charged magma body or the rapid mixing of magma with groundwater can cause rapid gas expansion and massive explosions. Large blocks ejected in these explosions are sometimes hurled as far as 20 kilometres from the explosive vent. A directed blast in which one side of a volcanic cone fails, as happened at Mt. St. Helens in 1980, can cause destruction of several hundred square kilometres on the failed flank of the volcano. This is especially true if the blast cloud is heavily laden with fragmental debris and becomes a dense fluidized flow. It then takes on characteristics similar to a pyroclastic flow. Even beyond the limit of explosive destruction, the hot, ash-laden gas clouds associated with an explosive eruption can scorch vegetation and kill people by suffocation.

Toxic gas emissions and ash falls

Gas clouds from a sudden increase in fumarole emissions may also contain poisonous or suffocating gases such as hydrogen sulfide, carbon monoxide, carbon dioxide, and sulfur dioxide. At a crater lake in Cameroon, West Africa, more than 1,700 people were killed by a sudden release of carbon dioxide in August 1986.

Ash falls from continued explosive jetting of fine volcanic particles into high ash clouds generally do not cause any direct fatalities. Yet, where the ash accumulates more than a few centimetres, collapsing roofs from ash loads and failure of crops are major secondary hazards. Crop failure can occur over large areas downwind from major Plinian eruptions, and widespread famine and disease may result, especially in poorly developed countries. Lightning strikes are common close to Plinian ash clouds, adding one more element of terror to an explosive eruption.

Pyroclastic flows

Pyroclastic flows are the most dangerous and destructive aspect of explosive volcanism. They occur in many sizes and types, but their common characteristic is that they form a fluidized emulsion of volcanic particles, eruption gases, and entrapped air, resulting in a flow of low enough viscosity to be very mobile and of high enough density to hug the ground surface. A pyroclastic flow can pour over the lip of an erupting vent, or it may form when an ash cloud becomes too dense to continue rising and falls back to the ground. In major caldera collapses associated with explosive volcanoes, huge pyroclastic flows may issue from the ring fractures as the caldera block subsides.

Pyroclastic flows can move at speeds up to a few kilometres per minute and have temperatures ranging from 100° to 700° C. They sweep away and incinerate nearly everything in their path. Smaller pyroclastic flows are confined to valleys and are called *nuées ardentes*. Large pyroclastic flows may spread out as a blanket deposit across many hundreds or even thousands of square kilometres around a major caldera collapse. During the past 2,000,000 years, the Yellowstone National Park area has undergone three major caldera collapses involving pyroclastic eruptions of 300 to 3,000 cubic kilometres of rhyolitic magma.

Avalanches of rock and ice also are common on active volcanoes. They may occur with or without an eruption. Those without an eruption are often triggered by earthquakes, weakening of rock into clay by hydrothermal activity, or by heavy rainfall or snowfall. Those associated with eruptions are sometimes caused by oversteepening of a volcano's flank by intrusion of a shallow body of magma within or just beneath the volcanic cone, as happened at Mt. St. Helens.

A caldera collapse that is in part or entirely submarine usually generates a tsunami; the larger and more rapid the collapse, the larger the tsunami. Tsunamis also can be caused by avalanches or large pyroclastic flows rapidly entering the sea on the flank of a volcano.

Giant mudflows

Mudflows are common hazards associated with stratovolcanoes and can happen even without an eruption. They occur whenever floods of water mixed with ash, loose soil, or rocks that have been altered to clay by hydrothermal activity sweep down the valleys draining the sides of large stratovolcanoes. The huge mudflows generated by meltwater from the ice cap of Nevado del Ruiz Volcano in 1985 are classic examples of mudflows associated with eruptions. Heavy rainfall or earthquake-induced avalanches of ice or hydrothermal clay also can cause mudflows on steep volcanoes during periods of repose.

Lava flows

A lava flow engulfs and buries the land it covers, but new soil and vegetation eventually develop again. In warm humid climates the recovery is rapid; a few decades will hide the rocky surface of flows. In desert or arctic climates recovery is slower; flows more than 1,000 years old may still retain their barren appearance.

The greatest hazard at potentially active volcanoes is human complacency. The physical hazards can be reliably estimated by studying the past eruptive activity as recorded in history or in the prehistoric deposits around a volcano. Volcano observatories can monitor the local earthquake activity and the surface deformation of a potentially active volcano and make useful, if not yet precise, forecasts of eruptions. Increased earthquake activity beneath Mauna Loa in 1983 led to a forecast of an increase in probability of an eruption for 1984 or 1985; an eruption occurred in March 1984. The major eruption of Mt. St. Helens on May 18, 1980, was much larger than anticipated, but the high level of local earthquakes and the bulge forming on the north flank of the mountain provided enough warning to encourage a partial evacuation of the surrounding area. Some lives were lost, but the toll would have been much higher if access to the area had not been restricted by the local authorities. A major problem in reducing volcanic risk is that most explosive volcanoes have such long repose periods that people living nearby consider them extinct rather than dormant.

EFFECTS OF VOLCANIC ERUPTIONS ON SOIL AND CLIMATE

Not all volcanic phenomena are destructive. The oceans, atmosphere, and continents owe their origin and evolution in large measure to volcanic processes throughout geologic time.

Soil enrichment

Volcanic ash slowly weathers to form rich loamy soils. On volcanic Java, terraced rice paddies support a dense population. Across the Java Sea is Kalimantan (part of Borneo) with a similar climate but no volcanoes. The jungles of Kalimantan provide only temporary slash-and-burn agriculture and support a much smaller population.

Climate, too, is subject to the effects of volcanic activity. High ash clouds from Plinian eruptions, especially if they are rich in sulfur dioxide, can inject much fine dust and aerosol droplets of sulfuric acid into the stratosphere. Above rain clouds, these fine particles are not washed quickly back to Earth but spread slowly into haze layers that can blanket a hemisphere or even the entire Earth.

World climate seems to have been affected by the eruptions of Krakatoa in 1883 and of Agung Volcano in Bali in 1963 in the form of a lowering of average world temperature by about 0.5° C over the few years following these eruptions. Although world temperature data was poorly recorded in the early 1800s, the eruption of Tambora in April 1815 was followed in 1816 in North America and Europe by what was called "the year without a summer." Other large volcanic eruptions such as Katmai in Alaska in 1912, however, appear to have produced no cooling effect. Records of average world temperature over the past several decades often show changes of 0.1°–0.3° C from year to year unrelated to any known volcanic eruptions, so it is difficult to establish clearly whether or not volcanic eruptions have a major impact on climate.

Direct sampling of the stratosphere has shown that the major haze-forming agent from volcanic eruptions is not fine dust but an aerosol of tiny sulfuric acid droplets. This indicates that the composition of high volcanic ash clouds may be as important as their volume in affecting climate. Atmospheric chemists are discovering that not only volcanic eruptions but also man-made aerosols of chlorofluoromethanes (commonly called Freon), exhaust from high-altitude jet aircraft, and a general increase in carbon dioxide from burning fossil fuels are perturbing the atmosphere. The Earth has many buffers that maintain its environment, but their interactions are not well understood. Questions as to how volcanic and human activity affect climate remain essentially unanswered, and they are important problems for future research and for survival.

GEOTHERMAL ENERGY

Geothermal energy is plentiful, but geothermal power is not. Temperatures increase below the Earth's surface at a rate of about 30° C per kilometre in the first 10 kilometres. This internal heat of the Earth is an immense store of energy. In the upper 10 kilometres of rock beneath the conterminous United States, it amounts to 3.3×10^{25} joules, or about 6,000 times the energy contained in the world's oil reserves. The problem in utilizing geothermal energy is extracting it.

The natural escape of the Earth's heat through its surface averages only 0.06 watt per square metre. To make geothermal power practical, some special situation must exist to concentrate the Earth's heat energy in a small area. Underground reservoirs of steam or hot water that can be funneled into a drill hole provide this special situation.

Geothermal steam wells

Some geothermal steam wells can produce six megawatts of thermal power, an amount equal to the normal heat flux from 100 square kilometres of land surface. The key to this concentration is the mass transfer of heat from deeper levels to the near surface by volcanism. Magma at temperatures close to 1,200° C moves upward to depths of only a few kilometres, where it transfers heat by conduction to groundwater. The groundwater then circulates by convection and forms large underground reservoirs of hot water and steam. Some of this thermal water may escape to the surface as hot springs or geysers.

Holes drilled into a subsurface geothermal system allow rapid mass transfer of hot water or steam to the surface. At The Geysers geothermal field north of San Francisco, superheated steam is directly tapped from porous underground reservoirs. In most other geothermal fields, the hot water is at or below its subsurface boiling temperature—about 300° C at a depth of one kilometre. Hot water that is near its subsurface boiling temperature partially flashes to steam at the reduced pressures in a drill hole. This mixture of steam and hot water is separated at the well head, and the steam is used to drive the turbine of an electric generator. Hot water from lower temperature geothermal reservoirs can be used for space heating. This form of geothermal power is utilized extensively in Iceland.

Hydrothermal ore deposits

Some geothermal systems act as natural distilleries in the subsurface, dissolving trace amounts of gold, silver, and other rare elements from their host rocks. These elements may then be deposited at places where changes in temperature, pressure, or composition favour precipitation. Many hydrothermal ore deposits have been formed by once active—and in a few cases still active—geothermal systems. Gold is one more legacy of volcanism.

BIBLIOGRAPHY. Because volcanoes are studied by scientists in a wide range of specialties, discussions of volcanism appear in many diverse publications. This bibliography provides only the more general sources of information on volcanism. Many of these, however, contain extensive references that can lead the reader to more specialized data.

Coverage at a nontechnical level can be found in ROBERT DECKER and BARBARA DECKER, *Volcanoes* (1981), emphasizing physical processes and the relationship of volcanoes to plate tectonics; PETER FRANCIS, *Volcanoes* (1976), concentrating on volcanic rock types; FRED M. BULLARD, *Volcanoes of the Earth*, 2nd rev. ed. (1984), describing many volcanic eruptions on a worldwide basis; and GORDON A. MACDONALD, AGATIN T. ABBOTT, and FRANK L. PETERSON, *Volcanoes in the Sea*, 2nd ed. (1983), discussing Hawaiian volcanoes and the geology of Hawaii. On a fairly high technical level are HOWELL WILLIAMS and ALEXANDER R. MCBIRNEY, *Volcanology* (1979); and GORDON A. MACDONALD, *Volcanoes* (1972), with emphasis on descriptions of eruptions and the character of volcanic rocks. R.V. FISHER and H.-U. SCHMINCKE, *Pyroclastic Rocks* (1984), is a comprehensive treatise on the origin and character of this important but diverse group of volcanic deposits. GRANT HEIKEN and KENNETH WOHLETZ, *Volcanic Ash* (1985), provides a thorough discussion of the origin and character of airfall deposits. See also T. SIMKIN *et al.*, *Volcanoes of the World: A Regional Directory, Gazetteer, and Chronology of Volcanism During the Last 10,000 Years* (1981).

Useful popular books include HAROUN TAZIEFF, *Volcanoes*, trans. from French (1961), an account of the study of volcanoes and volcanic eruptions from Greek and Roman to modern times; *Volcano*, rev. ed. (1982), a well-illustrated account of major volcanic eruptions with a chapter on monitoring active volcanoes, prepared by the editors of Time-Life Books; and MAURICE KRAFFT and KATIA KRAFFT, *Volcano* (1975; originally published in French, 1975), a book of colour photographs of volcanoes in action and a text on the mechanism of volcanoes.

General research articles about volcanism are published in *Science* (weekly) and *Nature* (weekly); and more specialized articles in *Journal of Geophysical Research: Solid Earth and Planets* (monthly); *Bulletin of Volcanology* (bimonthly); and *Journal of Volcanology and Geothermal Research* (monthly). *SEAN Bulletin* (monthly), published by the Scientific Event Alert Network of the Smithsonian Institution, provides preliminary but up-to-date data on eruptions worldwide.

Collections of articles can be found in NATIONAL RESEARCH COUNCIL. GEOPHYSICS STUDY COMMITTEE, *Explosive Volcanism: Inception, Evolution, and Hazards* (1984); ROBERT DECKER and BARBARA DECKER, *Volcanoes and the Earth's Interior: Readings from Scientific American* (1982); PETER W. LIPMAN and DONAL R. MULLINEAUX (eds.), *The 1980 Eruptions of Mount St. Helens, Washington* (1981); and ROBERT W. DECKER, THOMAS L. WRIGHT, and PETER H. STAUFFER (eds.), *Volcanism in Hawaii*, 2 vol. (1987).

(R.W.De./B.B.De.)

Voltaire

Voltaire, one of the greatest French authors, though only a few of his works are still read, is held in worldwide repute as a courageous crusader against tyranny, bigotry, and cruelty. He embodies characteristic qualities of the French mind—a critical capacity, wit, and satire. His whole work vigorously propagates an ideal of progress to which men of all nations have remained responsive. His long life spans the last years of classicism and the eve of the revolutionary era; during this age of transition his works and activities influenced the direction taken by European civilization.

Cliche Musees Nationaux, Paris

Voltaire, portrait by an unknown artist after a portrait by Nicolas de Largillière, 1718. In the Château de Versailles.

Heritage and youth. Voltaire's background was middle class. According to his birth certificate he was born François-Marie Arouet in Paris on November 21, 1694, but the hypothesis that his birth was kept secret cannot be dismissed, for he stated on several occasions that in fact it took place on February 20, 1694. He believed that he was the son of an officer named Rochebrune, who was also a songwriter. He had no love for either his putative father, François Arouet, a onetime notary who later became receiver in the Cour des Comptes (audit office), or his elder brother Armand. Almost nothing is known about his mother of whom he hardly said anything. Having lost her when he was seven, he seems to have become an early rebel against family authority. He attached himself to his godfather, the Abbé de Châteauneuf, a freethinker and epicurean who presented the boy to the famous courtesan Ninon de Lenclos when she was in her 84th year. But it is doubtless that he owed his positive outlook and his sense of reality to his bourgeois origins.

He attended the Jesuit college of Louis-le-Grand in Paris, where he learned to love literature, the theatre, and social life. While he appreciated the classical taste the college instilled in him, the religious instruction of the fathers served only to arouse his skepticism and mockery. He witnessed the last sad years of Louis XIV and was never to forget the distress and the military disasters of 1709 nor the horrors of religious persecution. He retained, however, a degree of admiration for the sovereign, and he remained convinced that the enlightened kings are the indispensable agents of progress.

He decided against the study of law after he left college. Employed as secretary at the French embassy in The Hague, he became infatuated with the daughter of an adventurer. Fearing scandal, the French ambassador sent him back to Paris. Despite his father's wishes, he wanted to devote himself wholly to literature, and he frequented the Temple, then the centre of free-thinking society. After the death of Louis XIV, under the morally relaxed Regency, Voltaire became the wit of Parisian society, and his epigrams were widely quoted. But when he dared to mock the dissolute regent, the Duc d'Orléans, he was banished from Paris and then imprisoned in the Bastille for nearly a year (1717). Behind his gay facade, he was fundamentally serious and set himself to learn the accepted literary forms. In 1718, after the success of *Oedipe,* the first of his tragedies, he was acclaimed as the successor of the great classical dramatist Jean Racine and thenceforward adopted the name of Voltaire. The origin of this pen name remains doubtful. It is not certain that it is the anagram of Arouet le jeune (*i.e.,* the younger). Above all he desired to be the Virgil that France had never known. He worked at an epic poem whose hero was Henry IV, the king beloved by the French people for having put an end to the wars of religion. This *Henriade* is spoiled by its pedantic imitation of Virgil's *Aeneid,* but his contemporaries saw only the generous ideal of tolerance that inspired the poem. These literary triumphs earned him a pension from the regent and the warm approval of the young queen, Marie. He thus began his career of court poet.

Early drama and poetry

United with other thinkers of his day—literary men and scientists—in the belief in the efficacy of reason, Voltaire was a Philosophe, as the 18th century termed it. In the salons he professed an aggressive Deism, which scandalized the devout. He became interested in England, the country that tolerated freedom of thought; he visited the Tory leader Viscount Bolingbroke, exiled in France—a politician, an orator, and a philosopher whom Voltaire admired to the point of comparing him to Cicero. On Bolingbroke's advice he learned English in order to read the philosophical works of John Locke. His intellectual development was furthered by an accident: as the result of a quarrel with a member of one of the leading French families, the Chevalier de Rohan, who had made fun of his adopted name, he was beaten up, taken to the Bastille, and then conducted to Calais on May 5, 1726, from where he set out for London. His destiny was now exile and opposition.

Exile to England. During a stay that lasted more than two years he succeeded in learning the English language; he wrote his notebooks in English and to the end of his life he was able to speak and write it fluently. He met such English men of letters as Alexander Pope, Jonathan Swift, and William Congreve, the philosopher George Berkeley, and Samuel Clarke, the theologian. He was presented at court, and he dedicated his *Henriade* to Queen Caroline. Though at first he was patronized by Bolingbroke, who had returned from exile, it appears that he quarrelled with the Tory leader and turned to Sir Robert Walpole and the liberal Whigs. He admired the liberalism of English institutions, though he was shocked by the partisan violence. He envied English intrepidity in the discussion of religious and philosophic questions and was particularly interested in the Quakers. He was convinced that it was because of their personal liberty that the English, notably Sir Isaac Newton and John Locke, were in the forefront of scientific thought. He believed that this nation of merchants and sailors owed its victories over Louis XIV to its economic advantages. He concluded that even in literature France had something to learn from England; his experience of Shakespearean theatre was overwhelming, and, however much he was shocked by the "barbarism" of the productions, he was struck by the energy of the characters and the dramatic force of the plots.

Meetings with English men of letters

Return to France. He returned to France at the end of 1728 or the beginning of 1729 and decided to present En-

gland as a model to his compatriots. His social position was consolidated. By judicious speculation he began to build up the vast fortune that guaranteed his independence. He attempted to revive tragedy by discreetly imitating Shakespeare. *Brutus,* begun in London and accompanied by a *Discours à milord Bolingbroke,* was scarcely a success in 1730; *La Mort de César* was played only in a college (1735); in *Eriphyle* (1732), the apparition of a ghost, as in *Hamlet,* was booed by the audience. *Zaïre,* however, was a resounding success. The play, in which the sultan Orosmane, deceived by an ambiguous letter, stabs his prisoner, the devoted Christian-born Zaïre, in a fit of jealousy, captivated the public with its exotic subject.

Interest in history

At the same time Voltaire had turned to a new literary genre: history. In London he had made the acquaintance of Fabrice, a former companion of the Swedish king Charles XII. The interest he felt for the extraordinary character of this great soldier impelled him to write his life, *Histoire de Charles XII* (1731), a carefully documented historical narrative that reads like a novel. Philosophic ideas began to impose themselves as he wrote: the King of Sweden's exploits brought desolation, whereas his rival Peter the Great brought Russia into being, bequeathing a vast, civilized empire. Great men are not warmongers; they further civilization—a conclusion that tallied with the example of England. It was this line of thought that Voltaire brought to fruition, after prolonged meditation, in a work of incisive brevity: the *Lettres philosophiques* (1734). These fictitious letters are primarily a demonstration of the benign effects of religious toleration. They contrast the wise Empiricist psychology of Locke with the conjectural lucubrations of René Descartes. A philosopher worthy of the name, such as Newton, disdains empty, a priori speculations; he observes the facts and reasons from them. After elucidating the English political system, its commerce, its literature, and the Shakespeare almost unknown to France, Voltaire concludes with an attack on the French mathematician and religious philosopher Pascal: the purpose of life is not to reach heaven through penitence but to assure happiness to all men by progress in the sciences and the arts, a fulfillment for which their nature is destined. This small, brilliant book is a landmark in the history of thought: not only does it embody the philosophy of the 18th century, but it also defines the essential direction of the modern mind.

Life with Mme du Châtelet. Scandal followed publication of this work that spoke out so frankly against the religious and political establishment. When a warrant of arrest was issued in May of 1734, Voltaire took refuge in the château of Mme du Châtelet at Cirey in Champagne and thus began his liaison with this young, remarkably intelligent woman. He lived with her in the château he had renovated at his own expense. This period of retreat was interrupted only by a journey to the Low Countries in December 1736—an exile of a few weeks became advisable after the circulation of a short, daringly epicurean poem called "Le Mondain."

The life these two lived together was both luxurious and studious. After *Adélaïde du Guesclin* (1734), a play about a national tragedy, he brought *Alzire* to the stage in 1736 with great success. The action of *Alzire*—in Lima, Peru, at the time of the Spanish conquest—brings out the moral superiority of a humanitarian civilization over methods of brute force. Despite the conventional portrayal of "noble savages," the tragedy kept its place in the repertory of the Comédie-Française for almost a century. Mme du Châtelet was passionately drawn to the sciences and metaphysics and influenced Voltaire's work in that direction. A "gallery" or laboratory of the physical sciences was installed at the château, and they composed a memorandum on the nature of fire for a meeting of the Académie des Sciences.

Interest in science and metaphysics

While Mme du Châtelet was learning English in order to translate Newton and *The Fable of the Bees* of Bernard de Mandeville, Voltaire popularized, in his *Éléments de la philosophie de Newton* (1738), those discoveries of English science that were familiar only to a few advanced minds in France, such as the astronomer and mathematician Pierre-Louis de Maupertuis. At the same time, he continued to pursue his historical studies. He began *Le Siècle de Louis XIV,* sketched out a universal history of kings, wars, civilization and manners that became the *Essai sur les moeurs,* and plunged into biblical exegesis. Mme du Châtelet herself wrote an *Examen,* highly critical of the two Testaments. It was at Cirey that Voltaire, rounding out his scientific knowledge, acquired the encyclopaedic culture that was one of the outstanding facets of his genius.

Because of a lawsuit, he followed Mme du Châtelet to Brussels in May 1739, and thereafter they were constantly on the move between Belgium, Cirey, and Paris. Voltaire corresponded with the crown prince of Prussia, who, rebelling against his father's rigid system of military training and education, had taken refuge in French culture. When the prince acceded to the throne as Frederick II (the Great), Voltaire visited his disciple first at Cleves (Kleve, West Germany), then at Berlin. When the War of the Austrian Succession broke out, Voltaire was sent to Berlin (1742–43) on a secret mission to rally the King of Prussia—who was proving himself a faithless ally—to the assistance of the French Army. Such services—as well as his introduction of his friends the brothers d'Argenson, who became ministers of war and foreign affairs, respectively, to the protection of Mme de Pompadour, the mistress of Louis XV—brought him into favour again at Versailles. After his poem celebrating the victory of Fontenoy (1745), he was appointed historiographer, gentleman of the king's chamber, and academician. His tragedy *Mérope,* about the mythical Greek queen, won public acclaim on the first night (1743). The performance of *Mahomet,* in which Voltaire presented the founder of Islām as an imposter, was forbidden, however, after its successful production in 1742. He amassed a vast fortune through the manipulations of Joseph Pâris Duverney, the financier in charge of military supplies, who was favoured by Mme de Pompadour. In this ambience of well-being, he began a liaison with his niece Mme Denis, a charming widow, without breaking off his relationship with Mme du Châtelet.

Yet he was not spared disappointments. Louis XV disliked him, and the pious Catholic faction at court remained acutely hostile. He was guilty of indiscretions. When Mme du Châtelet lost large sums at the Queen's gaming table, he said to her in English: "You are playing with cardsharpers"; the phrase was understood, and he was forced to go into hiding at the country mansion as the guest of the Duchesse du Maine in 1747. Ill and exhausted by his restless existence, he at last discovered the literary form that ideally fitted his lively and disillusioned temper: he wrote his first *contes* (stories). *Micromégas* (1752) measures the littleness of man in the cosmic scale; *Vision de Babouc* (1748) and *Memnon* (1749) dispute the philosophic optimism of Gottfried Wilhelm Leibniz and Alexander Pope. *Zadig* (1747) is a kind of allegorical autobiography: like Voltaire, the Babylonian sage Zadig suffers persecution, is pursued by ill fortune, and ends by doubting the tender care of Providence for human beings.

The great crisis of his life was drawing near. In 1748 at Commercy, where he had joined the court of Stanisław (the former king of Poland), he detected the love affair of Mme du Châtelet and the poet Saint-Lambert, a slightly ludicrous passion that ended tragically. On September 10, 1749, he witnessed the death in childbirth of this uncommonly intelligent woman who for 15 years had been his guide and counsellor. He returned in despair to the house in Paris where they had lived together; he rose in the night and wandered in the darkness, calling her name.

Death of Mme du Châtelet

Later travels. The failure of some of his plays aggravated his sense of defeat. He had attempted the *comédie larmoyante,* or "sentimental comedy," that was then fashionable: after *L'Enfant prodigue* (1736), a variation of the prodigal son theme, he adapted William Wycherley's satiric Restoration drama *The Plain-Dealer* to his purpose, entitling it *La Prude;* he based *Nanine* (1749) on a situation taken from Samuel Richardson's novel *Pamela,* but all without success. The court spectacles he directed gave him a taste for scenic effects, and he contrived a sumptuous decor, as well as the apparition of a ghost, for *Sémiramis* (1748), but his public was not captivated. His enemies compared him with Prosper Jolyot, sieur de Crébillon, who was pre-eminent among French writers of

tragedy at this time. Though Voltaire used the same subjects as his rival (*Oreste, Sémiramis*), the Parisian audience preferred the plays of Crébillon. Exasperated and disappointed, he yielded to the pressing invitation of Frederick II and set out for Berlin on June 28, 1750.

At the moment of his departure a new literary generation, reacting against the ideas and tastes to which he remained faithful, was coming to the fore in France. Disseminators of the philosophical ideas of the time, such as Denis Diderot, Baron d'Holbach, and their friends, were protagonists of a thoroughgoing Materialism and regarded Voltaire's Deism as too timid. Others had rediscovered with Jean-Jacques Rousseau the poetry of Christianity. All in fact preferred the charm of sentiment and passion to the enlightenment of reason. As the years passed, Voltaire became increasingly more isolated in his glory.

At first he was enchanted by his sojourn in Berlin and Potsdam, but soon difficulties arose. After a lawsuit with a moneylender, and quarrels with prominent noblemen, he started a controversy with Maupertuis (the president of Frederick's academy of science, the Berlin Academy) on scientific matters. In a pamphlet entitled "Diatribe du docteur Akakia" (1752), he covered him with ridicule. The King, enraged, consigned "Akakia" to the flames and gave its author a thorough dressing down. Voltaire left Prussia on March 26, 1753, leaving Frederick exasperated and determined to punish him. On the journey he was held under house arrest at an inn at Frankfurt, by order of the Prussian resident. Louis XV forbade him to approach Paris. Not knowing where to turn, he stayed at Colmar for more than a year. At length he found asylum at Geneva, where he purchased a house called Les Délices, at the same time securing winter quarters at Lausanne.

Major historical studies

He now completed his two major historical studies. *Le Siècle de Louis XIV* (1751), a book on the century of Louis XIV, had been prepared after an exhaustive 20-year interrogation of the survivors of *le grand siècle.* Voltaire was particularly concerned to establish the truth by collecting evidence from as many witnesses as possible, evidence that he submitted to exacting criticism. His desire was to write the nation's history by means of an examination of its arts and sciences and of its social life, but military events and politics still occupy a large place in his survey. The *Essai sur les moeurs,* the study on customs and morals that he had begun in 1740 (first complete edition, 1756), traced the course of world history since the end of the Roman Empire and gave an important place to the Eastern and Far Eastern countries. Voltaire's object was to show humanity slowly developing beyond barbarism. He supplemented these two works with one on Russian history during the reign of Peter the Great, *Histoire de l'empire de Russie sous Pierre le Grand* (1759–63), the *Philosophie de l'histoire* (1765), and the *Précis du siècle de Louis XV* (1768).

At Geneva, he had at first been welcomed and honoured as the champion of tolerance. But soon he made those around him feel uneasy. At Les Délices his presentation of plays was stopped, in accordance with the law of the republic of Geneva, which forbade both public and private theatre performances. Then there was his mock-heroic poem "La Pucelle" (1755), a most improper presentation of Joan of Arc (*La Pucelle d'Orléans),* which the booksellers printed in spite of his protests.

Attracted by his volatile intelligence, Calvinist pastors as well as women and young people thronged to his salon. Yet he soon provoked the hostility of important Swiss intellectuals. The storm broke in November 1757, when volume seven of Diderot's *Encyclopédie* was published. Voltaire had inspired the article on Geneva that his fellow philosopher Jean d'Alembert had written after a visit to Les Délices; not only was the city of Calvin asked to build a theatre within its walls but also certain of its pastors were praised for their doubts of Christ's divinity. The scandal sparked a quick response: the *Encyclopédie* was forced to interrupt publication, and Rousseau attacked the rational philosophy of the Philosophes in general in a polemical treatise on the question of the morality of theatrical performances, *Lettre à d'Alembert sur les spectacles* (1758). Rousseau's view that drama might well be abolished marked a final break between the two writers.

Voltaire no longer felt safe in Geneva, and he longed to retire from these quarrels. In 1758 he wrote what was to be his most famous work, *Candide.* In this philosophical fantasy, the youth Candide, disciple of Doctor Pangloss (himself a disciple of the philosophical optimism of Leibniz), saw and suffered such misfortune that he was unable to believe that this was "the best of all possible worlds." Having retired with his companions to the shores of the Propontis, he discovered that the secret of happiness was "to cultivate one's garden," a practical philosophy excluding excessive idealism and nebulous metaphysics. Voltaire's own garden became Ferney, a property he bought at the end of 1758, together with Tourney in France, on the Swiss border. By crossing the frontier he could thus safeguard himself against police incursion from either country.

Candide

Achievements at Ferney. At Ferney, Voltaire entered on one of the most active periods of his life. Both patriarch and lord of the manor, he developed a modern estate, sharing in the movement of agricultural reform in which the aristocracy was interested at the time. He could not be true to himself, however, without stirring up village feuds and went before the magistrates on a question of tithes, as well as about the beating of one of his workmen. He renovated the church and had *Deo erexit Voltaire* ("Voltaire erected this to God") carved on the facade. At Easter Communion, 1762, he delivered a sermon on stealing and drunkenness and repeated this sacrilegious offense in the following year, flouting the prohibition by the bishop of Annecy, in whose jurisdiction Ferney lay. He meddled in Genevan politics, taking the side of the workers (or *natifs,* those without civil rights), and installed a stocking factory and watchworks on his estate in order to help them. He called for the liberation of serfs in the Jura, but without success, though he did succeed in suppressing the customs barrier on the road between Gex in the Jura and Geneva, the natural outlet for the produce of Gex. Such generous interventions in local politics earned him enormous popularity. In 1777 he received a popular acclamation from the people of Ferney. In 1815 the Congress of Vienna halted the annexation of Ferney to Switzerland in his honour.

His fame was now worldwide. "Innkeeper of Europe"—as he was called—he welcomed such literary figures as James Boswell, Giovanni Casanova, Edward Gibbon, the Prince de Ligne, and the fashionable philosophers of Paris. He kept up an enormous correspondence—with the Philosophes, with his actresses and actors, and with those high in court circles, such as the Duc de Richelieu (grandnephew of the Cardinal de Richelieu), the Duc de Choiseul, and Mme du Barry, Louis XV's favourite. He renewed his correspondence with Frederick II and exchanged letters with Catherine II of Russia.

Scientific, political, and religious controversies

There was scarcely a subject of importance on which he did not speak. In his political ideas, he was basically a liberal, though he also admired the authority of those kings who imposed progressive measures on their people. On the question of fossils, he entered into foolhardy controversy with the famous French naturalist Comte de Buffon. On the other hand, he declared himself a partisan of the Italian scientist Abbé Lazzaro Spallanzani against the hypothesis of spontaneous generation, according to which microscopic organisms are generated spontaneously in organic substances. He busied himself with political economy and revived his interest in metaphysics by absorbing the ideas of 17th-century philosophers Benedict de Spinoza and Nicolas Malebranche.

His main interest at this time, however, was his opposition to *l'infâme,* a word he used to designate the church, especially when it was identified with intolerance. For mankind's future he envisaged a simple theism, reinforcing the civil power of the state. He believed this end was being achieved when, about 1770, the courts of Paris, Vienna, and Madrid came into conflict with the pope; but this was to misjudge the solidarity of ecclesiastical institutions and the people's loyalty to the traditional faith. Voltaire's beliefs prompted a prodigious number of polemical writings. He multiplied his personal attacks, often stooping to low cunning; in his sentimental comedy *L'Écossaise* (1760),

he mimicked the eminent critic Élie Fréron, who had attacked him in reviews, by portraying his adversary as a rascally journalist who intervenes in a quarrel between two Scottish families. He directed *Le Sentiment des Citoyens* (1764) against Rousseau. In this anonymous pamphlet, which supposedly expressed the opinion of the Genevese, Voltaire, who was well informed, revealed to the public that Rousseau had abandoned his children. As author he used all kinds of pseudonyms: Rabbi Akib, Pastor Bourn, Lord Bolingbroke, M. Mamaki "interpreter of Oriental languages to the king of England," Clocpitre, Cubstorf, Jean Plokof—a nonstop performance of puppets. As a part-time scholar he constructed a personal *Encyclopédie,* the *Dictionnaire philosophique* (1764), enlarged after 1770 by *Questions sur l'Encyclopédie.* Among the mass of writings of this period are *Le Blanc et le noir* ("The White and the Black"), a philosophical tale in which Oriental fantasy contrasts with the realism of *Jeannot et Colin; Princesse de Babylone,* a panorama of European philosophies in the fairyland of *The Thousand and One Nights;* and *Le Taureau blanc,* a biblical tale.

Again and again Voltaire returned to his chosen themes: the establishment of religious tolerance, the growth of material prosperity, respect for the rights of man by the abolition of torture and useless punishments. These principles were brought into play when he intervened in some of the notorious public scandals of these years. For instance, when the Protestant Jean Calas, a merchant of Toulouse accused of having murdered his son in order to prevent his conversion to the Roman Catholic Church, was broken on the wheel while protesting his innocence (March 10, 1762), Voltaire, livid with anger, took up the case and by his vigorous intervention obtained the vindication of the unfortunate Calas and the indemnification of the family. But he was less successful in a dramatic affair concerning the 19-year-old Chevalier de La Barre, who was beheaded for having insulted a religious procession and damaging a crucifix (July 1, 1766). Public opinion was distressed by such barbarity, but it was Voltaire who protested actively, suggesting that the Philosophes should leave French territory and settle in the town of Cleves offered them by Frederick II. Although he failed to obtain even a review of this scandalous trial, he was able to reverse other judicial errors.

Late dramatic works

By such means he retained leadership of the philosophic movement. On the other hand, as a writer, he wanted to halt a development he deplored—that which led to Romanticism. He tried to save theatrical tragedy by making concessions to a public that adored scenes of violence and exoticism. For instance, in *L'Orphelin de la Chine* (1755), Lekain (Henri-Louis Cain), who played the part of Genghis Khan, was clad in a sensational Mongol costume. Lekain, whom Voltaire considered the greatest tragedian of his time, also played the title role of *Tancrède,* which was produced with a sumptuous decor (1760) and which proved to be Voltaire's last triumph. Subsequent tragedies, arid and ill-constructed and overweighted with philosophic propaganda, were either booed off the stage or not produced at all. He became alarmed at the increasing influence of Shakespeare; when he gave a home to a grandniece of the great 17th-century classical dramatist Pierre Corneille and on her behalf published an annotated edition of the famous tragic author, he inserted, after *Cinna,* a translation of *Julius Caesar,* convinced that such a confrontation would demonstrate the superiority of the French dramatist. He was infuriated by the Shakespearean translations of Pierre Le Tourneur in 1776, which stimulated French appreciation of this more robust, nonclassical dramatist, and dispatched an abusive *Lettre à l'Académie.* He never ceased to acknowledge a degree of genius in Shakespeare, yet spoke of him as "a drunken savage." He returned to a strict classicism in his last plays, but in vain, for the audacities of his own previous tragedies, timid as they were, had paved the way for Romantic drama.

It was the theatre that brought him back to Paris in 1778. Wishing to direct the rehearsals of *Irène,* he made his triumphal return to the city he had not seen for 28 years on February 10. More than 300 persons called on him the day after his arrival. On March 30 he went to the Académie amid acclamations, and, when *Irène* was played before a delirious audience, he was crowned in his box. His health was profoundly impaired by all this excitement. On May 18 he was stricken with uremia. He suffered much pain on his deathbed, about which absurd legends were quickly fabricated; on May 30 he died, peacefully it seems. His nephew, the Abbé Mignot, had his body, clothed just as it was, swiftly transported to the Abbey of Scellières, where he was given Christian burial by the local clergy; the prohibition of such burial arrived after the ceremony. His remains were transferred to the Panthéon during the Revolution in July 1791.

Assessment. Voltaire's name has always evoked vivid reactions. Toward the end of his life he was attacked by the followers of Rousseau, and after 1800 he was held responsible for the Revolution. But the excesses of clerical reactionaries under the Restoration and the Second Empire rallied the middle and working classes to his memory. At the end of the 19th century, though conservative critics remained hostile, scientific research into his life and works was given impetus by Gustave Lanson. Voltaire himself did not hope that all his vast quantity of writings would be remembered by posterity. His epic poems and lyrical verse are virtually dead, as are his plays. But his *contes* are continually republished, and his letters are regarded as one of the great monuments of French literature. He bequeathed a lesson to humanity, which has lost nothing of its value. He taught men to think clearly; his was a mind at once precise and generous. "He is the necessary philosopher," wrote Lanson, "in a world of bureaucrats, engineers, and producers." (R.H.Po.)

MAJOR WORKS

STORIES: *Les Voyages du baron de Gangan* (1739; an earlier version of *Micromégas,* 1752); *Zadig* (1747; Eng. trans. by J. Butt, 1964); *Vision de Babouc* (1748); *Candide* (1759; Eng. trans. by J. Spencer, 1966, and by Robert M. Adams, 1966); *Le Blanc et le noir* (1764; *The Two Genies,* 1895); *Jeannot et Colin* (1764); *La Princesse de Babylone* (1768; Eng. trans. 1969); *Le Taureau blanc* (1774; Eng. trans. 1774—attributed to Jeremy Bentham).

HISTORY: *Histoire de Charles XII* (1731; Eng. trans. by C. Messiter, 1860); *Le Siècle de Louis XIV* (1751; Eng. trans. by J.H. Brumfitt, 1966); *Histoire de l'empire de Russie sous Pierre le Grand* (1759–63; Eng. trans. 1763); *Précis du siècle de Louis XV* (1768; Eng. trans. 1774, 2 vol.).

PHILOSOPHY: *Lettres philosophiques* (1734); *Éléments de la philosophie de Newton* (1738; Eng. trans. 1738); *Essai sur les moeurs* (1756; Eng. trans. by Nugent, 1759); *Dictionnaire philosophique* (1764; Eng. trans. by H.I. Woolf, 1945); *Philosophie de l'histoire* (1765; Eng. trans. 1965).

PLAYS: *Zaïre* (1732; Eng. trans. by A. Wallace, 1854); *Alzire* (1736; *Tragedy of Alzira,* trans. by A. Hill, 1817); *Mérope* (1743; Eng. trans. by A. Hill, 1803).

OTHER WORKS: *La Henriade* (1728; Eng. trans. by R.G. Macgregor, 1854), an epic poem.

EDITIONS: *Les Oeuvres complètes* is an edition begun in 1968, directed first by THEODORE BESTERMAN, then by GILES BARBER, with introductions and notes in English and French. Among the works published are 51 volumes of Voltaire's complete correspondence and related documents. Theodore Besterman also translated and edited selections of Voltaire's correspondence, such as *The Love Letters of Voltaire to His Niece* (1958). See also THEODORE BESTERMAN and ANDREW BROWN (eds.), *Concordance to the Correspondence of Voltaire* (1977).

BIBLIOGRAPHY. A summary of scholarship on Voltaire is presented in MARY MARGARET H. BARR, *Century of Voltaire Study: A Bibliography of Writings on Voltaire, 1825–1925* (1929, reissued 1972), continued in her *Quarante Années d'études voltairiennes: bibliographie analytique des livres et articles sur Voltaire, 1926–1965* (1968). Since 1955 the Institut et Musée Voltaire, in Geneva, and the Voltaire Foundation, at the Taylor Institution in Oxford, Eng., have been publishing the multivolume series "Studies on Voltaire and the Eighteenth Century"; several works appear in the series each year, and many of the titles cited below belong to it.

Life: GUSTAVE DESNOIRESTERRES, *Voltaire et la société au XVIII siècle,* 2nd ed., 8 vol. (1871–76, reprinted 1967), is a narrative work that was the basis for most later Voltaire biographies but is outdated now because of the numerous documents discovered since its original publication; THEODORE BESTERMAN, *Voltaire,* 3rd rev. ed. (1976), offers a thorough, informed life story that is characterized by the author's full agreement with the views of Voltaire himself; IRA O. WADE, *The*

Intellectual Development of Voltaire (1969), is another broad-ranging, scholarly biography; A. OWEN ALDRIDGE, *Voltaire and the Century of Light* (1975), examines both the life and works; JOHN E.N. HEARSEY, *Voltaire* (1976), provides a highly readable introductory survey of important events of Voltaire's life; JEAN ORIEUX, *Voltaire* (1979; originally published in French, 1966), is a detailed biographical account by a French historian, written in the tradition of Desnoiresterres; PEYTON RICHTER and ILONA RICARDO, *Voltaire* (1980), is a short work that introduces the general reader to Voltaire; and HAYDN MASON, *Voltaire: A Biography* (1981), is a scholarly study that places particular emphasis on the philosopher's later years and serves as a companion to Mason's critical survey of Voltaire's work, cited below.

WAYNE ANDREWS, *Voltaire* (1981), provides a well-researched biography with a discussion of Voltaire's social and literary contacts; and JEAN MOHSEN FAHMY, *Voltaire et Paris* (1981), explores his Parisian life at different periods. Social and literary contacts and associates are studied in PAUL O. LECLERC, *Voltaire and Crébillon Père: History of an Enmity* (1973); and CHRISTIANE MERVAUD, *Voltaire et Frédéric II: une dramaturgie des lumières, 1736–1778* (1985). *Voltaire and the English* (1979) is a collection of scholarly articles, with an illustrated catalog of a documents exhibition on Voltaire's journeys and the intellectual influences stemming from them. English literary and social influences are also studied in AHMAD GUNNY, *Voltaire and English Literature* (1979). Russian influences and Voltaire's place in Russian historiography are explored in CAROLYN H. WILBERGER, *Voltaire's Russia: Window on the East* (1976). Voltaire's temperament and personality are studied in NORMAN L. TORREY, *The Spirit of Voltaire* (1938, reprinted 1968); and R.S. RIDGWAY, *Voltaire and Sensibility* (1973). NANCY MITFORD, *Voltaire in Love* (1957, reissued 1985); and SAMUEL EDWARDS, *The Divine Mistress* (1970), are substantial histories of Voltaire's involvement with Mme du Châtelet, who greatly influenced both his work and his life.

Views: RENÉ POMEAU, *La Religion de Voltaire,* rev. ed. (1969, reissued 1974), examines the religious beliefs of Voltaire within the intellectual background of the 18th century, concluding that he was indeed a deist; other studies of Voltaire's views on religion and religious matters include JOHN RENWICK, *Marmontel, Voltaire, and the Bélisaire Affair* (1974); RICHARD A. BROOKS, *Voltaire and Leibniz* (1964); WILLIAM H. TRAPNELL, *Christ and His "Associates" in Voltairian Polemic: An Assault on the Trinity and the Two Natures* (1982), and *Voltaire and the Eucharist* (1981); R.E. FLORIDA, *Voltaire and the Socinians* (1974); GRAHAM GARGETT, *Voltaire and Protestantism* (1980); DAVID D. BIEN, *The Calas Affair: Persecution, Toleration, and Heresy in Eighteenth-Century Toulouse* (1960, reprinted 1979), on Voltaire's intervention in a case of religious persecution; and BERTRAM EUGENE SCHWARZBACH, *Voltaire's Old Testament Criticism* (1971).

For political and social views, see HENRY MEYER, *Voltaire on War and Peace* (1976); and PETER GAY, *Voltaire's Politics: The Poet as Realist,* 2nd ed. (1988), dealing with Voltaire's politics in relation to England, France, Prussia, Russia, and Geneva and relating his ideas for social reform directly to events in 18th-century Europe. A.J. AYER, *Voltaire* (1986), assesses him as a historian and dramatist, with a focus on his religious and philosophical arguments. PATRICK HENRY, *Voltaire and Camus: The Limits of Reason and the Awareness of Absurdity* (1975), is a study of Voltaire's interpretation of the philosophical categories of rationalism and absurdity.

Critical studies: VIRGIL W. TOPAZIO, *Voltaire: A Critical Study of His Major Works* (1967), analyzes the thought and work of Voltaire the storyteller, historian, playwright, and philosopher. J.H. BRUMFITT, *Voltaire, Historian* (1958, reprinted 1985), studies the principal features of his historical works and the theories behind them. Other introductions to Voltaire's criticism, analyzing selections of representative works without detailed exploration of individual ones, include THEODORE BESTERMAN, *Voltaire Essays, and Another* (1962, reprinted 1980); WILLIAM F. BOTTIGLIA (ed.), *Voltaire: A Collection of Critical Essays* (1968); R.H. HOWELLS *et al.* (eds.), *Voltaire and His World* (1985); HAYDN MASON, *Voltaire* (1975); and JEAN SAREIL (ed.), *Voltaire et la critique* (1966), a collection of essays, both in English and in French, devoted to Voltaire as poet, historian, philosopher, storyteller, and propagandist. For separate analyses, in English, of major works, see DENNIS FLETCHER, *Voltaire, Lettres Philosophiques* (1986), a brief critical guide; and GEOFFREY MURRAY, *Voltaire's Candide: The Protean Gardener, 1755–1762* (1970).

Voltaire as a playwright is studied in JACK ROCHFORD VROOMAN, *Voltaire's Theatre: The Cycle from Œdipe to Mérope* (1970); ROBERT L. HERBERT, *David, Voltaire, "Brutus," and the French Revolution* (1972), an exploration of the political atmosphere at the time of the 1790 revival of Voltaire's play *Brutus* and its role in the French Revolution; LILIAN WILLENS, *Voltaire's Comic Theatre: Composition, Conflict, and Critics* (1975); and MARIE WELLINGTON, *The Art of Voltaire's Theatre: An Exploration of Possibility* (1987). Fiction and poetry are discussed in DOUGLAS A. BONNEVILLE, *Voltaire and the Form of the Novel* (1976); JACQUES VAN DEN HEUVEL, *Voltaire dans ses contes: De "Micromégas" à "L'Ingénu,"* 3rd ed. (1982), a new interpretation of the stories, showing how Voltaire both revealed and resolved his inner conflicts in these stories; CAROL SHERMAN, *Reading Voltaire's Contes: A Semiotics of Philosophical Narration* (1985); and RALPH ARTHUR NABLOW, *A Study of Voltaire's Lighter Verse* (1974). (R.H.Po./Ed.)

The Technology of War

Technology may be defined as the application of knowledge to obtain a physical effect by means of an object. It may also be defined as the object itself and as the knowledge required to design, produce, and employ the object. The technology of war encompasses the entire range of weapons, equipment, structures, and vehicles with which man has armed himself specifically for the purpose of fighting his fellow man. It also includes the knowledge required to construct them, to employ them in combat, and to repair and replenish them.

The technology of war may be divided into five categories. Offensive arms harm the enemy, while defensive weapons ward off offensive blows. Transportation technology moves soldiers and weaponry; communications coordinate the movements of armed forces; and sensors detect forces and guide weaponry.

From the earliest times, a critical relationship has existed between military technology, the tactics of its employment, and the psychological factors that bind its users into units. Success in combat, the sine qua non of military organizations and the ultimate purpose of military technology, depends on the ability of the combatant group to coordinate the actions of its members in a tactically effective manner. This coordination is a function of the strength of the forces that bind the unit together, inducing its members to set aside their individual interests—even life itself—for the welfare of the group. These forces, in turn, are directly affected both by tactics and by technology.

The influence of technology can be either positive or negative. The experience of the ancient Greek hoplite infantrymen is one example of positive influence. Their arms and armour were most effective for fighting in close formation, which led in turn to marching in step, which further augmented cohesion and made the phalanx a tactically formidable formation. The late medieval knight offers an example of the negative influence of technology. To wield his sword and lance effectively, he and his charger needed considerable space, yet his closed helmet made communication with his fellows extremely difficult. It is not surprising, then, that knights of the late Middle Ages tended to fight as individuals and were often defeated by cohesive units of less well-equipped opponents.

This article is about the physical artifacts of war, the objects used in combat and in supporting the application of military force. The article is divided into two parts. The first traces the development of military technology by historical period, from prehistory to the 18th century. The second part discusses individual weapons or weapon systems as they have been developed since the 18th century.

A full treatment of the actual waging of war—including discussions of strategy, tactics, and logistics—is found in the article WAR, THE THEORY AND CONDUCT OF. The social sciences of war, such as economics, law, and the theory of its origins, are also covered in that article. For a military history of World Wars I and II, see WORLD WARS. Accounts of other wars can be found in articles on individual countries, regions, and historical eras. For example, for the American Civil War, see UNITED STATES OF AMERICA; for the Punic Wars, see GREEK AND ROMAN CIVILIZATION, ANCIENT.

Warfare requires the use of technologies that also have nonmilitary applications. For descriptions of the propulsion systems used in military vehicles, ships, aircraft, and missiles, see ENERGY CONVERSION; for the manufacture of explosives, see INDUSTRIES, CHEMICAL PROCESS. The principles of radar, and its military applications, are covered in RADAR. For the principles of aircraft flight, see TRANSPORTATION.

For coverage of related topics in the *Macropædia* and *Micropædia*, see the *Propædia*, section 736, and the *Index*.

(J.F.G.)

This article contains the following sections:

Military technology before the modern era 530
- Prehistory 530
  - The earliest military weapons
  - From precious metals to base metals
- Antiquity and the classical age, *c.* 1000 BC–AD 400 531
  - Defensive weaponry
  - Offensive weaponry
  - Mechanical artillery
  - Fortification
  - Land transportation
- The age of cavalry, *c.* AD 400–1350 535
  - The war-horse
  - Personal armour
  - The castle
  - Siege weapons
  - The horse archer
- The infantry revolution, *c.* 1200–1500 539
  - The crossbow
  - The English longbow
  - Halberd and pike
- The gunpowder revolution, *c.* 1300–1650 541
  - Early gunpowder
  - The development of artillery
  - The first small arms
  - Fortification

Modern weapons and weapon systems 547
- Small arms 547
  - Shoulder weapons
  - Machine guns
  - Pistols
  - Grenade launchers
  - Antitank weapons
- Artillery 555
  - Cannons
  - Mortars
  - Antiaircraft artillery
  - Antitank guns
  - Recoilless guns
- Rockets and missile systems 560
  - Military rockets
  - Tactical guided missiles
  - Strategic missiles
- Nuclear weapons 575
  - The development of fission weapons
  - The development of fusion weapons
  - Refinements in design
- Chemical and biological weapons 580
  - Chemical warfare
  - Biological warfare
- Fortification 581
  - Trench warfare, 1860–1918
  - Permanent fortification, 1914–45
  - Nuclear fortification
- Tanks and armoured vehicles 583
  - Tanks
  - Amphibious vehicles
  - Self-propelled guns
  - Armoured personnel carriers
- Naval ships and craft 589
  - The age of oar and ram
  - The age of gun and sail
  - The age of steam and iron
  - The age of big gun and torpedo
  - The age of the aircraft carrier
  - The age of the guided missile
- Submarines 602
  - Early hand-powered submersibles
  - Toward diesel-electric power
  - Nuclear propulsion
- Military aircraft 608
  - Early history
  - World War I

Interwar developments
World War II
The jet age
Warning and detection systems 618
History
Electromagnetic sensors
Warning systems
Military communication 622
Early development
The advent of electrical signaling
From World War I to 1940
World War II and after
Bibliography 625

MILITARY TECHNOLOGY BEFORE THE MODERN ERA

In the remote past, the diffusion of military technology was gradual and uneven. There were several reasons for this. First, transport was slow and its capacity small. Second, the technology of agriculture was no more advanced than that of war, so that, with most of their energy devoted to feeding themselves and with little economic surplus, people had few resources available for specialized military technology. Low economic development meant that even the benefits of conquest would not pay off a heavy investment in weaponry. Third, and most important, the absolute level of technological development was low. A heavy dependence on human muscle was the principal cause and a major effect of this low level of development. With human ingenuity bound by the constraints of the human body, both technology and tactics were heavily shaped by geography, climate, and topography.

The emergence of military ecospheres

The importance of geographic and topographic factors, along with limited means of communication and transportation, meant that separate geographic regions tended to develop unique military technologies. Such areas are called military ecospheres. The boundaries of a military ecosphere might be physical barriers, such as oceans or mountain ranges; they might also be changes in the military topography, that combination of terrain, vegetation, and man-made features that could render a particular technology or tactic effective or ineffective.

Until the late 15th century AD, when advances in transportation technology broke down the barriers between them, the world contained a number of military ecospheres. The most clearly defined of these were based in Mesoamerica, Japan, India–Southeast Asia, China, and Europe. (In this context, Europe includes all of the Mediterranean basin and the watershed of the Tigris and Euphrates rivers.) With the appearance of the horse archer in late antiquity, the Eurasian Steppe became a well-defined military ecosphere as well.

Those ecospheres with the most enduring impact on the technology of war were the European and Chinese. Though Japan possessed a distinctive, coherent, and effective military technology, it had little influence on developments elsewhere. India–Southeast Asia and Mesoamerica developed technologies that were well adapted to local conditions, but they were not particularly advanced. The Eurasian Steppe was a special case: usually serving as an avenue for a limited exchange of knowledge between Europe and China, in the late classical and medieval eras of Europe it developed an indigenous military technology based on the horse and composite recurved bow that challenged Europe and ultimately conquered China.

Improved methods of transportation and warfare led to the eventual disappearance of the regional ecospheres and their absorption into the European ecosphere. This process began in the 12th century with the Mongol conquest of China and invasions of Europe, and it quickened and assumed a more pronounced European flavour in the 15th and 16th centuries with the development of oceangoing ships armed with gunpowder weapons.

Because European methods of warfare ultimately dominated the world, and because the technology of war, with few exceptions, advanced first and fastest in Europe, this article devotes most of its attention to the European military ecosphere. This first part traces the technology of land war in that ecosphere from Stone Age weapons to the early guns. For reasons of continuity, warships from before the gunpowder era are discussed with modern naval ships and craft in the second part, *Modern weapons and weapon systems.*

Prehistory

THE EARLIEST MILITARY WEAPONS

The earliest evidence for a specialized technology of war dates from the period before knowledge of metalworking had been acquired. The stone walls of Jericho, which date from about 8000 BC, represent the first technology that can be ascribed unequivocally to purely military purposes. These walls, at least 13 feet (four metres) in height and backed by a watchtower or redoubt some 28 feet tall, were clearly intended to protect the settlement and its water supply from human intruders.

When the defenses of Jericho were built, humans had used sharpened stone heads for axes, spears, and arrows for thousands of years and had no doubt mastered the use of wood for clubs, ax handles, and spear shafts. Flint daggers, first of chipped and later of polished stone, were also an established technology. Missile weapons included the simple bow, the javelin, the spear thrower, and the sling. All of these hunting tools had serious military potential, but the first known implements designed purposely as offensive weapons were maces dating from the Chalcolithic period, or early Bronze Age. The mace was a simple rock, shaped for the hand and intended to smash bone and flesh, to which a handle had been added to increase the velocity and force of the blow.

The stone mace

It is evident that the technical problems of hafting a stone onto a handle were not easily solved. Well-made maces were for a long time few in number and were, by and large, wielded only by champions and rulers. The earliest known inscription identifying a historical personage by name is on the palette of King Narmer, a small, low-relief slate sculpture dating from about 3100 BC. The palette depicts Menes, the first pharaoh of a unified Egypt, ritually smashing the forehead of an enemy with a mace.

The advent of the mace as a purposely designed offensive weapon opened the door to the conscious innovation of specialized military technology. By the middle of the 3rd millennium BC, mace heads were being cast of copper, first in Mesopotamia and then in Syria, Palestine, and Egypt. The copper mace head, yielding higher density and greater crushing power, represents one of the earliest significant uses of metal for other than ornamental purposes.

FROM PRECIOUS METALS TO BASE METALS

The dividing line between the utilitarian and the symbolic in warfare has never been clear and unequivocal, and this line is particularly difficult to find in the design and construction of early weaponry. The engineering principles that dictated functional effectiveness were not understood in any systematic fashion, yet the psychological reality of victory or defeat was starkly evident. The result was an "unscientific" approach to warfare and technology, in which materials appear to have been applied to military purposes as much for their presumed mystical or magical properties as for their functional worth.

Symbolism and utility

This overlapping of symbolism and usefulness is most evident in the smith's choice of materials. Ornaments and ceremonial artifacts aside, metalworking was applied to the production of weaponry as early as, or earlier than, any other economically significant pursuit. Precious metals, with their low melting points and great malleability, were worked first; next came copper—at first pure, then alloyed with arsenic or tin to produce bronze—and then iron. A remarkable phenomenon was the persistence of weaponry made of the soft, rare metals such as gold, silver, and electrum (a naturally occurring alloy of gold and

silver) long after mechanically superior materials had become available. Although they were functionally inferior to bronze or copper, precious metals were widely valued for their mystical or symbolic importance, and smiths continued to make weapons of them long after they had mastered the working of functionally superior base metals. Some of these weapons were plainly ceremonial, but in other cases they appear to have been functional. For example, helmets and body armour of electrum, which were probably intended for actual use, have been found in Egyptian and Mesopotamian burials dating from the 2nd and 3rd millennia BC.

Antiquity and the classical age, c. 1000 BC–AD 400

From the appearance of iron weaponry in quantity during late antiquity until the fall of Rome, the means with which war was waged and the manner in which it was conducted displayed many enduring characteristics that gave the period surprising unity. Prominent features of that unity were a continuity in the design of individual weaponry, a relative lack of change in transportation technology, and an enduring tactical dominance of heavy infantry.

Reliance on human muscle

Perhaps the strongest underlying technological feature of the period was the heavy reliance on human muscle, which retained a tactical primacy that contrasted starkly with medieval times, when the application of horse power became a prime ingredient of victory. (There were two major, if partial, exceptions to this prevailing feature: the success of horse archers in the great Eurasian Steppe during late classical times, and the decisive use in the 4th century BC of shock cavalry by the armies of Philip II of Macedon and his son Alexander the Great. However, the defeat of Roman legions by Parthian horse archers at Carrhae in western Mesopotamia in 53 BC marked merely a shifting of boundaries between ecospheres on topographical grounds rather than any fundamental change within the core of the European ecosphere itself. Also, the shock cavalry of Philip and Alexander was an exception so rare as to prove the rule; moreover, their decisiveness was made possible by the power of the Macedonian infantry phalanx.) Heavy infantry remained the dominant European military institution until it was overthrown in the 4th century AD by a system of war in which shock cavalry played the central role.

Classical technologists never developed an efficient means of applying animal traction to haulage on land, no doubt because agricultural resources in even the most advanced areas were incapable of supporting meaningful numbers of horses powerful enough to make the effort worthwhile. Carts were heavy and easily broken, and the throat-and-girth harness for horses, mules, and donkeys put pressure on the animals' windpipes and neck veins, severely restricting the amount they could pull. The yoke-and-pole harness for oxen was relatively efficient and oxen could pull heavy loads, but they were extremely slow. A human porter, on the other hand, was just as efficient as a pack horse in weight carried per unit of food consumed. The best recipe for mobility, therefore, was to restrict pack animals to the minimum needed for carrying bulky items such as essential rations, tents, and firewood, to use carts only for items such as siege engines that could be carried in no other way, and to require soldiers to carry all their personal equipment and some of their food.

On the other hand, mastery of wood and bronze for military purposes reached a level during this period that was seldom, if ever, attained afterward. Surviving patterns for the Roman military boot, the caliga, suggest equally high standards of craftsmanship in leatherworking, and the standards of carpentry displayed on classical ships were almost impossibly high when measured against those of later eras.

DEFENSIVE WEAPONRY

The design and production of individual defensive equipment was restricted by the shape of the human form that it had to protect; at the same time, it placed heavy demands on the smith's skills. The large areas to be protected, restrictions on the weight that a combatant could carry, the difficulty of forging metal into the complex contours required, and cost all conspired to force constant change.

Competition between offensive and defensive weaponry

The technology of defensive weapons was rarely static. Evidence exists of an ancient contest between offensive and defensive weaponry, with defensive weaponry at first leading the way. By 3000 BC Mesopotamian smiths had learned to craft helmets of copper-and-arsenic bronze, which, no doubt worn with a well-padded leather lining, largely neutralized the offensive advantages of the mace. By 2500 BC the Sumerians were making helmets of bronze, along with bronze spearheads and ax blades. The weapon smiths' initial response to the helmet was to augment the crushing power of the mace by casting the head in an ellipsoidal form that concentrated more force at the point of impact. Then, as technical competence increased, the ellipsoidal head became a cutting edge, and by this process the mace evolved into the ax. The contest between mace and helmet initiated a contest between offensive and defensive technology that continued throughout history.

Helmets. The helmet, though arguably the earliest focus of the armourer's craft, was one of the most demanding challenges. Forging an integral, one-piece dome of metal capable of covering the entire head was extremely difficult. The Corinthian Greek helmet, a deep, bowl-shaped helmet of carefully graduated thickness forged from a single piece of bronze, probably represented the functional as well as aesthetic apex of the bronze worker's art. Many classical Greek helmets of bronze were joined by a seam down the crown.

Iron helmets followed the evolution of iron mail, itself a sophisticated and relatively late development. The legionnaire of the early Roman Republic wore a helmet of bronze, while his successor in the Empire of the 1st century AD wore one of iron.

Shields. Shields were used for hunting long before they were used for warfare, partly for defense and partly for concealment in stalking game, and it is likely that the military shield evolved from that of the hunter and herdsman. The size and composition of shields varied greatly, depending on the tactical demands of the user. In general, the more effective the protection afforded by body armour, the smaller the shield; similarly, the longer the reach of the soldier's weapon, the smaller his shield. The Greek hoplite, a heavy infantryman who fought in closely packed formation, acquired his name from the *hoplon,* a convex, circular shield, approximately three feet (90 centimetres) in diameter, made of composite wood and bronze (see Figure 1). It was carried on the left arm by means of a bronze strap that passed across the forearm and a rope looped around the inner rim with sufficient slack to be gripped in the fist. In the 4th century BC the soldier of the Roman Republic, who fought primarily with the spear, carried an oval shield, while the later imperial legionnaire, who closed in with a short sword, protected himself with the *scutum,* a large cylindrical shield of leather-clad wood that covered most of his body (see Figure 2).

The Greek *hoplon*

Body armour. Padded garments, and perhaps armour of hardened leather, preceded edged metal weapons. It was then a logical, if expensive, step to cast or forge small metal plates and sew them onto a protective garment. These provided real protection against arrow, spear, or mace, and the small scales, perforated for attachment, were a far less demanding technical challenge than even the simplest helmet. Armour of overlapping scales of bronze, laced together or sewn onto a backing of padded fabric, is well represented in pictorial evidence and burial items from Mesopotamia, Palestine, and Egypt from about 1500 BC, though its use was probably restricted to a small elite.

Bronze. By classical times, breastplates of bronze, at first beaten and then cast to the warrior's individual shape, were commonplace among heavy infantry and elite cavalry. Greaves, defenses for the lower leg, closely followed the breastplate. At first these were forged of bronze plates; some classical Greek examples were cast to such fine tolerances that they sprang open and could be snapped onto the calf. Defenses for more remote portions of the body, such as vambraces for the forearm and defenses for the ankle resembling spats, were included in Greek

Figure 1: *Greek hoplite of the 6th century BC.* His bronze shield, called the *hoplon,* protects him from knee to neck. Bronze greaves, breastplate, and helmet protect his shins, trunk, and head. Detail from a bronze krater from Vix, Fr., 6th century BC. In the Musée Archéologique, Châtillon-sur-Seine, Fr.

By courtesy of the Musee Archeologique, Chatillon-sur-Seine, France

temple dedications, but they were probably not common in field service.

Bronze was the most common metal for body defenses well into the Iron Age, a consequence of the fact that it could be worked in large pieces without extended hand forging and careful tempering, while iron had to be forged from relatively small billets.

Mail. The first practical body armour of iron was mail, which made its appearance in Hellenistic times but became common only during the Roman Imperial period.

Deutsches Archaologisches Institut, Rome

Figure 2: *Roman legionnaires of the 2nd century AD.* The soldiers at top wear a segmented iron breastplate called the *lorica segmentata,* carry a leather-covered shield called the *scutum,* and are armed with a short iron sword called the *gladius.* The soldiers at bottom put up wooden fortification—a task at which the Romans excelled. All wear iron helmets. Detail of Trajan's Column, Rome, AD 106–113.

(Bronze mail was impractical because of the insufficient strength of the alloy.) Mail, or chain mail, was made of small rings of iron, typically of one-half-inch diameter or less, linked into a protective fabric. The rings were fastened together in patterns of varying complexity depending on the degree of protection desired; in general, smaller, lighter rings fastened in dense, overlapping patterns meant lighter, better protection. The fabrication of mail was extremely labour-intensive. The earliest mail was made of hand-forged links, each individual link riveted together. Later, armourers used punches of hardened iron to cut rings from sheets: this reduced the labour involved and, hence, the cost.

The earliest evidence of mail is depicted on Greek sculpture and friezes dating from the 3rd century BC, though this kind of protection might be considerably older (there was some evidence that it might be of Celtic origin). Little else is known about the use of mail by the Greeks, but the Roman legionnaire was equipped with a *lorica hamata,* a mail shirt, from a very early date. Mail was extremely flexible and provided good protection against cutting and piercing weapons. Its main disadvantage was its weight, which tended to hang from the shoulders and waist. In addition, strips of mail tended to curl at the edges; the Romans solved this problem by lacing mail shoulder defenses to leather plates. In the 1st century AD the legionnaire's mail shirt gave way to a segmented iron torso defense, the *lorica segmentata.*

The Roman *lorica segmentata*

Plate-iron armour. While some early forged bronze armour was technically plate, the introduction of the *lorica segmentata* heralded the production of practical plate armour on a large scale. In general, the term plate would imply a uniform thickness of metal, and only iron could provide reasonably effective protection with uniform thickness without excessive weight.

While the Republican legionnaire's *lorica hamata* hung to the midthigh, his imperial successor's *lorica segmentata* covered only the shoulders and torso. On the whole, classical plate armour probably provided better protection against smashing and heavy piercing blows, while a shirt of well-made mail covered more of the body and, hence, afforded better protection against slashing blows and missiles.

OFFENSIVE WEAPONRY

Development of the offensive technology of war was not as constrained by technological and economic limitations as was defensive weaponry. Every significant offensive weapon was widely available, while defensive equipment of high quality was almost always confined to the elite. Perhaps as a consequence, a wide variety of individual offensive weapons appeared in antiquity. One of the most striking facets of ancient military technology is the early date by which individual weapons attained their form and the longevity of early offensive weapons concepts. Some of the weapons of antiquity disappeared as practical military implements in classical and medieval times, and all underwent modification, but, with the exception of the halberd and crossbow, virtually every significant pre-gunpowder weapon was known in antiquity.

The ax. Limitations on the strength of bronze and difficulties in casting and hafting restricted the ax at first to a relatively broad blade mortised into a handle at three points and secured with bindings or rivets. The hafting problem became acute as improvements in armour dictated longer, narrower blades designed primarily for piercing rather than cutting. This led to the development of socketed axes, in which the handle passed through a tubular hole cast in the ax head; both hole and head were tapered from front to rear to prevent the head from flying off. This far stronger hafting technique must have been accompanied by a significant improvement in the quality of the metal itself. The pace and timing of these developments varied enormously from place to place, depending on the local level of technology. Sumerian smiths were casting socketed ax heads with narrow piercing blades by 2500 BC, while simple mortise-and-tenon hafting was still being used in Egypt 1,000 years later.

Socketed ax heads

The spear. Though early man probably employed spears

of fire-hardened wood, spearheads of knapped stone were used long before the emergence of any distinction between hunting and military weapons. Bronze spearheads closely followed the development of alloys hard enough to keep a cutting edge and represented, with the piercing ax, the earliest significant military application of bronze. Spearheads were also among the earliest militarily significant applications of iron, no doubt because existing patterns could be directly extrapolated from bronze to iron. Though the hafting is quite different, bronze Sumerian spearheads of the 3rd millennium BC differ only marginally in shape from the leaf-shaped spearheads of classical Greece.

The spears of antiquity were relatively short, commonly less than the height of the warrior, and typically were wielded with one hand. As defensive armour and other weapons of shock combat (notably the sword) improved, spear shafts were made longer and the use of the spear became more specialized. The Greek hoplite's spear was about nine feet long; the Macedonian *sarissa* was twice that length in the period of Alexander's conquests and it grew to some 21 feet in Hellenistic times.

The javelin. Javelins, or throwing spears, were shorter and lighter than spears designed for shock combat and had smaller heads. The distinction between javelin and spear was slow to develop, but by classical times the heavy spear was clearly distinguished from the javelin, and specialized javelin troops were commonly used for skirmishing. A throwing string was sometimes looped around the shaft and tied to the thrower's finger to impart spin to the javelin on release. This improved the weapon's accuracy and probably increased the range and penetrating power by permitting a harder cast.

A significant refinement of the javelin was the Roman pilum. The pilum was relatively short, about five feet long, and had a heavy head of soft iron that made up nearly one-third of the weapon's total length. The weight of this weapon restricted its range but gave it greater impact. Its head of soft iron was intended to bend on impact, preventing an enemy from throwing it back.

Like the spear, the javelin was relatively unaffected by the appearance of iron and retained its characteristic form until it was finally abandoned as a serious weapon in the 16th century.

The sling. The sling was the simplest of the missile weapons of antiquity in principle and the most difficult in practice. It consisted of two cords or thongs fastened to a pouch. A small stone was placed in the pouch, and the slinger whirled the whole affair around to build up velocity before letting go of one of the cord ends to release the projectile. While considerable velocity could be imparted to a projectile in this way, the geometry of the scheme dictated that the release be timed with uncanny precision to achieve even rudimentary accuracy. Almost always wielded by tribal or regionally recruited specialists who acquired their skills in youth, the sling featured prominently in warfare in antiquity and classical times. It outranged the javelin and even—at least at some times and places—the bow (a point confirmed in the 4th century BC by the Greek historian Xenophon). By classical times, lead bullets, often with slogans or epigrams cast into them—"A nasty present!"—were used as projectiles.

The sling vanished as a weapon of war in the Old World by the end of the classical period, owing mainly to the disappearance of the tribal cultures in which it originated. (In the New World, on the other hand, both the Aztecs and Incas used the sling with great effect against Spanish conquistadores in the 16th century.)

The sword. The advantages of a long, sharp blade had to await advanced smelting and casting technology before they could be realized. By about 1500 BC the cutting ax had evolved into the sickle sword, a bronze sword with a curved, concave blade and a straight, thickened handle. Bronze swords with straight blades more than three feet long have been found in Greek grave sites; however, because this length exceeded the structural capabilities of bronze, these swords were not practical weapons. As a serious military implement, the sword had to await the development of iron forging, and the first true swords date from about 1200 BC.

Swords in antiquity and classical times tended to be relatively short, at first because they were made of bronze and later because they were rarely called upon to penetrate iron armour. The blade of the classic Roman stabbing sword, the *gladius,* was only some two feet long, though in the twilight years of the empire the *gladius* gave way to the spatha, the long slashing sword of the barbarians.

The Roman *gladius*

The bow. The bow was simple in concept, yet it represented an extremely sophisticated technology. In its most basic form, the bow consisted of a stave of wood slightly bent by the tension of a bowstring connecting its two ends. The bow stored the force of the archer's draw as potential energy, then transferred it to the bowstring as kinetic energy, imparting velocity and killing power to the arrow. The bow could store no more energy than the archer was capable of producing in a single movement of the muscles of his back and arms, but it released the stored energy at a higher velocity, thus overcoming the arm's inherent limitations.

Though not as evident, the sophistication of arrow technology matched that of the bow. The effectiveness of the bow depended on the arrow's efficiency in retaining kinetic energy throughout its trajectory and then transforming it into killing power on impact. This was not a simple problem, as it depended on the mass, aerodynamic drag, and stability of the arrow and on the hardness and shape of the head. These factors were related to one another and to the characteristics of the bow in a complex calculus. The most important variables in this calculus were arrow weight and the length and stiffness of the bow.

Assuming the same length of draw and available force, the total amount of potential energy that an archer could store in a bow was a function of the bow's length; that is, the longer the arms of the bow, the more energy stored per unit of work expended in the draw and, therefore, the more kinetic energy imparted to the string and arrow. The disadvantage of a long bow was that the stored energy had to serve not only to drive the string and arrow but also to accelerate the mass of the bow itself. Because the longer bow's more massive arms accelerated more slowly, a longer bow imparted kinetic energy to the string and arrow at a lower velocity. A shorter bow, on the other hand, stored less energy for the same amount of work expended in the draw, but it compensated for this through its ability to transmit the energy to the arrow at a higher velocity. In sum, the shorter bow imparted less total energy to the arrow, but it did so at a higher velocity. Therefore, in practice maximum range was attained by a short, stiff bow shooting a very light arrow, and maximum killing power at medium ranges was attained by a long bow driving a relatively heavy arrow.

Tactical advantages of short and long bows

The early bow. The simple bow, made from a single piece of wood, was known to Neolithic hunters; it is clearly depicted in cave paintings of 30,000 BC and earlier. The first improvement was the reflex bow, a bow that was curved forward, or reflexively, near its centre so that the string lay close against the grip before the bow was drawn. This increased the effective length of the draw since it began farther forward, close to the archer's left hand.

The composite recurved bow. The next major improvement, one that was to remain preeminent among missile weapons until well into the modern era, was the composite recurved bow. This development overcame the inherent limitations of wood in stiffness and tensile strength. The composite bow's resistance to bending was increased by reinforcing the rear, or belly, of the bow with horn; its speed and power in recoil were increased by overlaying the front of the bow with sinew, usually applied under tension. The wooden structure of this composite thus consisted of little more than thin wooden strips supporting the horn and sinew. The more powerful composite bows, being very highly stressed, reversed their curvature when unstrung. They acquired the name recurved since the outer arms of the bow curved away from the archer when the bow was strung, which imparted a mechanical advantage at the end of the draw. Monumental and artistic evidence suggest that the principle of the composite recurved bow was known as early as 3000 BC.

A prime advantage of the composite bow was that it

could be engineered to essentially any desired strength. By following the elaborate but empirically understood trade-off between length and stiffness referred to above, the bowyer could produce a short bow capable of propelling light arrows to long ranges, a long, heavy bow designed to maximize penetrative power at relatively short ranges, or any desired compromise between.

Arrows. Arrow design was probably the first area of military technology in which production considerations assumed overriding importance. As a semi-expendable munition that was used in quantity, arrows could not be evaluated solely by their technological effectiveness; production costs had to be considered as well. As a consequence, the materials used for arrowheads tended to be a step behind those used for other offensive technologies. Arrowheads of flint and obsidian, knapped to remarkably uniform standards, survived well into the Bronze Age, and bronze arrowheads were used long after the adoption of iron for virtually every other military cutting or piercing implement.

Arrow shafts were made of relatively inexpensive wood and reed throughout history, though considerable labour was involved in shaping them. Remarkably refined techniques for fastening arrowheads of flint and obsidian to shafts were well in hand long before recorded history. (The importance of arrow manufacturing techniques is reflected in the survival in modern English of the given name Fletcher, the title of a specialist in attaching feathers to the arrow shaft.)

MECHANICAL ARTILLERY

In contrast to individual weaponry, there was little continuity from classical to medieval times in mechanical artillery. The only exception—and it may have been a case of independent reinvention—was the similarity of the Roman onager to the medieval catapult.

Tension-driven *gastrophetes* and torsion-driven catapult

Mechanical artillery of classical times was of two types: tension and torsion. In the first, energy to drive the projectile was provided by the tension of a drawn bow; in the other, it was provided by torsional energy stored in bundles of twisted fibres.

The invention of mechanical artillery was ascribed traditionally to the initiative of Dionysius the Elder, tyrant of Syracuse, in Sicily, who in 399 BC directed his engineers to construct military engines in preparation for war with Carthage. Dionysius' engineers surely drew on existing practice. The earliest of the Greek engines was the *gastrophetes,* or "belly shooter." In effect a large crossbow, it received its name because the user braced the stock against his belly to draw the weapon. Though Greek texts did not go into detail on construction of the bow, it was based on a composite bow of wood, horn, and sinew. The potential of such engines was apparent, and the demand for greater power and range quickly exceeded the capabilities of tension. By the middle of the 3rd century BC, the bow had been replaced by rigid wooden arms constrained in a wooden box and drawn against the force of tightly twisted bundles of hair or sinew. The overall concept was similar to the *gastrophetes,* but the substitution of torsion for tension permitted larger and more powerful engines to be made. Such catapults (from Greek *kata,* "to pierce," and *pelte,* "shield"; a "shield piercer") could throw a javelin as far as 800 yards (700 metres). The same basic principle was applied to large stone-throwing engines. The Jewish historian Josephus referred to Roman catapults used in the siege of Jerusalem in AD 70 that could throw a one-talent stone (about 55 pounds, or 25 kilograms) two *stades* (400 yards) or more.

The terminology of mechanical artillery is confusing. Catapult is the general term for mechanical artillery; however, the term also narrowly applies to a particular type of torsion engine with a single arm rotating in a vertical plane. Torsion engines with two horizontally opposed arms rotating in the horizontal plane, such as that described above, are called ballistae. There is no evidence that catapults in the narrow sense were used by the Greeks; the Romans called their catapults onagers, or wild asses, for the way in which their rears kicked upward under the recoil force. The Romans used large ballistae and onagers effectively in siege operations, and a complement of *carroballistae,* small, wheel-mounted torsion engines, was a regular part of the legion. The onager and the medieval catapult were identical in concept, but ballistae were not used after the classical era.

FORTIFICATION

Fortress design. Fortifications in antiquity were designed primarily to defeat attempts at escalade, though cover was provided for archers and javelin throwers along the ramparts and for enfilade fire from flanking towers. By classical Greek times, fortress architecture had attained a high level of sophistication; both the profile and trace (that is, the height above ground level and the outline of the walls) of fortifications were designed to achieve overlapping fields of fire from ballistae mounted along the ramparts and in supporting towers. Roman fortresses of the 2nd century AD, largely designed for logistic and administrative convenience, tended to have square or rectangular outlines, and were situated along major communication routes. By the late 3rd century, their walls had become thicker and had flanking towers strengthened to support mechanical artillery. The number of gates was reduced, and the ditches were dug wider. By the late 4th and 5th centuries, Roman fortresses were being built on easily defensible ground with irregular outlines that conformed to the topography; clearly, passive defense had become the dominant design consideration.

Roman fortress design

In general, the quality of masonry that went into permanent defensive works of the classical period was very high by later standards. Fortifications were almost exclusively of dressed stone, though by Roman times concrete mortar was used on occasion.

Field fortification. The main purpose of early field fortifications, particularly among the Greeks, was to secure an advantage by standing on higher ground so that the enemy was forced to attack uphill. The Romans were especially adept at field fortifications, preparing fortified camps at the close of each day's march (see Figure 2). The troops usually required three to four hours to dig a ditch around the periphery, erect a rampart or palisade from timbers carried by each man, lay out streets, and pitch tents. During extended campaigns the Romans strengthened the camps with towers and outlying redoubts, or small forts, and used the camps as bases for offensive forays into the surrounding territory.

Siege towers. For breaching fortified positions, military engineers of the classical age designed assault towers that remain a wonder to modern engineers. So large was one siege tower used by Macedonians in an attack on Rhodes that 3,400 men were required to move it up to the city walls. Another 1,000 men were needed to wield a battering ram 180 feet (55 metres) long. The Romans constructed huge siege towers, one of which Caesar mentions as being 150 feet high. The lower stories housed the battering ram, which had either a pointed head for breaching or a ram-like head for battering. Archers in the upper stories shot arrows to drive the defenders from their ramparts. From the top of the tower, a hinged bridge might be lowered to serve a storming party. To guard the attackers against enemy missiles, the Romans used great wicker or wooden shields, called mantelets, which were sometimes mounted on wheels. In some cases the attackers might approach the fortress under the protection of wooden galleries.

LAND TRANSPORTATION

In antiquity and classical times the transportation technology of land warfare largely amounted to man's own powers of locomotion. This was due in part to limitations in the size, strength, and stamina of horses and in part to deficiencies in crucial supporting technologies, notably the inefficiency of harnesses for horses and nonpivoting front axles for wagons. A more basic underlying factor was the generally low level of economic development. The horse was an economically inefficient animal, consuming large quantities of food. Of more importance, keeping horses—let alone selectively breeding them for size, strength, and power—was a highly labour-intensive and capital-intensive enterprise for which the classical world was not organized.

The expense of horse-drawn transport

An efficient pulling harness for horses was unknown, and mules and donkeys fitted with carrying baskets, or panniers, balanced in pairs across the back, were the most common pack or dray animals. The ox, the heavy-duty dray animal of the Mediterranean world, was used for military purposes when heavy loads were involved and speed was not critical.

The horse. Because it was not possible to maintain a breed of war-horses sufficiently powerful to sustain mounted shock action, the horse was restricted to a subsidiary role in warfare from the eclipse of the chariot in the middle of the 2nd millennium BC until the rise of the horse archer in the 4th century AD. Evidence as to the size of horses in classical times is equivocal. Greek vase paintings from the 7th century BC depict Scythians riding tall, apparently powerful horses with long, slender legs, implying speed; however, this breed evidently collapsed and disappeared. Later Mongolian steppe ponies, though tough and tractable, were probably considerably smaller.

Horses were rarely if ever used for drayage. This was partly because their rarity and expense restricted them to combat roles, and partly because of the lack of a suitable harness. The prevalent harness consisted of a pole-and-yoke assembly, attached to the animal by neck and chest harness. This was developed for use with oxen, where the primary load was absorbed by the thrust of the animal's hump against the yoke. With a horse, most of the pulling load was borne by the neck strap, which tended to strangle the horse and constrict blood flow.

The elephant. The war elephant was first used in India and was known to the Persians by the 4th century BC. Though they accomplished little subsequently, their presence in Hannibal's army during its transit of the Alps into Italy in 218 BC underscored their perceived utility. The elephant's tactical importance apparently stemmed in large part from its willingness to charge both men and horses and from the panic that it inspired in horses.

The chariot. The chariot was the earliest means of transportation in combat other than man's own powers of locomotion. The earliest known chariots, shown in Sumerian depictions from about 2500 BC, were not true chariots but four-wheeled carts with solid wooden wheels drawn by a team of four donkeys or wild asses. They were no doubt heavy and cumbersome; lacking a pivoting front axle, they would have skidded through turns.

Chariots of the ancient Middle East

Around 1600 BC Iranian tribes introduced the war-horse into Mesopotamia from the north, along with the light two-wheeled chariot. The Hyksos apparently introduced the chariot into Egypt shortly thereafter, by which time it was a mature technology. By the middle of the 2nd millennium BC, Egyptian, Hittite, and Palestinian chariots were extraordinarily light and flexible vehicles, the wheels and tires in particular exhibiting great sophistication in design and fabrication. Light war chariots were drawn by either two or three horses, which were harnessed by means of chest girths secured by one or two poles and a yoke.

That horses were long used for pulling chariots rather than for riding is probably attributable to the horse's inadequate strength and incomplete domestication. The chariot was subject to mechanical failure and, more importantly, was immobilized when any one of its horses was incapacitated. Moreover, the art of riding astride in cavalry fashion had been mastered long before the chariot's eclipse as a tactically dominant weapon. The decline of the chariot by the end of the 2nd millennium BC was probably related to the spread of iron weaponry, but it was surely related also to the breeding of horses with sufficient strength and stamina to carry an armed man. Chariots lingered in areas of slower technological advance, but in the classical world they were retained mainly for ceremonial functions.

The age of cavalry, c. AD 400–1350

The beginning of the age of cavalry in Europe is traditionally dated to the destruction of the legions of the Roman emperor Valens by Gothic horsemen at the Battle of Adrianople in AD 378. The period that followed, characterized by the network of political and economic relationships called feudalism, was an age during which the mounted arm assumed an ascendancy that it began to relinquish only in the 14th century, with the appearance of infantry capable of taking the open field unsupported against mounted chivalry. Cavalry, however, was only part of the story of this era. However impressive the mounted knight may have been in battle, he required a secure place of replenishment and refuge. This was provided by the seigneurial fortress, or castle. In a military sense, European feudalism rested on a symbiotic relationship between armoured man-at-arms, war-horse, and castle.

Economic base of the cavalry elite

The tactical dominance in Europe of the heavy mounted elites had a number of complex causes. It is clear that a basic reorientation of the means of production and of the social distribution of the means of armed violence was involved. Horses required large quantities of grain, and in an agricultural economy where returns on seed grain were as little as 2 to 1, mounted shock action could not have solidified its dominance without improvements in agricultural production. Perhaps ironically, these improvements seem to have involved the development of a means of harnessing the horse to agricultural transport and the plow—particularly beginning in the 14th century, when seed-to-yield ratios began to improve.

The age of heavy shock cavalry did not come on suddenly, ushered in by the stirrup or any other single invention. Improvements in the breeding of war-horses played a major and perhaps dominant role. The Germanic tribes that pressed against the boundaries of Rome from the 3rd century on may have made a breakthrough in horse breeding, and, in the Arab conquests of the 7th century and following, the superior breed of the Arabian horse was a major determinant of tactical success. The stirrup alone meant little without powerful war-horses and supporting technologies such as saddle, girth, and bridle.

Using scattered artistic and archaeological evidence, historians have constructed an approximate chronology of technological innovation in medieval Europe. The war saddle with a single girth was introduced by the 6th century, and the iron stirrup was common by the 7th (having probably been known earlier in the East). The curb bit, vitally important for controlling a war-horse, probably dates from about the same time. According to literary evidence, iron horseshoes date from the end of the 9th century, and, based on pictorial evidence, spurs date from the 11th. By the 12th century the European knight was using a war saddle with high, wraparound cantle and pommel that protected the genitals and held him securely in his seat; the saddle itself was secured to the horse by a double girth that held it firmly in place fore and aft. These developments welded horse and rider into a single unit and enabled the knight to apply much of the force of his horse's charge to the point of the lance, held couched beneath the arm, without being driven over the horse's rump on impact. An associated development dating from the end of the 12th century was the incorporation of a rigid backplate into knightly armour; this, backed with several inches of padding, braced the man-at-arms against the shock of head-on impact and protected his kidneys from the cantle. These developments were accompanied, and in part caused, by increases in the size and power of war-horses and steady improvements in personal armour.

THE WAR-HORSE

The destrier, or medieval war-horse, was central to the tactical viability of European feudalism. This animal was a product of two great migrations of horses originating in Central Asia. One, moving westward, crossed into Europe and there originated the vast herds of primeval animals that eventually roamed almost the entire continent. The second flowed to the southwest and found its way into Asia Minor and the neighbouring lands of Persia, India, and Arabia. Ultimately it crossed into Egypt, then spread from that country along all of North Africa. At the same time it crossed from Asia Minor into Greece and spread along the northern shores of the Mediterranean.

There were two channels through which the horses of Arabia and North Africa were distributed into northern Europe. One was through the conquest of the Romans across the Alps into France and the Low Countries, where,

previously, descendants of the horses of Central Asia had constituted the equine population. The other channel led northward through Greece, Macedonia, and the Gothic countries into the land of the Vandals. When these barbarian peoples invaded the empire, the vast number of horses that they possessed helped them to overthrow the Romans. The era that followed witnessed the collapse of the Roman breeds and the gradual development—especially during the era of Charlemagne in the late 8th and early 9th centuries—of improved types, owing largely to the importation of Arabian stock. The most important of these was the "great horse," which originated in the Low Countries; its size and strength were required to carry the heavy load of the armoured knight. These horses, the ancestors of modern draft breeds, were bred from the largest and most powerful of the northern European horses, but there was apparently an admixture of Arabian breeds as well.

The Crusades of the 12th and 13th centuries took the nobility of Europe into the native land of the Arabian horse. The speed and agility of these light horses so impressed them that large numbers were imported into England and France. Over a long period of time the Moors took Arabian and North African horses into Spain, where they were crossed with the native stock and produced the superior breeds that were sought after by other nations. (Spanish horses were also taken to the New World, where they became the principal ancestors of the equine population of North and South America.)

Difficulties in breeding and training horses

The breeding, care, and maintenance of medieval warhorses, and the mastering of the skills of mounted combat, required immense amounts of time, skill, and resources. Horses strong enough to be ridden did not exist everywhere, and European horses in particular tended to revert in a feral state to a small animal not much larger than a Shetland pony. On the other hand, the horse was genetically tractable, and breeders learned that hard inbreeding could produce larger, more powerful animals. Still, it was difficult to establish a breed, and only careful control of bloodlines could maintain one. While crossbreeding could produce size and power, it also promoted instability and was best abandoned as soon as the desired traits were "fixed." This was not easy, particularly where the resources available to maintain a nonproductive breeding stock were limited. The net result was that breeds of large, powerful horses suitable for mounted combat were difficult to establish and expensive to maintain, and they were often lost in the turmoil of war. Even when herds were not dispersed or destroyed, a breed could be lost through indiscriminate breeding arising from a need for numbers.

PERSONAL ARMOUR

Mail. The availability to mounted warrior elites of iron armour of high quality, particularly mail, was instrumental in the fall of Rome and in the establishment of European feudalism. Until the 10th century, however, there was little qualitative difference between the body armour of the western European knight and the Roman legionnaire's *lorica hamata*. Then, during the 11th century, the sleeves of the knight's mail shirt, or byrnie, became longer and closer-fitting, extending downward from the middle of the upper arm to the wrist; at the same time, the hem of the byrnie dropped from just above to just below the kneecap. Knights began wearing the gambeson, a quilted garment of leather or canvas, beneath their mail for additional protection and to cushion the shock of blows. (Ordinary soldiers often wore a gambeson as their only protection.) Use of the surcoat, a light garment worn over the knight's armour, became general during this period. Both gambeson and surcoat may have been Arab imports, adopted as a result of exposure to Muslim technology during the Crusades.

The mail hauberk

Norman men-at-arms were protected by a knee-length mail shirt called a hauberk, which was a later version of the Saxon byrnie that was split to permit the wearer to sit astride his horse (see Figure 3). Though 11th-century men-at-arms probably did not have complete mail trousers, the hauberk apparently had inserts of cloth or leather, giving the same effect. It also included a hoodlike garment of mail worn over the head to protect the neck and throat; this had a hole for the face much like a modern ski cap. The hood was backed by padding of cloth or leather, and a pointed iron helmet with nasal (a vertical bar protecting the nose) was worn over it. The knight's defensive equipment was completed by a large, kite-shaped shield, nearly two-thirds the height of its owner. The size of this shield was testimony to the incomplete protection offered by the hauberk.

The British Library

Figure 3: *Crusader in iron mail.*
The mail shirt, or hauberk, is split below the waist for easier riding on horseback. Beneath the hauberk is a padded garment called the gambeson, and over it is worn the surcoat. The iron helm is pierced by slits for seeing and breathing. From the Westminster Psalter, 12th century. In the British Library (Royal MS. 2A xxii, fol. 220).

During the 12th century the open helmet with nasal evolved into the pot helm, or casque. This was an involved process, with the crown of the helmet losing its pointed shape to become flat and the nasal expanding to cover the entire face except for small vision slits and breathing holes. The late 12th-century helm was typically a barrel-shaped affair; however, more sophisticated designs with hinged visors appeared as well. The helm was extremely heavy, and the entire weight was borne by the neck; for this reason it was only donned immediately before combat. Some knights preferred a mail coif, no doubt with heavy padding and perhaps an iron cap beneath. One 12th-century depiction shows an iron visor worn over a coif of mail.

By the early 13th century European amourers had learned to make mail with a sufficiently fine mesh to provide protection to the hand. At first this was in the form of mittens with a leather-lined hole in the palm through which the knight could thrust his hand when out of action; by mid-century the armourer's skill had developed to the point of making complete gloves of mail.

Plate. The earliest knightly plate armour appeared shortly after 1200 in the form of thin plates worn beneath the gambeson. External plate armour began to appear around the middle of the century, at first for elbows, kneecaps, and shins. The true plate cuirass appeared about 1250, though it was at first unwieldy, covering only the front of the torso and no doubt placing considerable stress on the underlying garments to which it was attached. Perhaps in part for this reason, the breastplate was followed shortly by the backplate. From the late 13th century,

plate protection spread from the knees and elbows to encompass the extremities; square plates called ailettes, which protected the shoulder, made a brief appearance between about 1290 and 1325 before giving way to jointed plate defenses that covered the gap between breastplate and upper-arm defenses. Helmets with hinged visors appeared about 1300, and by mid-century armourers were constructing closed, visored helms that rested directly on the shoulder defenses. Plate armour, at first worn above mail as reinforcement, began to replace it entirely except in areas such as the crotch, the armpits, and the back of the knees, where the armourer's skill could not devise a sufficiently flexible joint. In response to this enhanced coverage, the knight's large, kite-shaped shield evolved into a much smaller implement.

Gothic plate armour

The first suits of full plate armour date from the first decades of the 15th century. By 1440 the Gothic style of plate armour was well developed, representing the ultimate development of personal armour protection (see Figure 4). Armourers were making gloves with individually jointed fingers, and shoulder defenses had become particularly sophisticated, permitting the man-at-arms full freedom to wield sword, lance, or mace with a minimum of exposure. Also during the 15th century the weight of personal armour increased, partly because of the importance of shock tactics in European warfare and partly because of the demands of jousting, a form of mock combat in which two armoured knights, separated by a low fence or barrier, rode at each other head-on and attempted to unseat each other with blunted lances. As armour protection became more complete and heavier, larger breeds of horses appeared. Mail protection for horses became common in the 13th century; by the 15th, plate horse armour was used extensively.

Reproduced by permission of the Trustees of the Wallace Collection, London

Figure 4: *Gothic-style armour for man and horse.* Made at Landshut, Bavaria, about 1480, this armour represented the high point of armourers' art, but within a century the crossbow bolt and harquebus ball made full plate armour obsolete. In the Wallace Collection, London.

The unprecedented protection that plate armour gave the man-at-arms did not come without tactical, as well as economic, cost. A closed helm seriously interfered with vision and made voice communication in battle impossible. No doubt in response to this, heraldry emerged during this period and the armorial surcoat became a standard item of knightly dress. Ultimately, the thickness of iron needed to stop missiles—at first arrows and crossbow bolts, then harquebus and musket balls—made armour so heavy as to be impractical for active service. By the 16th century, armour was largely ceremonial and decorative, with increasingly elaborate ornamentation.

THE CASTLE

The motte-and-bailey castle. The earliest distinctive European fortification characteristic of feudal patterns of social organization and warfare was the motte-and-bailey castle, which appeared in the 10th and 11th centuries between the Rhine and Loire rivers and eventually spread to most of western Europe. The motte-and-bailey castle consisted of an elevated mound of earth, called the motte, which was crowned with a timber palisade and surrounded by a defensive ditch that also separated the motte from a palisaded outer compound, called the bailey. Access to the motte was by means of an elevated bridge across the ditch from the bailey. The earliest motte-and-bailey castles were built where the ground was suitable and timber available, these factors apparently taking precedence over considerations such as proximity to arable land or trade routes. Later on, as feudal social and economic relationships became more entrenched, castles were sited more for economic, tactical, and strategic advantage and were built of imported stone. The timber palisade was replaced with a keep, or donjon, of dressed stone, and the entire enclosure, called the enceinte, was surrounded by a wall.

The motte-and-bailey castle was not the only pattern of European fortification. There was, for example, a tradition of fortified towns, stemming from Roman fortification, that enjoyed a tenuous existence throughout the Dark Ages, particularly in the Mediterranean world.

Stone fortifications. The greatest weakness of timber fortifications was vulnerability to fire; in addition, a determined attacker, given enough archers to achieve fire dominance over the palisade, could quickly chop his way in. A stone curtain wall, on the other hand, had none of these deficiencies. It could be made high enough to frustrate improvised escalade and, unlike a wooden palisade, could be fitted with a parapet and crenellated firing positions along the top to give cover to defending archers and crossbowmen (see Figure 5). Stone required little maintenance or upkeep, and it suffered by comparison with timber only in the high capital investment required to build with it.

The stone curtain wall

Given walls high enough to defeat casual escalade, the prime threats to stone fortresses were the battering ram and attempts to pry chunks out of the wall or undermine it. Since these tactics benefited from an unprotected footing at the base of the wall, most of the refinements of medieval fortress architecture were intended to deny an undisturbed approach. Where terrain permitted, a moat was dug around the enceinte. Towers were made with massive, protruding feet to frustrate attempts at mining. Protruding towers also enabled defenders to bring flanking fire along the face and foot of the wall, and the towers were made higher than the wall to give additional range to archers and crossbowmen. The walls themselves were fitted with provisions for hoardings, which were overhanging wooden galleries from which arrows, stones, and unpleasant substances such as boiling tar and pitch could be dropped or poured on an attacker. Hoardings gave way to machicolations, permanent overhanging galleries of stone that became a distinctive feature of medieval European fortress architecture.

Castle entrances, which were few and small to begin with, were protected by barbicans, low-lying outworks dominated by the walls and towers behind. Gates were generally deeply recessed and backed by a portcullis, a latticework grate suspended in a slot that could be dropped quickly to prevent surprise entry. The gate could also be sealed by means of a drawbridge. These measures were sufficiently effective that medieval sieges were settled more often by treachery, starvation, or disease than by breached walls and undermined towers.

SIEGE WEAPONS

The most basic means of taking a fortress were to storm the gate or go over the wall by simple escalade using ladders, but these methods rarely succeeded except by surprise or treachery. Beginning in the 9th century, European engineers constructed wheeled wooden siege towers, called

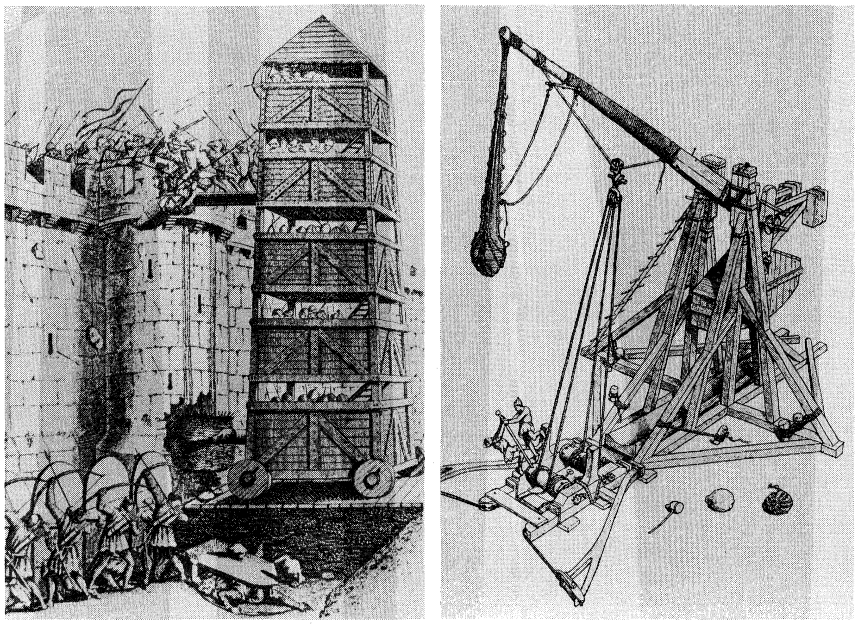


Figure 5: *Engines of medieval siegecraft.*
(Left) A siege tower, or belfroy, has been brought to the castle wall, and a storming party crosses the drawbridge under cover from archers. (Right) Soldiers winch down the arm of a trebuchet. Large trebuchets, powered by 10-ton counterweights, could hurl 300-pound wall-smashing boulders as far as 300 yards.
Ian V. Hogg

belfroys. These were fitted with drawbridges, which could be dropped onto the parapet, and with protected firing positions from which the defending parapets could be swept by arrow fire (see Figure 5). Constructing one of these towers and moving it forward against an active defense was a considerable feat of engineering and arms. Typically, the moat had to be filled and leveled, all under defensive fire, and attempts to burn or dismount the tower had to be prevented. The wooden towers were vulnerable to fire, so that their faces were generally covered with hides.

Battering rams were capable of bringing down sections of wall, given sufficient time, manpower, and determination. Large battering rams were mounted on wheels and were covered by a mobile shed for protection from defensive fire.

Mining castle walls

The most powerful method of direct attack on the structure of a fortress was mining, digging a gallery beneath the walls and supporting the gallery with wooden shoring. Once completed, the mine was fired to burn away the shoring; this collapsed the gallery and brought down the walls. Mining, of course, required suitable ground and was susceptible to countermining by an alert defender.

The trebuchet. In general, the mechanical artillery of medieval times was inferior to that of the Classical world. The one exception was the trebuchet, an engine worked by counterpoise. Counterpoise engines appeared in the 12th century and largely replaced torsion engines by the middle of the 13th. The trebuchet worked something like a seesaw. Suspended from an elevated wooden frame, the arm of the trebuchet pivoted from a point about one-quarter of the way down its length. A large weight, or counterpoise, was suspended from the short end, and the long end was fitted with a hollowed-out spoonlike cavity or a sling. (A sling added substantially to the trebuchet's range.) The long end was winched down, raising the counterpoise; a stone or other missile was put into the spoon or sling, and the arm was released to fly upward, hurling the missile in a high, looping arc toward its target. Though almost anything could be thrown, spherical projectiles of cut stone were the preferred ammunition (see Figure 5).

Trebuchets might have a fixed counterpoise, a pivoted counterpoise, or a counterpoise that could be slid up and down the arm to adjust for range. Ropes were frequently attached to the counterpoise to be pulled on for extra power. Modern experiments suggest that a trebuchet with an arm about 50 feet (15 metres) long would have been capable of throwing a 300-pound (135-kilogram) stone to a distance of 300 yards (275 metres); such a trebuchet would have had a counterpoise of about 10 tons. Though the rate of fire was slow, and prodigious quantities of timber and labour were required to build and serve one, a large trebuchet could do serious damage to stone fortifications. The machines were apparently quite accurate, and small trebuchets were useful in sweeping parapets of archers and crossbowmen.

Greek fire. Greek fire was a weapon that had a decisive tactical and strategic impact in the defense of the Byzantine Empire. It was first used against the Arabs at the siege of Constantinople of 673. Greek fire was a liquid that ignited on contact with seawater. It was viscous and burned fiercely, even in water. Sand and—according to legend—urine were the only effective means of extinguishing the flames. It was expelled by a pumplike device similar to a 19th-century hand-pumped fire engine, and it may also have been thrown from catapults in breakable containers. Although the exact ingredients of Greek fire were a Byzantine state secret, other powers eventually developed and used similar compositions. The original formula was lost and remains unknown. The most likely ingredients were colloidal suspensions of metallic sodium, lithium, or potassium—or perhaps quicklime—in a petroleum base.

Greek fire was particularly effective in naval combat, and it constituted one of the few incendiary weapons of warfare afloat that were used effectively without backfiring on their users. It may have been used following the sack of Constantinople by Venetian-supported crusaders in 1204, but it probably disappeared from use after the fall of Constantinople to the Turks in 1453.

THE HORSE ARCHER

The age of cavalry came to be viewed from a European perspective, since it was there that infantry was overthrown and there that the greatest and most far-reaching changes occurred. But it was by no means an exclusively European phenomenon; to the contrary, the mounted warrior's tactical supremacy was less complete in western Europe than in any other region of comparably advanced technology save Japan, where a strikingly parallel feudal situation prevailed. Indeed, from the 1st century AD nomadic horse archers had strengthened their hold over the Eurasian Steppes, the Iranian plateau, and the edges of the Fertile Crescent, and, in a series of waves extending

Mounted warfare in Eurasia

through medieval times, they entered Europe, China, and India and even touched Japan briefly in the 13th century. The most important of these incursions into the European and Chinese military ecospheres left notable marks on the military technology of East Asia and the Byzantine Empire, as well as on the kingdoms of Europe.

The Huns and Avars. The first of the major horse nomad incursions into Europe were the Hunnish invasions of the 4th century. The Huns' primary significance in the history of military technology was in expanding the use of the composite recurved bow into the eastern Roman Empire. This important instance of technological borrowing constituted one of the few times in which a traditional military skill as physiologically and economically demanding as composite archery was successfully transplanted out of its original cultural context.

The Avars of the 6th and 7th centuries were familiar with the stirrup, and they may have introduced it into Europe. Some of the earliest unequivocal evidence of the use of the stirrup comes from Avar graves.

The Byzantine cataphract. Although they continued to make effective use of both shock and missile infantry, the Byzantines turned to cavalry earlier and more completely than did the western Roman Empire. After an extended period of dependence on Teutonic and Hunnish mercenary cavalry, the reforms of the emperors Maurice and Heraclius in the 6th and 7th centuries developed an effective provincial militia based on the institution of *pronoia,* the award of nonhereditary grants of land capable of supporting an armoured horse archer called a cataphract. *Pronoia,* which formed the core of the Byzantine army's strength during the period of its greatest efficiency in the 8th through 10th centuries, entailed the adoption of the Hunnish composite recurved bow by native troopers.

The Byzantine cataphract was armed with bow, lance, sword, and dagger; he wore a shirt of mail or scale armour and an iron helm and carried a small, round, ironbound shield of wood that could be strapped to the forearm or slung from the waist. The foreheads and breasts of officers' horses and those of men in the front rank were protected with frontlets and poitrels of iron. The militia cataphracts were backed by units of similarly armed regulars and mercenary regiments of Teutonic heavy shock cavalry of the imperial guard. Mercenary horse archers from the steppe continued to be used as light cavalry.

The Turks. The infiltration of Turkish tribes into the Eurasian military ecosphere was distinguished from earlier steppe nomad invasions in that the raiders were absorbed culturally through Islāmization. The long-term results of this wave of nomadic horse archers were profound, leading to the extinction of the Byzantine Empire.

Armour and weaponry of the Seljuq Turks

Turkish horse archers, of whom the Seljuqs were representative, were lightly armoured and mounted but extremely mobile. Their armour generally consisted of an iron helmet and, perhaps, a shirt of mail or scale armour (called brigandine). They carried small, light, one-handed shields, usually of wicker fitted with an iron boss. Their principal offensive arms were lance, sabre, and bow. The Turkish bow developed in response to the demands of mounted combat against lightly armoured adversaries on the open steppe; as a consequence, it seems to have had greater range but less penetrative and knockdown power at medium and short ranges than its Byzantine equivalents. Turkish horses, though hardy and agile, were not as large or powerful as Byzantine chargers. Therefore, Turkish horse archers could not stand up to a charge of Byzantine cataphracts, but their greater mobility generally enabled them to stay out of reach and fire arrows from a distance, wearing their adversaries down and killing their horses.

The Mongols. The 13th-century Mongol armies of Genghis Khan and his immediate successors depended on large herds of grass-fed Mongolian ponies, as many as six or eight to a warrior. The ponies were relatively small but agile and hardy, well-adapted to the harsh climate of the steppes. The Mongol warrior's principal weapon was the composite recurved bow, of which he might carry as many as three. Characteristically, each man carried a short bow for use from the saddle and a long bow for use on foot. The former, firing light arrows, was for skirmishing and long-range harassing fire; the latter had the advantage in killing power at medium ranges. The saddle bow was probably capable of sending a light arrow more than 500 yards; the heart of the long bow's engagement envelope would have been about 100–350 yards, close to that of the contemporary English longbow. Each warrior carried several extra quivers of arrows on campaign. He also carried a sabre or scimitar, a lasso, and perhaps a lance. Personal armour included a helmet and breastplate of iron or lacquered leather, though some troops wore shirts of scale or mail.

Mongol armies were proficient at military engineering and made extensive use of Chinese technology, including catapults and incendiary devices. These latter probably included predecessors of gunpowder, of which the Mongols were the likely vehicle of introduction into western Europe.

The infantry revolution, c. 1200–1500

The appearance of the crossbow as a serious military implement along the northern rim of the western Mediterranean at about the middle of the 9th century marked a growing divergence between the technology of war in Europe and that of the rest of the world. It was the first of a series of technological and tactical developments that culminated in the rise of infantry elites to a position of tactical dominance. This infantry revolution began when the crossbow spread northward into areas that were peripheral to the economic, cultural, and political core of feudal Europe and where the topography was unfavourable for mounted shock action and the land too poor to support an armoured elite. Within this closed military topography, the crossbow soon proved itself the missile weapon par excellence of positional and guerrilla warfare.

The efficiency of the crossbow

The reasons for the crossbow's success were simple: crossbows were capable of killing the most powerful of mounted warriors, yet they were far cheaper than warhorses and armour and were much easier to master than the skills of equestrian combat. Also, it was far easier to learn to fire a crossbow than a long bow of equivalent power. Serious war bows had significant advantages over the crossbow in range, accuracy, and maximum rate of fire, but crossbowmen could be recruited and trained quickly as adults, while a lifetime of constant practice was required to master the Turkish or Mongol composite bow or the English longbow.

The crossbow directly challenged the mounted elite's dominance of the means of armed violence—a point that the lay and ecclesiastical authorities did not miss. In 1139 the second Lateran Council banned the crossbow under penalty of anathema as a weapon "hateful to God and unfit for Christians," and Emperor Conrad III of Germany (reigned 1138–52) forbade its use in his realms. But the crossbow proved useful in the Crusades against the infidel and, once introduced, could not be eradicated in any event. This produced a grudging acceptance among the European mounted elites, and the crossbow underwent a continuous process of technical development toward greater power that ended only in the 16th century, with the replacement of the crossbow by the harquebus and musket.

An independent, reinforcing, and almost simultaneous development was the appearance of the English longbow as the premier missile weapon of western Europe. The signal victory of an outnumbered English army of longbowmen and dismounted men-at-arms over mounted French chivalry supported by mercenary Genoese crossbowmen at Crécy on Aug. 26, 1346, marked the end of massed cavalry charges by European knights for a century and a half (see Figure 6).

Another important and enduring discovery was made by the Swiss. At the Battle of Morgarten in 1315, Swiss *eidgenossen,* or "oath brothers," learned that an unarmoured man with a seven-foot (200-centimetre) halberd could dispatch an armoured man-at-arms. Displaying striking adaptability, they replaced some of their halberds with the pike, an 18-foot spear with a small, piercing head. No longer outreached by the knight's lance, and displaying

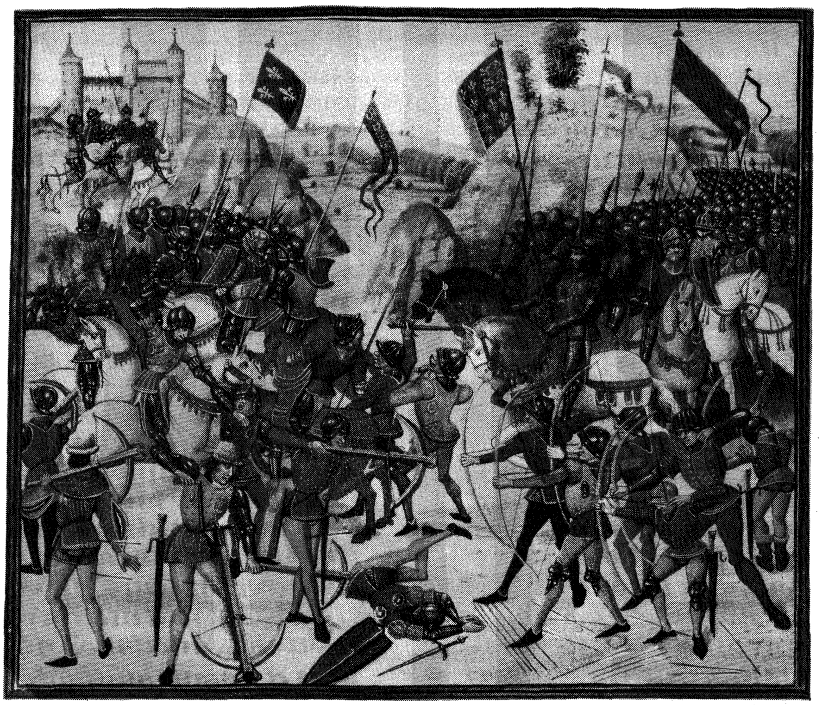


Figure 6: *Crossbow and longbow at the Battle of Crécy, 1346.*
Although the crossbow was deadly at short range, the longbow's greater range and faster rate of fire enabled English longbowmen (right) to sweep the field at Crécy of Italian crossbowmen and French horsemen (left). Both weapons signaled the end of medieval chivalry. From the *Chronicles* of Jean Froissart, 14th century. In the Bibliothèque Nationale, Paris (MS. Français 2643, fol. 165v).
Bibliotheque Nationale, Paris

far greater cohesion than any knightly army, the Swiss soon showed that they could defeat armoured men-at-arms, mounted or dismounted, given anything like even numbers. With the creation of the pike square tactical formation, the Swiss provided the model for the modern infantry regiment.

THE CROSSBOW

The idea of mounting a bow permanently at right angles across a stock that was fitted with a trough for the arrow, or bolt, and a mechanical trigger to hold the drawn string and release it at will was very old. Crossbows were buried in Chinese graves in the 5th century BC, and the crossbow was a major factor in Chinese warfare by the 2nd century BC at the latest. The Greeks used the crossbow principle in the *gastrophetes,* and the Romans knew the crossbow proper as the *manuballista,* though they did not use it extensively. The European crossbow of the Middle Ages differed from all of these in its combination of power and portability.

In Europe, crossbows were progressively developed to penetrate armour of increasing thicknesses. In China, on the other hand, crossbow development emphasized rapidity of fire rather than power; by the 16th century, Chinese artisans were making sophisticated lever-actuated rapid-fire crossbows that carried up to 10 bolts in a self-contained magazine. These, however, were feeble weapons by contemporary European standards and had relatively little penetrating power.

Mechanical cocking aids freed the crossbow from the limitations of simple muscular strength. If the bow could be held in a drawn state by a mechanical trigger, then the bow could be drawn in progressive stages using levers, cranks, and gears or windlass-and-pulley mechanisms, thereby multiplying the user's strength. The power of such a weapon, unlike that of the bow, was thus not limited by the constraints of a single muscular spasm.

Tactical deficiencies of the crossbow

The crossbowman, unlike the archer, did not have to be particularly strong or vigorous, and his volume of fire was not as limited by fatigue. Nevertheless, the crossbow had serious tactical deficiencies. First, ordinary crossbows for field operations (as opposed to heavy siege crossbows) were outranged by the bow. This was because crossbow bolts were short and heavy, with a flat base to absorb the initial impact of the string. The flat base and relatively crude leather fins (crossbow bolts were produced in volume and were not as carefully finished as arrows) were aerodynamically inefficient, so that velocity fell off more quickly than that of an arrow. These factors, combined with the inherent lack of precision in the trigger and release mechanism, made the ordinary military crossbow considerably shorter-ranged and less accurate than a serious military bow in the hands of a skilled archer. Also, the advantage of the crossbow's greater power was offset by its elaborate winding mechanisms, which took more time to use. The combination of short range, inaccuracy, and slow rate of fire meant that crossbowmen in the open field were extremely vulnerable to cavalry.

The earliest crossbows had a simple bow of wood alone. However, such bows were not powerful enough for serious military use, and by the 11th century they gave way to composite bows of wood, horn, and sinew. The strength of crossbows increased as knightly armour became more effective, and, by the 13th century, bows were being made of mild steel. (The temper and composition of steel used for crossbows had to be precisely controlled, and the expression "crossbow steel" became an accepted term designating steel of the highest quality.) Because composite and steel crossbows were too powerful to be cocked by the strength of the arms alone, a number of mechanical cocking aids were developed. The first such aid of military significance was a hook suspended from the belt: the crossbowman could step down into a stirrup set in the front of the bow's stock, loop the bowstring over the hook, and by straightening up use the powerful muscles of his back and leg to cock the weapon. The belt hook was inadequate for cocking the steel crossbows required to penetrate plate armour, and by the 14th century military crossbows were being fitted with removable windlasses and rack-and-pinion winding mechanisms called cranequins. Though slow, these devices effectively freed the crossbow from limitations on its strength: draw forces well in excess of 1,000 pounds became common, particularly for large siege crossbows.

THE ENGLISH LONGBOW

The longbow evolved during the 12th century in response to the demands of siege and guerrilla operations in the Welsh Marches, a topographically close and economically marginal area that was in many ways similar to the regions in which the crossbow had evolved three centuries earlier. It became the most effective individual missile weapon of western Europe until well into the age of gunpowder and was the only foot bow since classical times to equal the composite recurved bow in tactical effectiveness and power.

While it was heavily dependent on the strength and competence of its user, the longbow in capable hands was far superior to the ordinary military crossbow in range, rate of fire, and accuracy. Made from a carefully cut and shaped stave of yew or elm, it varied in length, according to the height of the user, from about five to seven feet. The longbow had a shorter maximum range than the short, stiff composite Turkish or Mongol saddle bows of equivalent draw force, but it could drive a heavy arrow through armour with equal efficiency at medium ranges of 150–300 yards. Each archer would have carried a few selected light arrows for shooting at extreme ranges and could probably have reached 500 yards with these.

Training the longbowman

The longbow's weakness was that of every serious military bow: the immense amounts of time and energy needed to master it. Confirmation of the extreme demands placed on the archer was found in the skeletal remains of a bowman who went down with the English ship *Mary Rose,* sunk in Portsmouth Harbour in 1545. The archer (identified as such by a quiver, its leather strap still circling his spine) exhibited skeletal deformations caused by the stresses of archery: the bones of his left forearm showed compression thickening, his upper backbone was twisted radially, and the tips of the first three fingers of his right hand were markedly thickened, plainly the results of a lifetime of drawing a bow of great strength. The longbow was dependent upon the life-style of the English yeomanry, and, as that life-style changed to make archery less remunerative and time for its practice less available, the quality of English archery declined. By the last quarter of the 16th century there were few longbowmen available, and the skill and strength of those who responded to muster was on the whole well below the standards of two centuries earlier. An extended debate in the 1580s between advocates of the longbow and proponents of gunpowder weapons hinged mainly on the small numbers and limited skills of available archers, not around any inherent technical deficiency in the weapon itself.

HALBERD AND PIKE

The halberd. The halberd was the only significant medieval shock weapon without classical antecedents. In its basic form, it consisted of a six-foot shaft of ash or another hardwood, mounted by an ax blade that had a forward point for thrusting and a thin projection on the back for piercing armour or pulling a horseman off balance. The halberd was a specialized weapon for fighting armoured men-at-arms and penetrating knightly armour. With the point of this weapon, a halberdier could fend off a mounted lancer's thrusts and, swinging the cutting edge with the full power of his arms and body, could cleave armour, flesh, and bone. The halberd's power was counterbalanced by the vulnerability of taking a full swing with both arms; once committed, the halberdier was totally dependent upon his comrades for protection. This gave halberd fighting a ferocious all-or-nothing quality and placed a premium on cohesion (see Figure 7).

The pike. While the halberd could penetrate the best plate armour, allowing infantrymen to inflict heavy casualties on their mounted opponents, the lance's advantage in length meant that men-at-arms could inflict heavy casualties in return. The solution was the pike, a staff, usually of ash, that was twice the length of the halberd and had a small piercing head about 10 inches (25 centimetres) long. Sound infantry armed with the pike could fend off cavalry with ease, even when outnumbered. As with the halberd, effectiveness of shock action with the pike was heavily dependent upon the cohesion and solidity of the troops wielding it. The pike remained a major factor in European warfare until, late in the 17th century, the bayonet gave missile-armed infantry the ability to repel charging cavalry.

Pikeman versus cavalryman

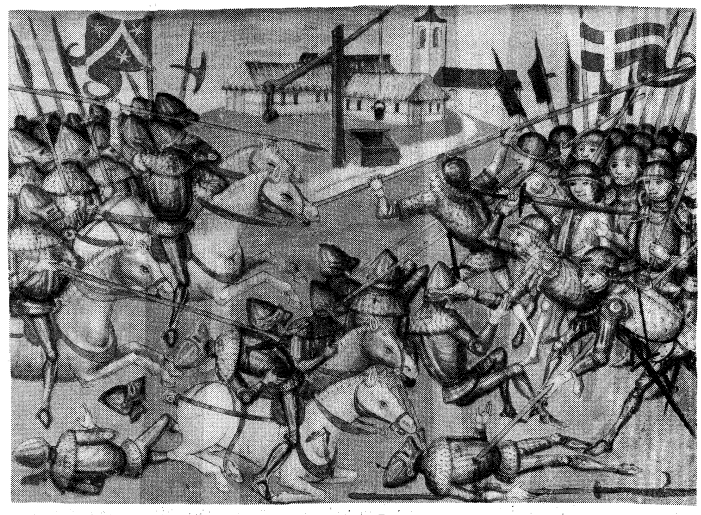


Figure 7: *Halberd and pike in battle near Ins, Berne canton, in 1375.*
Encumbered by heavy armour, the mounted French and English mercenaries at left are cut down by disciplined Swiss infantrymen wielding long, armour-piercing weapons. From the *Amtliche Chronik* by Diebold Schilling, 15th century. In the Burgerbibliothek Bern (MSS. hist. helv. 1.1, fol. 205).
Burgerbibliothek Bern

The gunpowder revolution, c. 1300–1650

Few inventions have had an impact on human affairs as dramatic and decisive as that of gunpowder. The development of a means of harnessing the energy released by a chemical reaction in order to drive a projectile against a target marked a watershed in the harnessing of energy to human needs. Before gunpowder, weapons were designed around the limits of their users' muscular strength; after gunpowder, they were designed more in response to tactical demand.

Technologically, gunpowder bridged the gap between the medieval and modern eras. By the end of the 19th century, when black powder was supplanted by nitrocellulose-based propellants, steam power had become a mature technology, the scientific revolution was in full swing, and the age of electronics and the internal combustion engine was at hand. The connection between gunpowder and steam power is instructive. Steam power as a practical reality depended on the ability to machine iron cylinders precisely and repetitively to predetermined internal dimensions; the methods for doing this were derived from cannon-boring techniques.

Gunpowder bridged the gap between the old and the new intellectually as well as technologically. Black powder was a product of the alchemist's art, and although alchemy presaged science in believing that physical reality was determined by an unvarying set of natural laws, the alchemist's experimental method was hardly scientific. Gunpowder was a simple mixture combined according to empirical recipes developed without benefit of theoretical knowledge of the underlying processes. The development of gunpowder weapons, however, was the first significant success in rationally and systematically exploiting an energy source whose power could not be perceived directly with the ordinary senses. As such, early gunpowder technology was an important precursor of modern science.

EARLY GUNPOWDER

Chinese origin of black powder

Chinese alchemists discovered the recipe for what became known as black powder in the 9th century AD; this was a mixture of finely ground potassium nitrate (also called saltpetre), charcoal, and sulfur in approximate proportions of 75:15:10 by weight. The resultant gray powder behaved differently from anything previously known; it exploded

on contact with open flame or a red-hot wire, producing a bright flash, a loud report, dense white smoke, and a sulfurous smell. It also produced considerable quantities of superheated gas, which, if confined in a partially enclosed container, could drive a projectile out of the open end. The Chinese used the substance in rockets, in pyrotechnic projectors much like Roman candles, in crude cannon, and, according to some sources, in bombs thrown by mechanical artillery. This transpired long before gunpowder was known in the West, but development in China stagnated. The development of black powder as a tactically significant weapon was left to the Europeans, who probably acquired it from the Mongols in the 13th century (though diffusion through the Arab Muslim world is also a possibility).

Chemistry and internal ballistics. Black powder differed from modern propellants and explosives in a number of important particulars. First, only some 44 percent by weight of a properly burned charge of black powder was converted into propellant gases, the balance being solid residues. The high molecular weights of these residues limited the muzzle velocities of black-powder ordnance to about 2,000 feet (600 metres) per second. Second, unlike modern nitrocellulose-based propellants, the burning rate of black powder did not vary significantly with pressure or temperature. This occurred because the reaction in an exploding charge of black powder was transmitted from grain to grain at a rate some 150 times greater than the rate at which the individual grains were consumed and because black powder burned in a complex series of parallel and mutually dependent exothermal (heat-producing) and endothermal (heat-absorbing) reactions that balanced each other out. The result was an essentially constant burning rate that differed only with the grain size of the powder; the larger the grains, the less surface area exposed to combustion and the slower the rate at which propellant gases were produced.

Nineteenth-century experiments revealed sharp differences in the amount of gas produced by charcoal burned from different kinds of wood. For example, dogwood charcoal decomposed with potassium nitrate was found to yield nearly 25 percent more gas per unit weight than fir, chestnut, or hazel charcoal and some 17 percent more than willow charcoal. These scientific observations confirmed the insistence of early—and thoroughly unscientific—texts that gunpowder made of charcoal burned from different kinds of wood was suited to different applications. Willow charcoal, for example, was preferred for cannon powder and dogwood charcoal for small arms—a preference substantiated by 19th-century tests. (A preference for urine instead of water as the incorporation agent might have had some basis in fact because urine is rich in nitrates; so might the view that a beer-drinker's urine was preferable to that of an abstemious person and a wine-drinker's urine best of all. It is difficult, however, to hypothesize a rational basis for the presumed superiority of clerical urine to that of laypersons.) For all this, the empirically derived recipe for gunpowder was fixed during the 14th century and hardly varied thereafter. Subsequent improvements were almost entirely concerned with the manufacturing process and with the ability to purify and control the quality of the ingredients.

Varying the ingredients

Serpentine powder. The earliest gunpowder was made by grinding the ingredients separately and mixing them together dry. This was known as serpentine. The behaviour of serpentine was highly variable, depending on a number of factors that were difficult to predict and control. If packed too tightly and not confined, a charge of serpentine might fizzle; conversely, it might develop internal cracks and detonate. When subjected to vibration, as when being transported by wagon, the components of serpentine separated into layers according to relative density, the sulfur settling to the bottom and the charcoal rising to the top. Remixing at the battery was necessary to maintain the proper proportions—an inconvenient and hazardous procedure producing clouds of noxious and potentially explosive dust.

Corned powder. Shortly after 1400, smiths learned to combine the ingredients of gunpowder in water and grind them together as a slurry. This was a significant improvement in several respects. Wet incorporation was more complete and uniform than dry mixing, the process "froze" the components permanently into a stable grain matrix so that separation was no longer a problem, and wet slurry could be ground in large quantities by water-driven mills with little danger of explosion. The use of waterpower also sharply reduced cost.

After grinding, the slurry was dried in a sheet or cake. It was then processed in stamping mills, which typically used hydraulically tripped wooden hammers to break the sheet into grains. After being tumbled to wear the sharp edges off the grains and impart a glaze to their surface, they were sieved. The grain size varied from coarse—about the size of grains of wheat or corn (hence the name corned powder)—to extremely fine. Powder too fine to be used was reincorporated into the slurry for reprocessing. Corned powder burned more uniformly and rapidly than serpentine; the result was a stronger powder that rendered many older guns dangerous.

Uses of different grains of powder

Refinements in ballistics. Late medieval and early modern gunners preferred large-grained powder for cannon, medium-grained powder for shoulder arms, and fine-grained powder for pistols and priming—and they were correct in their preferences. In cannon the slower burning rate of large-grained powder allowed a relatively massive, slowly accelerating projectile to begin moving as the pressure built gradually, reducing peak pressure and putting less stress on the gun. The fast burning rate of fine-grained powders, on the other hand, permitted internal pressure to peak before the light, rapidly accelerating projectile of a small arm had exited the muzzle. But the early modern gunner had no provable rationale for his preferences, and in the 18th century European armies standardized on fine-grained musket powder for cannon as well as small arms.

Then, beginning in the late 18th century, the application of science to ballistics began to produce practical results. The ballistic pendulum, invented by the English mathematician Benjamin Robins, provided a means of measuring muzzle velocity directly and, hence, of accurately gauging the effective power of a given quantity of powder. The ballistic pendulum involved a swinging mass into which a projectile was fired horizontally and lodged, absorbing the projectile's kinetic energy and converting it into upward movement against the force of gravity (that is, potential energy). According to the law of the conservation of energy, the potential energy imparted by the projectile to the pendulum should equal the projectile's kinetic energy. The velocity of the projectile could then be calculated from the formula for kinetic energy, $E_k = \frac{1}{2}mv^2$ (kinetic energy equals one-half the mass times the velocity squared).

The initial impact of science on internal ballistics was to show that traditional powder charges for cannon were much larger than necessary. Refinements in the manufacture of gunpowder followed. In about 1800 the British introduced cylinder-burned charcoal—that is, charcoal burned in enclosed vessels rather than in pits. With this method, wood was converted to charcoal at a uniform and precisely controlled temperature. The result was greater uniformity and, since fewer of the volatile trace elements were burned off, more powerful powder. Later, powder for very large ordnance was made from charcoal that was deliberately "overburned" to reduce the initial burning rate and, hence, the stress on the gun.

Beginning in the mid-19th century, the use of extremely large guns for naval warfare and coastal defense pressed existing materials and methods of cannon construction to the limit. This led to the development of methods for measuring pressures within the gun, which involved cylindrical punches mounted in holes drilled at right angles through the barrel. The pressure of the propellant gases forced the punches outward against soft copper plates, and the maximum pressure was then determined by calculating the amount of pressure needed to create an indentation of equal depth in the copper. The ability to measure pressures within a gun led to the design of cannon made thickest where internal pressures were greatest—that is, near the breech. The resultant "soda bottle" cannon of the mid- to

late 19th century, which had fat breeches curving down to short, slim muzzles, bore a strange resemblance to the very earliest European gun of which a depiction survives, that of the Walter de Millimete manuscript of 1327.

THE DEVELOPMENT OF ARTILLERY

The de Millimete gun

The earliest known gunpowder weapons vaguely resembled an old-fashioned soda bottle or a deep-throated mortar and pestle. The earliest such weapon, depicted in the English de Millimete manuscript, was some three feet long with a bore diameter of about two inches (five centimetres). The projectile resembled an arrow with a wrapping around the shaft, probably of leather, to provide a gas seal within the bore. Firing was apparently accomplished by applying a red-hot wire to a touchhole drilled through the top of the thickest part of the breech. The gun was laid horizontally on a trestle table without provision for adjusting elevation or absorbing recoil—a tribute to its modest power, which would have been only marginally greater than that of a large crossbow.

The breakthrough that led to the emergence of true cannon derived from three basic perceptions. The first was that gunpowder's propellant force could be used most effectively by confining it within a tubular barrel. This stemmed from an awareness that gunpowder's explosive energy did not act instantaneously upon the projectile but had to develop its force across time and space. The second perception was that methods of construction derived from cooperage could be used to construct tubular wrought-iron gun barrels. The third perception was that a spherical ball was the optimal projectile. The result was modern artillery.

Wrought-iron muzzle-loaders. The earliest guns were probably cast from brass or bronze. Bell-founding techniques would have sufficed to produce the desired shapes, but alloys of copper, tin, and zinc were expensive and, at first, not well adapted to the containment of high-temperature, high-velocity gases. Wrought iron solved both of these problems. Construction involved forming a number of longitudinal staves into a tube by beating them around a form called a mandrel and welding them together. (Alternatively, a single sheet of iron could be wrapped around the mandrel and then welded closed; this was particularly suitable for smaller pieces.) The tube was then reinforced with a number of rings or sleeves (in effect, hoops). These were forged with an inside diameter about the same as the outside of the tube, raised to red or white heat, and slid into place over the cooled tube, where they were held firmly in place by thermal contraction. The sleeves or rings were butted against one another and the gaps between them sealed by a second layer of hoops. Forging a strong, gastight breech presented a particular problem that was usually solved by welding a tapered breech plug between the staves.

Hoop-and-stave construction permitted the fabrication of guns far larger than had been made previously. By the last quarter of the 14th century, wrought-iron siege bombards were firing stone cannonballs of 450 pounds (200 kilograms) and more. These weapons were feasible only with projectiles of stone. Cast iron has more than two and a half times the density of marble or granite, and gunners quickly learned that a cast-iron cannonball with a charge of good corned powder behind it was unsafe in any gun large enough for serious siege work.

Wrought-iron breechloaders. Partly because of the difficulties of making a long, continuous barrel, and partly because of the relative ease of loading a powder charge into a short breechblock, gunsmiths soon learned to make cannon in which the barrel and powder chamber were separate. Since the charge and projectile were loaded into the rear of the barrel, these were called breechloaders. The breechblock was mated to the barrel by means of a recessed lip at the chamber mouth. Before firing, it was dropped into the stock and forced forward against the barrel by hammering a wedge into place behind it; after the weapon was fired, the wedge was knocked out and the block was removed for reloading. This scheme had significant advantages, particularly in the smaller classes of naval swivel guns and fortress wallpieces, where the use of multiple breechblocks permitted a high rate of fire. Small breechloaders continued to be used in these ways well into the 17th century.

Problems with early breech-loading

The essential deficiency of early breechloaders was the imperfect gas seal between breechblock and barrel, a problem that was not solved until the advent of the brass cartridge late in the 19th century. Hand-forging techniques could not produce a truly gastight seal, and combustion gases escaping through the inevitable crevices eroded the metal, causing safety problems. Wrought-iron cannon must have required constant maintenance and care, particularly in a saltwater environment.

Wrought-iron breechloaders were the first cannon to be produced in significant numbers. Their tactical viability was closely linked to the economics of cannonballs of cut stone, which, modern preconceptions to the contrary, were superior to cast-iron projectiles in many respects. Muzzle velocities of black-powder weapons were low, and smoothbore cannon were inherently inaccurate, so that denser projectiles of iron had no advantage in effective range. Cannon designed to fire a stone projectile were considerably lighter than those designed to fire an iron ball of the same weight; as a result, stone-throwing cannon were for many years cheaper. Also, because stone cannonballs were larger than iron ones of the same weight, they left larger holes after penetrating the target. The principal deficiency of stone-throwing cannon was the enormous amount of skilled labour required to cut a sphere of stone accurately to a predetermined diameter. The acceleration of the wage–price spiral in the 15th and 16th centuries made stone-throwing cannon obsolete in Europe.

Cast bronze muzzle-loaders. The advantages of cast bronze for constructing large and irregularly shaped objects of a single piece were well understood from sculpture and bell founding, but a number of problems had to be overcome before the material's plasticity could be applied to ordnance. Most important, alloys had to be developed that were strong enough to withstand the shock and internal pressures of firing without being too brittle. This was not simply a matter of finding the optimal proportions of copper and tin; bronze alloys used in cannon founding were prone to internal cavities and "sponginess," and foundry practices had to be developed to overcome the inherent deficiencies of the metal. The essential technical problems were solved by the first decades of the 15th century, and, by the 1420s and '30s, European cannon founders were casting bronze pieces that rivaled the largest of the wrought-iron bombards in size.

The importance of the trunnion

Developments in foundry practice were accompanied by improvements in weapon design. Most notable was the practice of casting cylindrical mounting lugs, called trunnions, integral with the barrel. Set just forward of the centre of gravity, trunnions provided the principal point for attaching the barrel to the carriage and a pivot for adjusting the vertical angle of the gun. This permitted the barrel to be adjusted in elevation by sliding a wedge, or quoin, beneath the breech. At first, trunnions were supplemented by lifting lugs cast atop the barrel at the centre of gravity; by the 16th century most European founders were casting these lugs in the shape of leaping dolphins, and a similarly shaped fixture was often cast on the breech of the gun.

Toward the end of the 15th century, French founders combined these features with efficient gun carriages for land use. French carriage design involved suspending the barrel from its trunnions between a pair of heavy wooden side pieces; an axle and two large wheels were then mounted forward of the trunnions, and the rear of the side pieces descended to the ground to serve as a trail. The trail was left on the ground during firing and absorbed the recoil of the gun, partly through sliding friction and partly by digging into the ground. Most important, the gun could be transported without dismounting the barrel by lifting the trail onto the limber, a two-wheeled mount that served as a pivoting front axle and point of attachment for the team of horses. This improved carriage, though heavy in its proportions, would have been familiar to a gunner of Napoleonic times. Sometime before the middle of the 16th century, English smiths developed a highly compact four-wheeled truck carriage for mounting trunnion-equipped

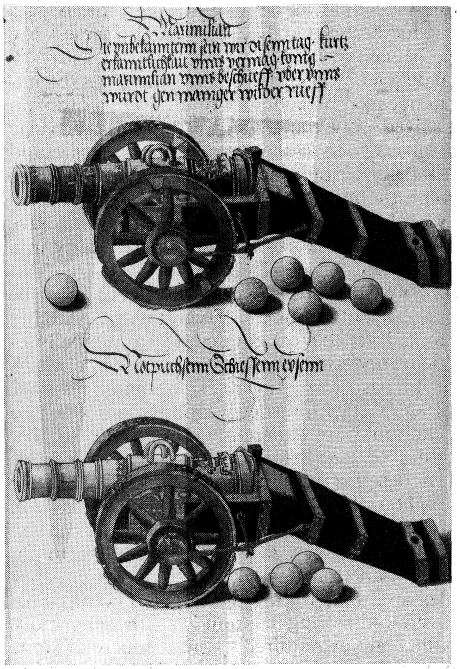


Figure 8: *Cast-bronze cannon of Maximilian I of Bavaria.*
Designed to smash fortress walls with a solid shot of iron, these large artillery pieces were lifted onto their wooden carriages by handles cast into the top of the barrel. Each carriage was carved from a single timber. From *Geschützbuch oder Zeugbuchs Kaiser Maximilian*, about 1500. In the Bayerische Staatsbibliothek, Munich (Cod. icon. 222, fol. 45v).
Bayerische Staatsbibliothek, Munich

shipboard ordnance, resulting in cannon that would have been familiar to a naval gunner of Horatio Nelson's day.

By the early 1500s, cannon founders throughout Europe had learned to manufacture good ordnance of cast bronze (see Figure 8). Cannon were cast in molds of vitrified clay, suspended vertically in a pit. Normally, they were cast breech down; this placed the molten metal at the breech under pressure, resulting in a denser and stronger alloy around the chamber, the most critical point. Subsequent changes in foundry practice were incremental and took effect gradually. As founders established mastery over bronze, cannon became shorter and lighter. In about 1750, advances in boring machines and cutting tools made it possible for advanced foundries to cast barrels as solid blanks and then bore them out. Until then cannon were cast hollow—that is, the bore was cast around a core suspended in the mold. Ensuring that the bore was precisely centred was a particularly critical part of the casting process, and small wrought-iron fixtures called chaplets were used to hold the core precisely in place. These were cast into the bronze and remained a part of the gun. Boring produced more accurate weapons and improved the quality of the bronze, since impurities in the molten metal, which gravitate toward the centre of the mold during solidification, were removed by the boring. But, while these changes were important operationally, they represented only marginal improvements to the same basic technology. A first-class bronze cannon of 1500 differed hardly at all in essential technology and ballistic performance from a cannon of 1850 designed to shoot a ball of the same weight. The modern gun would have been shorter and lighter, and it would have been mounted on a more efficient carriage, but it would have fired its ball no farther and no more accurately.

Enduring value of cast-bronze cannon

Cast-iron cannon. In 1543 an English parson, working on a royal commission from Henry VIII, perfected a method for casting reasonably safe, operationally efficient cannon of iron. The nature of the breakthrough in production technology is unclear, but it probably involved larger furnaces and a more efficient organization of resources. Cast-iron cannon were significantly heavier and bulkier than bronze guns firing the same weight of ball. Unlike bronze cannon, they were prone to internal corrosion. Moreover, when they failed, they did not tear and rupture like bronze guns but burst into fragments like a bomb. They possessed, however, the overwhelming advantage of costing only about one-third as much. This gave the English, who alone mastered the process until well into the 17th century, a significant commercial advantage by enabling them to arm large numbers of ships. The Mediterranean nations were unable to cast significant quantities of iron artillery until well into the 19th century.

Terminology and classification. Early gunpowder artillery was known by a bewildering variety of names. (The word cannon became dominant only gradually, and the modern use of the term to describe a gun large enough to fire an explosive shell did not emerge until the 20th century.) The earliest efficient wrought-iron cannon were called bombards or lombards, a term that continued in use well into the 16th century. The term basilisk, the name of a mythical dragonlike beast of withering gaze and flaming breath, was applied to early "long" cannon capable of firing cast-iron projectiles, but, early cannon terminology being anything but consistent, any particularly large and powerful cannon might be called a basilisk.

Founders had early adopted the practice of classifying cannon by the weight of the ball, so that, for example, a 12-pounder fired a 12-pound cannonball. By the 16th century, gunners had adopted the custom of describing the length of a cannon's bore in calibres, that is, in multiples of the bore diameter. These became basic tools of classification and remained so into the modern era with certain categories of ordnance such as large naval guns. Also by the 16th century, European usage had divided ordnance into three categories according to bore length and the type of projectile fired. The first category was the culverins, "long" guns with bores on the order of 30 calibres or more. The second was the cannons, or cannon-of-battery, named for their primary function of battering down fortress walls; these typically had barrels of 20 to 25 calibres. The third category of ordnance was the pedreros, stone-throwing guns with barrels of as little as eight to 10 calibres that were used in siege and naval warfare.

Mortars were a separate type of ordnance. With very wide bores of even fewer calibres than those of the pedreros, they were used in siege warfare for lobbing balls at a very high trajectory (over 45°). Mortars owed their name to the powder chamber of reduced diameter that was recessed into the breech; this made them similar in appearance to the mortars used to pulverize grain and chemicals by hand. Unlike the longer cannon, mortars were cast with trunnions at the breech and were elevated by placing wedges beneath the muzzle.

Special-purpose shot. Both culverins and cannon-of-battery generally fired cast-iron balls. When fired against masonry walls, heavy iron balls tended to pulverize stone and brick. Large stone cannonballs, on the other hand, were valued for the shock of their impact, which could bring down large pieces of wall. Undercutting the bottom of a wall with iron cannonballs, then using the heavy impact of large stone shot to bring it down, was a standard tactic of siege warfare. (Ottoman gunners were particularly noted for this approach.)

In the 15th century exploding shot was developed by filling hollow cast-iron balls with gunpowder and fitting a fuze that had to be lit just before firing. These ancestors of the modern exploding shell were extremely dangerous to handle, as they were known to explode prematurely or, with equally catastrophic results, jam in the gun barrel. For this reason they were used only in the short-bored mortars.

Exploding shot

For incendiary purposes, iron balls were heated red-hot in a fire before loading. (In that case, moist clay was sometimes packed atop the wadding that separated the ball from the powder charge.) Other projectiles developed for special purposes included the carcass, canister, grapeshot, chain shot, and bar shot. The carcass was a thin-walled shell containing incendiary materials. Rounds of canister and grapeshot consisted of numerous small missiles, usually iron or lead balls, held together in various ways for

simultaneous loading into the gun but designed to separate upon leaving the muzzle. Because they dispersed widely upon leaving the gun, the projectiles were especially effective at short range against massed troops. Bar shot and chain shot consisted of two heavy projectiles joined by a bar or a chain. Whirling in their trajectories, they were especially effective at sea in cutting the spars and rigging of sailing vessels.

Gunnery. During most of the black-powder era, with smoothbore cannon firing spherical projectiles, artillery fire was never precisely accurate at long ranges. (Aiming and firing were particularly difficult in naval gunnery, since the gunner had to predict the roll of the ship in order to hit the target.) Gunners aimed by sighting along the top of the barrel, or "by the line of metals," then stepped away before firing to avoid the recoil. The basic relationship between range and elevation being understood, some accuracy was introduced through the use of the gunner's quadrant, in which the angle of elevation of a gun barrel was measured by inserting one leg of the quadrant into the barrel and reading the angle marked on the scale by a vertically hanging plumb line. Nevertheless, the inherent inaccuracy of smoothbore artillery meant that most shooting was done at short ranges of 1,000 yards or less; at these ranges, estimating elevation by rule of thumb was sufficient. For attacking fortress walls, early modern gunners preferred a range of 60 to 80 yards; a range of 100 to 150 yards was acceptable, but 300 yards or more was considered excessive.

THE FIRST SMALL ARMS

Small arms did not exist as a distinct class of gunpowder weapon until the middle of the 15th century. Until then, hand cannon differed from their larger relatives only in size. They looked much the same, consisting of a barrel fastened to a simple wooden stock that was braced beneath the gunner's arm. A second person was required to fire the weapon. About the middle of the 15th century, a series of connected developments established small arms as an important and distinct category of weaponry. The first of these was the development of slow match—or match, as it was commonly called. This was cord or twine soaked in a solution of potassium nitrate and dried. When lit, match smoldered at the end in a slow, controlled manner. Slow match found immediate acceptance among artillerists and remained a standard part of the gunner's kit for the next four centuries.

The matchlock. Small arms appeared during the period 1460–80 with the development of mechanisms that applied match to hand-portable weapons. German gunsmiths apparently led the way. The first step was a simple S-shaped "trigger," called a serpentine, fastened to the side of a hand cannon's stock. The serpentine was pivoted in the middle and had a set of adjustable jaws, or dogs, on the upper end that held the smoldering end of a length of match. Pulling up on the bottom of the serpentine brought the tip of the match down into contact with powder in the flashpan, a small, saucer-shaped depression surrounding the touchhole atop the barrel. This arrangement made it possible for one gunner to aim and fire, and it was quickly improved on. The first and most basic change was the migration of the touchhole to the right side of the barrel, where it was served by a flashpan equipped with a hinged or pivoting cover that protected the priming powder from wind, rain, and rough handling. The serpentine was replaced by a mechanism, enclosed within the gunstock, that consisted of a trigger, an arm holding the match with its adjustable jaws at the end, a sear connecting trigger and arm, and a mechanical linkage opening the flashpan cover as the match descended. These constituted the matchlock, and they made possible modern small arms.

The first mechanical firing device

One final refinement was a spring that drove the arm holding the match downward into the pan when released by the sear. This mechanism, called the snap matchlock, was the forerunner of the flintlock. The fabrication of these devices fell to locksmiths, the only sizable body of craftsmen accustomed to constructing metal mechanisms with the necessary ruggedness and precision. They gave to the firing mechanism the enduring name lock.

The development of mechanical locks was accompanied by the evolution of gunstocks with proper grips and an enlarged butt to transmit the recoil to the user's body. The result was the matchlock harquebus, the dominant military small arm of the 15th century and the direct ancestor of the modern musket. The harquebus was at first butted to the breastbone, but, as the power of firearms increased, the advantages of absorbing the recoil on the shoulder came to be appreciated. The matchlock harquebus changed very little in its essentials until it was replaced by the flintlock musket in the final years of the 17th century.

The wheel lock. The principal difficulty with the matchlock mechanism was the need to keep a length of match constantly smoldering. German gunsmiths addressed themselves to this problem early in the 16th century. The result was the wheel lock mechanism, consisting of a serrated wheel rotated by a spring and a spring-loaded set of jaws that held a piece of iron pyrites against the wheel. Pulling the trigger caused the wheel to rotate, directing a shower of sparks into the flashpan. The wheel lock firearm could be carried in a holster and kept ready to fire indefinitely, but, being delicate and expensive, it did not spread beyond cavalry elites and had a limited impact on warfare as a whole.

The flintlock. Flintlock firing mechanisms were known by the middle of the 16th century, about a hundred years before they made their appearance in quantity in infantry muskets. A flintlock was similar to a wheel lock except that ignition came from a blow of flint against steel, with the sparks directed into the priming powder in the pan. This lock was an adaptation of the tinderbox used for starting fires.

In the several different types of flintlocks that were produced, the flint was always held in a small vise, called a cock, which described an arc around its pivot to strike the steel (generally called the frizzen) a glancing blow. A spring inside the lock was connected through a tumbler to the cock. The sear, a small piece of metal attached to the trigger, either engaged the tumbler inside the lock or protruded through the lock plate to make direct contact with the cock (see Figure 9).

Flintlocks were not as surefire as either the matchlock or the wheel lock, but they were cheaper than the latter,

From H. Peterson, *Arms and Armor in Colonial America*

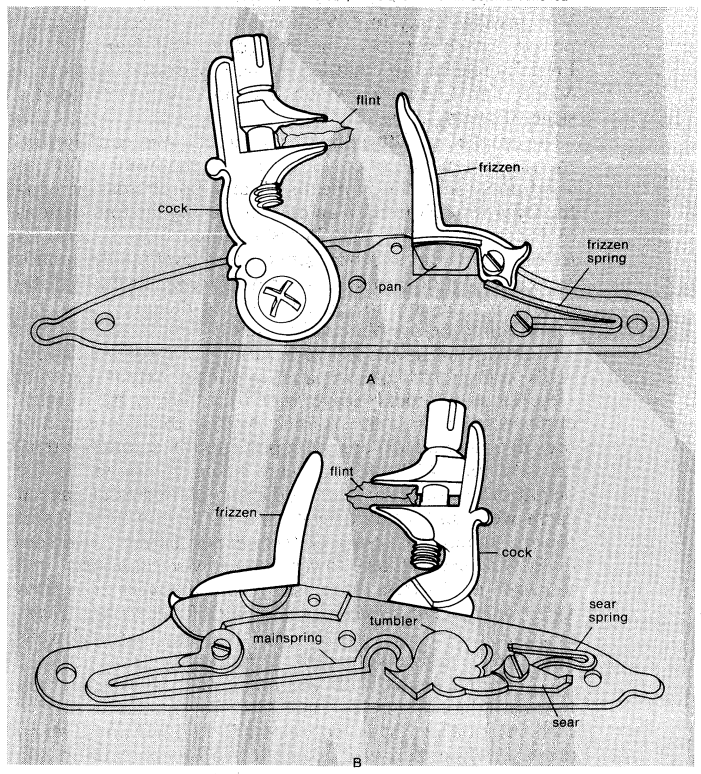


Figure 9: Typical flintlock mechanism, *c.* middle 17th century, with sear in safety position. Phantom view: (A) right side (outer), (B) left side (inner).

Tactical value of flint and steel

contained fewer delicate parts, and were not as difficult to repair in primitive surroundings. In common with the wheel locks they had the priceless advantage of being ready to fire immediately. A flintlock small arm was slightly faster to load than a matchlock, if the flint itself did not require adjustment.

FORTIFICATION

Before gunpowder artillery, a well-maintained stone castle, secured against escalade by high curtain walls and flanking towers, provided almost unbreachable security against attack. Artillery at first did little to change this. Large wrought-iron cannon capable of throwing wall-smashing balls of cut stone appeared toward the end of the 14th century, but they were neither efficient nor mobile. Indeed, the size and unwieldiness of early firearms and cannon suited them more for fortress arsenals than for the field, and adjustments to gunpowder by fortification engineers quickly tilted the balance of siege operations toward the defense. Gunports were cut low in walls for covering ditches with raking fire, reinforced platforms and towers were built to withstand the recoil shock of defensive cannon, and the special firing embrasures for crossbows were modified into gunports for hand cannon, with sophisticated vents to carry away the smoke. The name of the first truly effective small arm, the *hackenbüsche,* or hackbutt, is indicative: the weapon took its name, literally "hook gun," from a projection welded beneath the forward barrel that was hooked over the edge of a parapet in order to absorb the piece's recoil.

From medieval to modern. The inviolability of the medieval curtain wall came to an end in the 15th century, with the development of effective cast-bronze siege cannon. Many of the basic technical developments that led to the perfection of heavy bronze ordnance were pioneered by German founders. Frederick I, elector of Brandenburg from 1417 to 1425, used cannon systematically to defeat the castles of his rivals one by one in perhaps the earliest politically decisive application of gunpowder technology. The French and Ottomans were the first to bring siege artillery to bear in a decisive manner outside their own immediate regions. Charles VII of France (reigned 1422–61) used siege artillery to reduce English forts in the last stages of the Hundred Years' War. When his grandson Charles VIII invaded Italy in 1494, the impact of technically superior French artillery was immediate and dramatic; the French breached in eight hours the key frontier fortress of Monte San Giovanni, which had previously withstood a siege of seven years.

The fall of stone curtain walls

The impact of Ottoman siege artillery was equally dramatic. Sultan Mehmed II breached the walls of Constantinople in 1453 by means of large bombards, bringing the Byzantine Empire to an end and laying the foundations of Ottoman power. The Turks retained their superiority in siegecraft for another generation, leveling the major Venetian fortifications in southern Greece in 1499–1500 and marching unhindered through the Balkans before being repulsed before Vienna in 1529.

The shock of the sudden vulnerability of medieval curtain walls to French, Ottoman, and, to a lesser extent, German siege cannon quickly gave way to attempts by military engineers to redress the balance. At first, these consisted of the obvious and expensive expedients of counter-battery fire. By the 1470s, towers were being cut down to the height of the adjacent wall, and firing platforms of packed earth were built behind walls and in the lower stories of towers. Italian fortress architects experimented with specially designed artillery towers with low-set gunports sited to sweep the fortress ditch with fire; some were even sited to cover adjacent sections of wall with flanking fire. However, most of these fortresses still had high, vertical walls and were therefore vulnerable to battery.

A definitive break with the medieval past was marked by two Italian sieges. The first of these was the defense of Pisa in 1500 against a combined Florentine and French army. Finding their wall crumbling to French cannon fire, the Pisans in desperation constructed an earthen rampart behind the threatened sector. To their surprise and relief, they discovered not only that the sloping earthen rampart could be defended against escalade but that it was far more resistant to cannon shot than the vertical stone wall that it supplanted. The second siege was that of Padua in 1509. Entrusted with the defense of this Venetian city, a monk-engineer named Fra Giocondo cut down the city's medieval wall. He then surrounded the city with a broad ditch that could be swept by flanking fire from gunports set low in projections extending into the ditch. Finding that their cannon fire made little impression on these low ramparts, the French and allied besiegers made several bloody and fruitless assaults and then withdrew.

The sunken profile. While Pisa demonstrated the strength of earthen ramparts, Padua showed the power of a sunken profile supported by flanking fire in the ditch. With these two cities pointing the way, basic changes were undertaken in fortress design. Fortress walls, still essential for protection against escalade, were dropped into the ground behind a ditch and protected from battery by gradually sloping earthen ramparts beyond. A further refinement was the sloping of the glacis, or forward face of the ramparts, in such a manner that it could be swept by cannon and harquebus fire from the parapet behind the ditch. As a practical matter the scarp, or main fortress wall, now protected from artillery fire by the glacis, was faced with brick or stone for ease of maintenance; the facing wall on the forward side of the ditch, called the counterscarp, was similarly faced. Next, a level, sunken space behind the glacis, the covered way, was provided so that defenders could assemble for a sortie under cover and out of sight of the attackers. This, and the provision of firing embrasures for cannon in the parapet wall, completed the basics of the new fortress profile (see Figure 10).

Protecting the wall from cannon fire

Refinements of the basic sunken design included a palisade of sharpened wooden stakes either in the ditch or immediately behind the glacis and a sunken, level path behind the parapet for ammunition carts, artillery reinforcements, and relief troops. As attacking and defending batteries became larger, fortress designers placed greater emphasis on outworks intended to push the besieging batteries farther back and out of range.

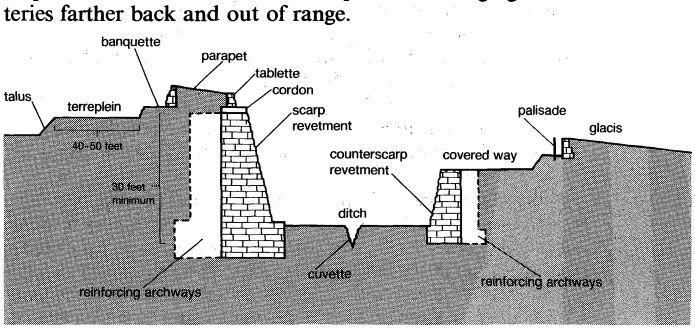


Figure 10: *Profile of the European fortress wall from the 16th century.*
The fortress wall, called the scarp, was sunk into the ground behind a ditch. A sloping earthen rampart, called the glacis, deflected cannonballs and gave defenders a free field of fire.

The profile of the outworks was designed according to the same basic principles applied to the fortress. Well established by 1520, these principles remained essentially unchanged until rifled artillery transformed positional warfare in the mid-19th century.

The bastioned trace. The sunken profile was only half the story of early modern fortress design; the other half was the trace, the outline of the fortress as viewed from above. The new science of trace design was based, in its early stages, on the bastion, a projection from the main fortress wall from which defending fire could sweep the face of adjacent bastions and the wall between. Actually, bastions had been introduced before engineers were fully aware of the power of artillery, so that some early 16th-century Italian fortifications combined sophisticated bastioned traces with outmoded high walls, a shallow ditch, and little or no protective glacis. After early experimentation with rounded contours, which were believed to be stronger, designers came to appreciate the advantages of bastions with polygonal shapes, which eliminated the dead space at the foot of circular towers and provided

uninterrupted fields of view and fire. Another benefit of the polygonal bastion's long, straight sections of wall was that larger defensive batteries could be mounted along the parapets.

Pushing back siege artillery

The relatively simple traces of the early Italian bastioned fortresses proved vulnerable to the ever larger armies and ever more powerful siege trains of the 16th century. In response, outworks were developed, such as ravelins (detached outworks in front of the bastions) and demilunes (semidetached outworks in the ditch between bastions), to shield the main fortess walls from direct battery. The increasing scale of warfare and the greater resources available to the besieger accelerated this development, and systems of outworks grew more and more elaborate and sprawling as a means of slowing the attacker's progress and making it more costly.

By the late 17th century, fortress profiles and traces were closely integrated with one another and with the ground on which they stood. The sophistication of their designs is frequently linked with the name of the French military engineer Sébastien Le Prestre de Vauban.

Duration of early modern fortification. With various refinements, the early modern fortress, based on a combination of the sunken profile and bastioned trace, remained the basic form of permanent fortification until the American Civil War, which saw the first extensive use of heavy rifled cannon made of high-quality cast iron. These guns not only had several times the effective range and accuracy of their predecessors, but they were also capable of firing explosive shells. They did to the early modern fortress what cast-bronze cannon had done to the medieval curtain wall. In 1862 the reduction by rifled Union artillery of Fort Pulaski, a supposedly impregnable Confederate fortification defending Savannah, Ga., marked the beginning of a new chapter in the design of permanent fortifications. (J.F.G.)

MODERN WEAPONS AND WEAPON SYSTEMS

Small arms

Since the introduction of the flintlock musket in the 17th century, military small arms have gone through a series of significant changes. By employing different projectiles and successively improved chemical propellants, the dual goal of most arms designers has been the creation of man-portable weapons of greater firepower and reduced weight. But the attainment of this goal has continually been hampered by an inescapable physical relationship between the recoil forces generated by gunpowder weapons and the mass and velocity of their projectiles. In order to reduce the weight of a weapon, its recoil energy has to be reduced, but reducing recoil also affects the killing power of the bullet. Given the constraints of this relationship, military small arms may well have reached a level where, within reasonable economic limits, significantly higher performance cannot be obtained merely by improving existing gunpowder-based technology.

SHOULDER WEAPONS

Muzzle-loaders. Practical shoulder-fired small arms started with the perfection of the flintlock ignition system in the mid-17th century (see above *Military technology before the modern era: The gunpowder revolution, c. 1300–1650*). Earlier gunpowder small arms, based on the matchlock or wheel lock mechanisms, were generally too heavy, too unreliable, or too expensive to allow for general issue to infantry forces. Indeed, the first matchlock *mosquetes* ("muskets") fielded by Spanish infantry weighed as much as 25 pounds (10 kilograms) and usually required a forked staff as a rest to enable a man of normal strength to fire them accurately from the shoulder. Nevertheless, they were capable of sending bullets through the best armour that could be worn by a mobile soldier. Almost overnight, firepower from muskets became the dominant force in war, and fully armoured soldiers almost disappeared from European battlefields toward the end of the 16th century. With armour-piercing power no longer necessary, muskets could be made smaller, and shoulder weapons without rests became the norm.

Early muskets

The introduction of new ignition systems did not immediately render older forms obsolete, however; all systems, in many variations, existed side by side. Wheel locks and matchlocks, for example, persisted into the 18th century, long after flintlocks had established their primacy in Europe and America.

Standardized patterns and parts. Flintlock small arms emerged at the start of industrialization, with weapons production becoming one of the first industrial sectors to exploit the transition from craft production to the large-scale production of the Industrial Revolution. On the military side, these weapons entered service at a time when the scale of ground forces employed in battle was increasing. The ability to manufacture large numbers of muskets enabled military leaders to equip these mass armies.

By the 1600s European military authorities had begun moving toward greater uniformity in order to eliminate mixed inventories of nonstandard weapons. England took the first steps toward creating a national system of small-arms manufacture. For years, completed muskets had been purchased from a variety of English, Irish, and Dutch gunmakers, who subcontracted for components and arranged for final assembly. Beginning in the early 1700s, ordnance officials, from their headquarters at the Tower of London, divided the manufacture of firearms into locks, stocks, barrels, ramrods, and furniture—all of which they sought to purchase directly from subcontractors. Since different components for the same weapon were made in different locations, Tower officials oversaw the establishment of "Sealed Patterns" (sample firearms) to serve as exact models for gunmakers.

An Ordnance Office decree of 1722 led to a standard army musket, called the "Long Land," which had a 46-inch (1,168-millimetre) barrel and a calibre, or bore diameter, of .75 inch (19 millimetres). The Long Land became popularly known in America as the first model Brown Bess musket. Fighting experience in the wilderness of North America during the Seven Years' War, or French and Indian War (1756–63), suggested the utility of lighter and shorter muskets, and in 1768 the Short Land musket, with a 42-inch barrel, became standard. Known as the second model Brown Bess, the Short Land became one of the basic weapons used in the American Revolution (1775–83). It was succeeded in 1797 by the "India Pattern," with a 39-inch barrel. During the wars with Napoleon from 1804 to 1815, more than 1.6 million of these muskets were assembled in Birmingham, and nearly 2.7 million muskets of all types were "fitted up" in London and at the Lewisham Royal Armoury Mills. In 1816 assembly work was divided between London and a new Royal Small Arms Factory at Enfield Lock, Middlesex.

The Short Land, or Brown Bess

In France, standard-pattern muskets did not exist prior to 1717, when the government specified a weapon with a 47-inch barrel and a calibre of .69 inch. (This calibre remained standard until the 19th century.) After the Seven Years' War, the French army introduced the Modèle 1763, with a stronger lock and shorter (45-inch) barrel—a length that remained standard to century's end. The Modèle 1777 musket represented a major step forward because of improved production techniques, with the French creating a rigorous system of patterns and gauges that yielded muskets with nearly interchangeable parts. This process was intended to produce less expensive muskets that were easier to make and repair, but worker resistance delayed large-scale manufacture of small arms using interchangeable parts until the early 1800s. Had the program succeeded earlier, France would have been better equipped to fight the Napoleonic Wars. As it was, French firms in such provincial cities as Charleville, Maubeuge, Saint-Étienne, and Tulle fabricated fewer than two million small arms.

The U.S. government created national armouries at Springfield, Mass., and at Harpers Ferry, Va., in 1794; work at Springfield commenced in 1795, and arms production began at Harpers Ferry in 1801. Both built an Americanized version of the French Modèle 1777 musket (known as the Model 1795 in the United States). These armouries and their private competitors later became important centres of technological innovation. With the adoption of the .69-inch Model 1842, the U.S. military introduced the large-scale assembly of weapons from uniform, interchangeable parts. By the mid-1850s arms makers around the world were beginning to copy this "American System" of manufacture, which contributed to the creation of the modern military small arm—especially after the introduction of percussion ignition and rifled barrels.

Percussion ignition. The Model 1842 was based on the Model 1840 flintlock, but it featured a switch to percussion ignition. This newer system was based on the explosive property of potassium chlorate and fulminate of mercury, both of which detonate when struck a small, sharp blow by a striker. Several Germans experimented with detonating fulminates in the late 17th century, and the French did likewise in the 18th century, but it was Alexander John Forsyth, a Scottish clergyman, who successfully wedded priming powders to the ignition of firearms in 1805, receiving a patent in April 1807. Forsyth invented the "scent bottle" type of lock mechanism, so called because rotating on a tapered steel plug at about the location of a flintlock touchhole was a powder-filled container that looked like a perfume bottle. Turning the bottle upside down released some detonator powder into a cavity at the top of the plug, and turning the bottle back left the striker mechanism, consisting of a hammer rather than the cock and jaws of the flintlock, free to operate. When the trigger was pulled, the hammer fell, detonating the compound.

Subsequent inventors simplified the percussion lock mechanism by using loose or pellet detonating powder. By 1830, percussion caps (attributed to the Philadelphian Joshua Shaw in 1815) were becoming the accepted system for igniting firearm powder charges. A percussion cap was a truncated cone of metal (preferably copper) that contained a small amount of fulminate of mercury inside its crown, protected by foil and shellac. This cap was fitted onto a steel nipple mounted at the weapon's breech, and a small channel in the nipple communicated the flash from the cap to the powder chamber. In the final form of this mechanism, a hollow-nosed percussion hammer came down over the percussion cap, thus eliminating the danger of flying copper when the powder detonated. Percussion cap ignition was easily adapted to existing flintlock muskets and pistols.

Rifling. As killing machines, smoothbore infantry muskets were relatively inefficient. Their heavy, round lead balls delivered bone-crushing and tissue-destroying blows when they hit a human body, but beyond 75 yards even trained infantrymen found it difficult to hit an individual adversary. Volley fire against massed troops delivered effective projectiles out to 200 yards, but at 300 yards balls from muzzle-loaders lost most of their lethality. Also, while well-trained soldiers could load and shoot their muskets five times per minute, volley fire led to a collective rate of only two to three shots per minute.

Shortcomings of smoothbore muskets

These ballistic shortcomings were a product of the requirement that the projectile, in order to be quickly rammed from muzzle to breech, had to fit loosely in the barrel. When discharged, it wobbled down the barrel, contributing to erratic flight after it left the muzzle. Rifled barrels, in which spiral grooves were cut into the bore, were known to improve accuracy by imparting a gyroscopic spin to the projectile, but reloading rifled weapons was slowed because the lead ball had to be driven into the barrel's rifling. Greased cloth or leather patches eased the problem somewhat, but the rate of fire of rifles was still much lower than that of smoothbore muskets.

One possible solution was the creation of mechanisms that allowed the bullet to be loaded at the breech instead of the muzzle. Many such ideas were tested during the 18th century, but, given the craftsman-based manufacture of the day, none was suited to large-scale production. Special army units in Europe and America used rifled muzzle-loaders, such as the flintlock British Baker rifle, to harass the enemy at long ranges, while most infantrymen continued to carry muzzle-loading smoothbores. For this reason, inventors concentrated on adapting rifled barrels to muzzle-loaders. In 1826 Henri-Gustave Delvigne of France, seeking a means of expanding the projectile without making it difficult to ram home, created a narrow powder chamber at the breech end of the barrel against which a loosely fitting lead ball came to rest. Ramrod blows expanded the soft lead at the mouth of the chamber so that, when fired, the bullet fit the rifling tightly. In 1844 another French officer, Louis-Étienne de Thouvenin, introduced yet a better method for expanding bullets. His *carabine à tige* embodied a post or pillar (*tige*) at the breech against which the bullet was expanded.

These rifles worked better than earlier types, but their deformed balls flew with reduced accuracy. Captain Claude-Étienne Minié, inspired by Delvigne's later work with cylindrical bullets, designed longer, smaller-diameter projectiles, which, having the same weight as larger round balls, possessed greater cross-sectional density and therefore retained their velocity better. Moreover, while the flat base of Minié's projectile was deformed against the pillar as in Thouvenin's weapon, the rest of the bullet maintained its shape and accuracy. The French army combined these ideas in the Carabine Modèle 1846 à tige and the Fusil d'Infanterie Modèle 1848 à tige.

In order to combat the tendency of muzzle-loading rifles to become difficult to load as gunpowder residue collected in the barrels, Minié suggested a major simplification—eliminating the pillar and employing in its place a hollow-based bullet with an iron expander plug that caused the projectile to engage the rifling when the weapon was fired. This new projectile could be loaded into dirty rifles with ease, and, because it was not deformed while loading, it had greater accuracy.

Officials in several countries, notably Britain and the United States, saw the significance of Minié's invention. In 1851 the Royal Small Arms Factory, Enfield, embarked upon production of the .702-inch Pattern 1851 Minié rifle (see Figure 11). In the Crimean War (1854–56), Russian troops armed with smoothbore muskets were no match for Britons shooting P/51 rifles. Massed formations were easy prey, as were cavalry and artillery units. A correspondent for the *Times* of London wrote: "The Minié is king of weapons . . . the volleys of the Minié cleft [Russian soldiers] like the hand of the Destroying Angel."

The Enfield Pattern 1851

Swiss experiments demonstrated that an expander plug was not necessary when a bullet's side walls were thin enough, and the British designed a smaller-calibre rifle using this type of Minié bullet. The result was a .577-inch weapon firing "cylindro-conoidal" projectiles—essentially a lead cylinder with a conical nose. "Enfield" as a weapon name was first generally applied to these Pattern 1853 rifles. Subsequent tests indicated that rifles with 33-inch barrels could provide accuracy equal to the 39-inch P/53 barrels. When the resulting P/53 Short Rifles were issued, there began a century-long trend toward shorter weapons.

In the United States, experiments undertaken in the late 1840s led to the adoption of a .58-inch Minié-type bullet and a family of arms designed to use it. The Model 1855 rifled musket, with a 40-inch barrel, produced a muzzle velocity of 950 feet (290 metres) per second. All Model 1855 weapons used mechanically operated tape priming, intended to eliminate the manual placement of percussion caps on the nipple, but this system proved too fragile and was eliminated with the introduction of a simplified Model 1861 rifled musket. During the American Civil War (1861–65), the Union government purchased both Model 1861 and Model 1863 rifled muskets as its basic infantry weapon. In the Confederacy, domestic production was supplemented by purchases of Enfield P/53 rifles and other European weapons.

The Civil War clearly demonstrated the deadly effect of rifled muskets, although many battlefield commanders only slowly appreciated the changing nature of warfare. Individual soldiers could hit their opposing numbers with

accurate fire out to 250 yards, so that frontal assaults, in which soldiers advanced in neat ranks across open fields, had to be abandoned. By 1862 both sides were building field entrenchments and barricades to provide protection from rifle and artillery fire.

Breechloaders. *The bolt action.* The American Civil War also previewed the importance of breech-loading rifles. For more than a century, soldiers carrying muzzle-loaders had been issued paper cartridges containing the musket ball and an appropriate powder charge. To use one of these cartridges, they simply bit off the end of the paper tube, poured a little powder into the pan (if the gun was a flintlock), dumped the rest down the barrel, and then rammed the ball and paper down on top. Some early breechloaders used slightly improved cartridges of nitrate-soaked paper or linen that contained the powder and ball and were inserted into the opened breech as a unit. The powder was set off when sparks from the flashpan ignited either the flammable case itself or exposed powder at the end of the cartridge. Other breechloaders employed metal cartridges that were pierced with holes or made with ends of flammable paper, so that the powder could be ignited by a percussion cap. But all of these systems, which relied upon externally mounted flintlock or percussion ignition mechanisms, were prone to misfiring, and they did little to prevent the leakage of gas and flame for which breechloaders were notorious. Breech-loading rifles became practical only with the design of cartridges that housed the primer as well as the propellant in a single case, and that provided an effective seal when the weapon was fired.

Early breech loading

The first such cartridge to be successfully employed in war was of the rimfire type, in which a ring of detonating fulminate was deposited in a hollow rim around the base of a thin copper case. An external hammer crushed the rim in one spot, firing the round. Unfortunately, some fulminate compounds detonated unpredictably, leading to both misfires and premature explosions. Also, a cartridge case that was soft enough to be crushed by a striker could not stand up to the heavy propellant charge necessary for a full-power infantry rifle. For this reason, rimfire cartridges were used most effectively in pistols or—during the American Civil War—in smaller repeating carbines such as the .56-inch Spencer and the .44-inch Henry.

In Europe, a milestone in the development of breech-loading infantry weapons was achieved by Johann Nikolaus Dreyse, a Prussian. His *Zündnadelgewehr* ("needle-fired gun"), introduced in 1838, used a paper cartridge with a priming pellet located at the base of a solid egg-shaped bullet. A long, needle-shaped firing pin, shot forward by a spring, pierced the cartridge and powder charge to detonate the primer. This needle was housed in a steel cylinder called the bolt, which slid forward in the frame of the receiver until it was locked firmly against the base of the cartridge in the chamber. Once the weapon was fired, the soldier released a latch with his thumb, grasped a knob at the end of a handle projecting from the bolt, turned it until locking lugs on the bolt were disengaged from slots in the receiver, and slid the bolt back to open the chamber for reloading. This bolt action, simple in concept and yet requiring precise workmanship, constituted a revolution in small-arms design.

The first Dreyse rifles were adopted by the Prussian army in 1843 and were used in campaigns in 1849 and 1864. In 1866, notably at the Battle of Königgrätz during the Seven Weeks' War, Prussian soldiers lying prone were able to fire six shots from their 15.43-millimetre (.607-inch) Zündnadelgewehr Modell 1862 for every one discharged from their Austrian opponents' muzzle-loading rifles.

Prussia's success encouraged other European states to adopt bolt-action breechloaders. The French employed Antoine-Alphonse Chassepot's 11-millimetre Fusil d'Infanterie Modèle 1866 to devastating effect in such battles of the Franco-German War (1870–71) as Mars-la-Tour and Gravelotte. Close-order troop formations disappeared from the European scene as a result of these fights, and the cavalry charge was relegated to the past. The Chassepot rifle employed a shorter firing pin than the Dreyse, because its cartridge was fitted with a detonating cap at the very base. About 1.03 million of these weapons were in hand when the war began, and Prussia had some 1.15 million Dreyse needle rifles—a quantity that demonstrated the value of machine production of weapons with interchangeable parts.

The French Chassepot

Needle rifles offered a faster rate of fire, but their paper cartridges provided a poor seal at the breech, and their long firing pins warped or broke under heavy use. One solution was the metallic centre-fire cartridge with a percussion cap centred in the base of a hard brass or copper case. A shorter, sturdier firing pin was sufficient to detonate the primer, and a metallic case that was strong enough to withstand a powerful propellant charge also provided effective closure of the breech. Adopting centre-fire cartridges, France transformed its Chassepots into the 11-millimetre Modèle 1866/67 and 1874 rifles, which were named after their designer, Basile Gras. Germany went to rifles designed by Peter Paul Mauser, first the 11-millimetre Modell 1871 Gewehr and then the Modell 1871/84 Infanterie-Repetier-Gewehr. The latter was a 10-shot repeater that ejected the spent case as the bolt was pulled back and fed a fresh cartridge into the chamber from a tubular magazine beneath the barrel as the bolt was pushed forward.

All other European countries soon adopted cartridge breech-loading rifles, usually by converting existing muzzle-loaders and then by purchasing purpose-built breechloaders. Many did not feature bolt action. For example, beginning in 1866, Britain converted its P/53 Enfields simply by hinging the top of the breech so that it could be opened sideways, the spent case extracted, and a fresh cartridge inserted. In 1871 the British went to new Martini-Henry breechloaders of .45-inch calibre. In these rifles, pushing down a lever attached to the trigger guard lowered the entire breechblock, exposing the chamber, and raised the breechblock back to firing position when it was pulled back. Russia adopted two new 10-millimetre breechloaders, the Model 1868 Berdan No. 1 and then the bolt-action Model 1870 Berdan No. 2, both of which were largely the work of American Civil War officer Hiram Berdan. The U.S.-made Remington Rolling Block Rifle, in which the breechblock was cocked back on a hinge like the hammer, was bought by a number of countries around the world. The United States itself adopted a series of single-shot rifles employing a hinged-breech "trap-door" mechanism, developed by Erskine S. Allin at the Springfield Armory, in which the top of the breech was flipped forward along the top of the barrel. The first Model 1866 was a converted .58-inch musket, the second Model 1866 was a new rifle in .50-inch calibre, and subsequent versions were built in .45-inch calibre. These weapons, born of postwar starvation budgets, continued to use components introduced with the Model 1855 muzzle-loaders.

The smokeless-powder revolution. All early breechloaders used black powder as their source of propellant energy, but in the early 1880s more powerful and cleaner-burning nitrocellulose-based propellants were perfected. Whereas black powder produced a large quantity of solid material upon combustion, quickly fouling barrels and pouring out huge clouds of smoke, nitrocellulose produced mostly gas and was therefore labeled "smokeless powder." Also, it produced three times the energy of black powder and burned at a more controllable rate. Such characteristics made possible a shift to longer and smaller-diameter projectiles. Bore diameters were again reduced, this time to calibres of about .30 inch, or 7.5 to 8 millimetres. Muzzle velocities ranged from 2,000 to 2,800 feet per second, and accurate range extended to 1,000 yards and beyond. Because lead projectiles were too soft to be used at such increased power and velocity, they were sheathed in harder metal. In 1881 a Swiss officer, Eduard Alexander Rubin, was the first to perfect a full-length, copper-jacketed bullet.

Problems with black powder

Magazine repeaters. France was the first country to issue a small-bore, high-velocity repeating rifle, the Modèle 1886 Lebel, which fired an 8-millimetre, smokeless-powder round. The tubular magazine of this rifle soon became obsolete, however. In 1885 Ferdinand Mannlicher of Austria had introduced a boxlike magazine fitted into the bottom of the rifle in front of the trigger guard. This magazine was easily loaded by a device called a clip, a

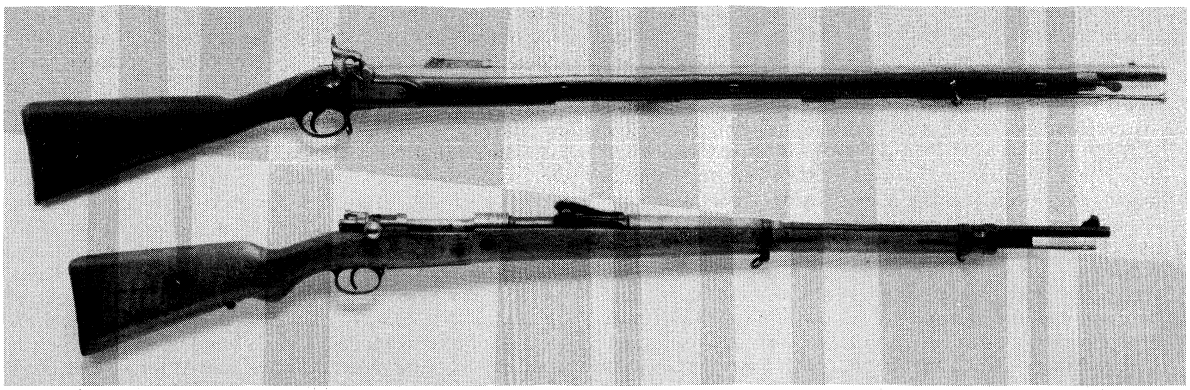


Figure 11: *Two manually operated infantry rifles of the 19th century.*
(Top) The British Enfield Pattern 1851, a percussion-ignition, Minié-type muzzle-loader.
(Bottom) The German 1898 Mauser, a bolt-action, magazine-fed repeater.
T. Jimbo

light metal openwork box that held five cartridges and fed them up into the chamber through the action of a spring as each spent case was ejected. Other magazine rifles, such as the Mauser, used a different loading device, called a charger. This was simply a flat strip of metal with its edges curled to hook over the rims or grooves of a row of cartridges (also usually five). To load his rifle, a soldier drew back the bolt, slipped the charger into position above the opened receiver, and pushed the cartridges down into the magazine, where they were held in tension against a spring. The efficiency of the box magazine was quickly recognized, as was its special compatibility with the bolt action, and all European states made the conversion. For example, Germany adopted the 8-millimetre Model 1888 Commission rifle, Belgium the 7.65-millimetre Model 1889 Mauser, Turkey the Model 1890 Mauser, and Russia the 7.62-millimetre Model 1891 Mosin-Nagant. In 1892 Britain abandoned movable-block action and went to the .303-inch, bolt-action Lee-Metford, and the United States began to purchase the .30-inch Model 1892 Krag-Jørgensen, a Danish design. In 1906 Japan adopted the 6.5-millimetre Year 38 Arisaka rifle.

By World War I (1914–18) all major powers adopted smokeless-powder, bolt-action, magazine-fed repeating rifles, and some had shifted to a second generation. Austria, for example, issued the Modell 1895 Mannlicher, firing an 8-millimetre round, and German troops carried the 7.92-millimetre Modell 1898, designed by Mauser (see Figure 11). For durability, safety, and efficiency, the 1898 Mauser was probably the epitome of bolt-action military rifles. It was sold and copied around the world. In the United States the Mauser was only slightly altered and issued as the .30-inch M1903 Springfield.

The 1898 Mauser

Also following Germany's lead in the design of ammunition, all armies replaced their blunt-nosed projectiles with aerodynamically superior pointed bullets (in German, *Spitzgeschossen*). Barrel lengths continued to decrease, partly in response to more efficient propellants and partly to make rifles easier to use in the field. The British .303-inch Short, Magazine, Lee-Enfield rifle, known as the SMLE, had a 25-inch barrel, while the M1903 Springfield's barrel measured just over 23.75 inches.

During the Great War, huge quantities of rifles were built. British factories made more than 3.9 million rifles, German sources produced about 5 million, and Russian factories built more than 9 million. Still, most armies suffered from shortages. Factories in the United States made 1.24 million rifles for the British and 280,000 for the Russians; for U.S. forces they produced 2.4 million between May 1917 and December 1918 alone.

Automatic weapons. *The self-loading rifle.* Magazine-fed rifles provided a radical increase in rate of fire. Indeed, by 1914 many British riflemen could fire 15 aimed shots per minute, and some very skillful individuals could exceed 30 shots per minute. Nevertheless, in order to transcend the limits imposed by manual operation, gun designers such as Mannlicher and the American Hiram Maxim came up with experimental self-loading, or semiautomatic, rifles, which used the energy generated by a fired round to load a fresh round into the chamber. However, only a handful of these weapons were adopted in very small numbers by the major armies, whose interest in automatic fire from the 1880s through World War I was directed primarily toward heavier infantry-support weapons (see below *Machine guns*).

After the war, all nations having an arms industry sought to produce a semiautomatic rifle, but only the United States was successful in developing and manufacturing a battle-worthy weapon. Adopted in 1936, the U.S. Rifle, Caliber .30 M1, designed by John C. Garand, was a technological tour de force (see Figure 12). A small hole or gas port on the underside of its barrel near the muzzle directed part of the propellant gases into a small cylinder holding a piston that was connected to the bolt. As gas pressure forced back the piston and bolt, the empty cartridge case was ejected and the hammer was cocked. A spring then forced the bolt forward. As it moved forward, the bolt stripped the top cartridge from an eight-round, clip-loaded magazine within the receiver and seated it in the chamber, ready to fire. Gas pressure thus performed automatically the reloading task formerly done by hand.

As the only semiautomatic rifle to become a standard-issue infantry weapon, the M1 was extremely durable and reliable in combat. Between 1937 and 1945, the Springfield Armory and the Winchester Repeating Arms Company produced 4.04 million of these rifles. Still, the infantry units of most other belligerents during World War II (1939–45) were armed with bolt-action rifles of the World War I era as their standard weapons.

Obsolescence of the bolt-action rifle

The submachine gun. The ballistic performance of infantry rifles was tailored to the long-range requirements of a bygone era when foot soldiers demanded weapons that could reach out to halt the dreaded cavalry charge. Beginning early in World War I, however, battlefields became no-man's-lands pockmarked by shell craters and crisscrossed by miles of barbed-wire entanglements, and machine guns dominated the 1,000 or more yards between trench lines. While rifles were shot at those extreme ranges, they could not equal the destructive power of artillery and machine guns, and they were too cumbersome and powerful for offensive assaults on enemy trenches. A generation later, in World War II, the greater mobility of troops accompanying armoured vehicles reinforced the need for lighter, more portable weapons of improved effectiveness at close quarters.

Such changing conditions led to experiments with automatic weapons firing rounds of lower velocity or lighter weight. One result, which saw its first use in World War I, was a new weapon called the machine carbine or submachine gun. Derived from the semiautomatic pistol and firing pistol-calibre ammunition with muzzle velocities of only about 1,000 feet per second, submachine guns were fitted with shoulder stocks (and sometimes forward hand grips). Such weapons offered easier handling than rifles while providing greater accuracy and more rapid fire than most handguns.

Figure 12: *Two autoloading infantry rifles of the 20th century.*
(Top) The U.S. M1, a gas-operated semiautomatic. (Bottom) The Soviet AK-47, a gas-operated assault rifle.
T. Jimbo

The first successful weapon of this type was the Maschinen Pistole 1918 Bergmann, designed by Hugo Schmeisser and employed by the Germans during the last few months of the war. The barrel of the MP18 was less than eight inches long, and it was chambered for nine-millimetre rounds introduced in 1908 for Parabellum, or Luger, pistols. It operated under a principle called blowback, in which the spent cartridge case, blown backward out of the chamber by the gases generated by the firing of the round, forced the bolt back against a spring and tripped the mechanism that ejected the case from the gun. The spring then forced the bolt forward as a fresh cartridge was fed into the chamber. If the trigger was kept depressed, the new round would be fired automatically, and the cycle would continue until the trigger was released or the ammunition was exhausted. In blowback actions, the bolt had to be quite heavy, or it had to be subjected to various devices that retarded its backward motion, in order to keep the mechanism from operating faster than was desired. In the MP18, a heavy bolt and spring limited the weapon's rate of fire to about 400 rounds per minute.

After the war, Vasily Degtyarev of the Soviet Union incorporated Schmeisser's principles into his own designs, culminating in the Pistolet Pulemyot Degtyarova of 1940. The PPD was fed by a drum-shaped magazine containing 71 7.62-millimetre cartridges, and it fired at a rate of 900 rounds per minute—far too fast for accuracy. In the United States, John T. Thompson's submachine gun, chambered for the .45-inch Colt pistol cartridge, was adopted by the army in 1928. Popularly called the "tommy gun," the M1928 was effective, but its blowback operation was modified by a complex retarding mechanism that was deleted from later versions, when its large drum magazine was also replaced by a box magazine.

Sheet-metal fabrication

Under the pressures of World War II, the major powers used millions of submachine guns. These included a second generation of simplified weapons that, being fabricated partly from sheet-metal stampings, could be produced in quantity almost anywhere and at little expense. The Germans led the way with the MP38 and MP40. Known to the Allies as "burp guns," these weapons operated at 450 to 550 rounds per minute, the optimal rate for controlled fire. Also, they were fed by a box magazine, which did not jam as often as a drum, and had a wire shoulder stock that could be folded against the receiver. Meanwhile, the Soviets issued en masse the PPSh of 1941 and the PPS of 1943. The latter closely resembled the new German guns, as did the United States' M3, called the "grease gun" for its resemblance to a mechanic's grease dispenser. The British Sten gun, extremely simple and inexpensive yet very effective, was issued to paratroops and commandos beginning in 1941 and was also smuggled to partisans in Europe.

After the war, almost all new submachine guns, such as the British Sterling and West German MP5, were chambered for nine-millimetre cartridges. As a class of weapon, they received a new lease on life with the telescoping bolt, pioneered by Václav Holek in the Czechoslovak Model 23 of 1948. This involved a hollowed-out bolt that slid partially over the barrel when a round was chambered, resulting in a much shorter weapon. A prominent example of this type was the Israeli Uzi, designed by Uziel Gal, which was only 25 inches long with its shoulder stock extended. The Uzi was adopted around the world as a police and counterterrorist weapon. Indeed, aside from arming special forces, the submachine gun lost importance as a military weapon. With an effective range limited to about 200 yards, it could not fill the broad gap between the low-power pistol cartridge and the full-power rifle cartridge. This gap, which constituted the ground upon which modern infantrymen found themselves fighting, had to be filled by another new weapon, which would fire a cartridge of intermediate power.

The assault rifle. A hint at this new weapon had been given during World War I, when Vladimir Grigorevich Fyodorov, father of Russian automatic weapons, married the 6.5-millimetre cartridge of the Japanese Arisaka rifle to an automatic rifle. In 1916 he unveiled his new weapon, the Avtomat Fyodorova. Owing to the turmoil of the Russian Revolution of 1917, only about 3,200 of Fyodorov's weapons were delivered. Nevertheless, they pointed the way to future infantry weapon design.

During World War II, Hugo Schmeisser designed a light rifle to fire the Germans' 7.92-millimetre Kurz, or "Short," cartridge, which was of the same calibre as the Mauser rifle cartridge but was lighter and shorter and was therefore of a less potent, "intermediate" power. The weapon, known variously as the MP43, MP44, or Sturmgewehr ("Assault Rifle") 44, was loaded by a curved box magazine holding 30 rounds and was designed for most effective fire at about 300 yards. Only 425,000 to 440,000 of these rifles were built—too few and too late for the German war effort—but they were based on a concept that would dominate infantry weapons for the rest of the century.

The AK-47

Late in the war the Soviets also began a search for a rifle to shoot their 7.62-millimetre intermediate cartridge, which produced a muzzle velocity of 2,330 feet per second. Historical evidence suggests that they were influenced by the Sturmgewehr, but to what extent remains uncertain. In 1947 they adopted a weapon designed by Mikhail Timofeyevich Kalashnikov, naming it the Avtomat Kalashnikova. Like the German weapon, the AK-47 (see Figure 12) was operated by diverting some of the propellant gases into a cylinder above the barrel; this drove a piston that forced the bolt back against its spring and cocked the hammer for the next round. At the turn of a selector switch, the action could be changed from semiautomatic to fully automatic, firing at a rate of 600 rounds per minute. The AK-47 was made of forged and milled steel, giving it a weight of 10.6 pounds (4.8 kilograms) with a loaded 30-round magazine.

The receiver of the AKM version, introduced in 1959, was made of lighter sheet metal, reducing the weight to 8.3 pounds, and the AK-74 version, following later trends in the West, switched to a 5.45-millimetre cartridge.

Kalashnikov's assault rifles became the most significant infantry weapons of the post-World War II era. In many variants, they were adopted and made by countries all over the world. Between 30 and 50 million AKs were produced within four decades of the series' introduction, more than any other firearm in history.

The development of Western small arms proceeded more slowly, mainly because the United States insisted upon maintaining a power level comparable to the M1. As a result, in 1953 the North Atlantic Treaty Organization (NATO) reluctantly agreed to standardize on a 7.62-millimetre cartridge that was a half-inch shorter than the M1 cartridge but of the same calibre and power. To fire this new round, the United States produced an improved version of the M1 rifle, featuring a 20-round detachable magazine and being capable of selective fire. Called the U.S. Rifle 7.62mm M14, it replaced the M1, beginning in 1957. As a self-loading rifle the M14 performed well, but it was too heavy as a close-quarters weapon, and the extreme recoil generated by the NATO round caused it to be totally unmanageable as an automatic rifle.

Other NATO armies adopted more satisfactory 7.62-millimetre rifles, although even these were employed as advanced self-loaders rather than automatics. Most commonly, they were either the gas-operated Fusil Automatique Léger (FAL), introduced by the Belgian Fabrique Nationale d'Armes de Guerre in 1957, or the blowback-operated Gewehr 3 (G3), produced by the West German firm of Heckler & Koch beginning in 1959. Millions of these weapons were sold to many countries.

After the Korean War (1950–53), U.S. military researchers dissatisfied with rifle-power ammunition began to test a .22-inch (5.56-millimetre) cartridge that propelled a lighter projectile at a much higher muzzle velocity of 3,000 feet per second. To fire this "small-calibre, high-velocity" round, in 1958 they chose the AR-15 rifle, designed by Eugene M. Stoner for the ArmaLite Division of Fairchild Engine and Airplane Corporation. The AR-15 was gas-operated, but it eliminated the piston in favour of a tube that directed propellant gases directly into an expansion chamber between the bolt and bolt carrier. By reducing the number of working parts and chambering the rifle for a smaller cartridge, Stoner had come up with a lightweight weapon that, even on automatic fire, produced a manageable recoil and yet was capable of inflicting fatal wounds at 300 yards and beyond. In 1961 the air force purchased the AR-15, renaming it the M16. Six years later, with units in Vietnam finding the weapon very effective under the close conditions of jungle warfare, the army adopted it as the M16A1.

The AR-15/M16

After U.S. troops in Europe were issued the M16, a series of trials ensued that ended with the decision, in 1980, to adopt a standard 5.56-millimetre NATO cartridge. This fired a brass-jacketed projectile that, having a heavier lead core and steel nose, was lethal at longer ranges than the original AR-15 bullet. The M16A2 was rifled to fire this round, and other NATO armies switched over. West Germany introduced the G41, a 5.56-millimetre version of the G3, and Belgium replaced the FAL with the FNC. The British and French armies developed new assault rifles with compact "bullpup" designs, in which the bolt, receiver, and magazine were behind the handgrip and trigger and much of the shoulder stock was occupied by the operating mechanism. This permitted a much shorter weapon than orthodox designs, in which the magazine and receiver were ahead of the trigger. As a result, the French FA MAS and British L85A1 were only some 30 to 31 inches long—compared with the M16, which was 39 inches overall. Many of the newer models were built with lightweight plastic shoulder stocks and magazines, as well as receivers made of aluminum. (E.C.E.)

MACHINE GUNS

The search for greater firepower has not been limited to shoulder firearms. In addition to personal-defense weapons, a variety of infantry-support weapons classed as machine guns have been subjected to intense experimentation.

Early manual weapons. During the flintlock era a number of heavy guns were developed that could fire several bullets either serially or in volley, but it was not until the mid-19th century, with the spread of centre-fire cartridge ammunition and better manufacturing techniques, that such weapons could be put to effective military use. The best known were the Gatling gun, invented by the American Richard J. Gatling, and the mitrailleuse, produced by the Belgian firm of Christophe & Montigny.

The Gatling gun. Gatling guns had several barrels (usually 6 or 10) mounted around a central axle and revolved by means of a hand crank. After a barrel fired a round, it went through successive unlocking, extracting, ejecting, reloading, and relocking. In the most successful Gatling guns, stacks of rounds could be fed by means of a feed device to give continuous fire for long periods. Gatling weapons were made to take a variety of ammunition, up to a full inch in calibre. A few were used by U.S. forces in Cuba in 1898 and in minor military operations around the world.

The mitrailleuse. The French mitrailleuse was also a multibarreled weapon, but it used a loading plate that contained a cartridge for each of its 25 barrels. The barrels and the loading plate remained fixed, and a mechanism (operated by a crank) struck individual firing pins simultaneously or in succession. The mitrailleuse issued to the French army fired 11-millimetre Chassepot rifle ammunition. Weighing more than 2,000 pounds, it was mounted on a wheeled carriage and was usually employed in volley fire, all barrels discharging at once. French forces in the Franco-German War endeavoured to use it in a manner similar to artillery, but it was no match for breech-loading cannon firing explosive shell.

Tactical misuse of the mitrailleuse

Heavy machine guns. Self-actuated machine guns, which operated under energy generated by a fired round, became militarily effective after the introduction of nitrocellulose propellants. These burned at a more controlled rate than did the older black-powder propellants, generating pressures that built up over a longer time. The first automatic weapons to take advantage of this were heavy guns firing the new, high-velocity rifle cartridges.

Recoil. The first successful automatic machine gun was invented by Hiram Stevens Maxim, an American working in Europe. Beginning about 1884, he produced a number of weapons in which the bullet's recoil energy was employed to unlock the breechblock from the barrel, to extract and eject the fired case from the gun, and to store sufficient energy in a main spring to push the bolt forward, pick up a fresh round, load the chamber, and lock the piece. Both barrel and breechblock recoiled a short distance to the rear locked together; then the barrel was stopped and the block continued back alone. If the trigger

Imperial War Museum

Figure 13: A Maxim machine gun, belt-fed and water-cooled, operated by German infantrymen, World War I.

was held in firing position, the weapon would continue to fire until it expended all of its ammunition. Rounds were fed to the gun on belts, which could be clipped together to provide continuous fire, and overheating was solved by surrounding the barrel in a metal jacket in which water was circulated from a separate container.

Maxim's salesmen provided armies with guns in any calibre, usually matching their current rifle cartridge. In Britain, Maxim guns were first chambered for the .45-inch Martini-Henry cartridge, but, as issued in 1891, they fired the .303-inch smokeless-powder round of the Lee-Metford rifle. During the Russo-Japanese War (1904–05), the Russians used English-made Maxim guns chambered for their 7.62-millimetre Mosin-Nagant round. Their Model 1910 weighed about 160 pounds, including mount, water-cooling apparatus, and a protective steel shield for the gunner. The German Model 1908, chambered for the 7.92-millimetre Mauser cartridge, weighed 100 pounds with its sled mount. Such light weights, made possible because the cartridge was the sole source of power, allowed these weapons to be operated by special infantry units (see Figure 13).

Power of the Maxim gun

Machine guns of the Maxim type had a destructive power never seen before in warfare. In the 1890s, British infantry units used Maxim guns, fabricated under contract by Vickers Sons, to cut down hordes of poorly armed rebels in Africa and Afghanistan. In World War I, a few of them could cause thousands of casualties. Their defensive fire so limited the offensive power of infantry that the entire Western Front, from the Swiss border to the English Channel, became one vast siege operation.

Gas operation. Not all the early heavy machine guns were of the recoil-operated Maxim type. Gas operation was also employed. In this system a piston located in a cylinder below the barrel was driven to the rear by gas diverted from the barrel through a port. The piston unlocked the breechblock and sent the bolt back against the main spring; a new round was then picked up, moved into the chamber, and fired on the forward stroke.

The best-known gas-operated heavy machine gun was the Hotchkiss, introduced in France in 1892 and modified several times until the definitive version of 1914. It was air-cooled, but the barrel itself was heavy and provided with metal fins to increase heat radiation. A slower method of feeding ammunition by short strips instead of long belts also helped to keep the weapon from overheating. The Japanese used Hotchkiss guns chambered for their 6.5-millimetre round against Russia in 1904–05. In World War I, two French Hotchkiss guns firing 8-millimetre Lebel cartridges were said to have fired 75,000 rounds each in the defense of Verdun and to have remained serviceable.

Blowback. A third principle of machine-gun operation was often called blowback. In this, the action and barrel were never locked rigidly together; the barrel did not move, nor was there a gas cylinder and piston. To prevent the breech from opening so early that propellant gases would rupture the spent cartridge case, the block was heavy and the main spring strong. Also, there was usually a linkage of parts not quite on centre to delay the actual opening. Finally, the barrel was shorter than usual, allowing the bullet and gases to leave the barrel quickly.

The Austrian Schwarzlose of 1907/12, firing eight-millimetre Mannlicher rounds, operated by delayed blowback. It was entirely satisfactory in combat during World War I.

Light machine guns. Heavy machine guns were satisfactory for defensive roles but were not really portable. A number of lighter machine guns (frequently called machine rifles or automatic rifles) began to be used in 1915. These included the British Lewis gun (invented in America but manufactured and improved in Great Britain), the French Chauchat, several German weapons, and the U.S. M1918 Browning automatic rifle (known as the BAR). Most, but not all, of these light weapons were gas-operated. Almost all were air-cooled. Generally, they fired from magazines rather than belts of ammunition because detachable magazines were more convenient and more easily transported. Weighing as little as 15 pounds, they were light enough to be carried by one man and fired rifle-fashion or from a prone position.

After World War I, light machine guns virtually took over the functions of their heavier counterparts, although the older weapons continued in service around the world through World War II and for decades thereafter. In Germany, where heavy, water-cooled Maxim-type guns had been forbidden by the victorious Allies, an entirely new generation of light machine guns was introduced by the Maschinengewehr 1934 and 1942. Recoil-operated and fed 7.92-millimetre rifle ammunition on belts, these were equally effective when fired from bipods or when mounted on tripods for sustained fire. Firing at an extremely high rate (as high as 1,000 rounds per minute), they dealt with the overheating problem by being built with barrels that could be changed in seconds. The MG34 pioneered the quick-change mechanism, while the MG42, being fabricated largely of stamped sheet-metal parts welded and riveted together, could be made cheaply and quickly even in factories designed for automobile manufacture.

The MG34 and MG42

The Soviets began to issue their Degtyarev Pekhotny (DP) in 1933 and supplied it to Loyalist forces in the Spanish Civil War. In 1944 it was modified into the DPM. British infantry units fought World War II with the Bren, a .303-inch version of a weapon designed by the Czech weapons maker Václav Holek, and U.S. troops relied on the BAR. All were gas-operated and magazine-fed and weighed from slightly over 20 pounds to more than 30 pounds loaded. They fired slowly enough to deliver accurate bursts from their bipods, 350–600 rounds per minute.

After the war, with assault-rifle cartridges becoming standard issue, terms such as automatic rifle, light machine gun, and medium machine gun gave way to general-purpose machine gun (GPMG) and squad automatic weapon (SAW). Most GPMGs were chambered for the intermediate-size 7.62-millimetre cartridges of NATO and the Soviet Union, while SAWs fired small-calibre, high-velocity rounds such as the 5.56-millimetre NATO or the 5.45-millimetre Kalashnikov. Significant belt-fed GPMGs included the West German MG3, a modernized version of the MG42; the Mitrailleuse d'Appui Général (MAG), built by Fabrique Nationale of Belgium; the U.S.-made M60; and the Soviet Pulemyot Kalashnikova (PK). Of the SAWs, the most prominent were the belt- or magazine-fed Minimi, built by Fabrique Nationale, and the magazine-fed Ruchnoy Pulemyot Kalashnikova (RPK).

Large-calibre machine guns. With the eclipse of the early water-cooled machine guns, the term heavy was applied to machine guns firing cartridges of several times rifle calibre—most often .50 inch or 12.7 millimetres.

Even before World War I, fully automatic weapons were used with ammunition more powerful than rifle cartridges, but such ammunition was not necessary for infantry missions until foot soldiers encountered armoured vehicles. During the 1930s, many higher-powered weapons were adopted, although only two had outstanding success. One was the United States' M2 Heavy Barrel Browning. Essentially a .50-inch version of the .30-inch M1917 Browning (a Maxim-type machine gun produced too late to see much fighting in World War I), the M2 was still widely used throughout the noncommunist world decades after World War II. Its cartridge delivered bullets of various weights and types at high muzzle velocities, with roughly five to seven times the energy of full rifle-power ammunition. The weapon was recoil-operated and air-cooled, and it fired at about 450 rounds per minute. The Soviet 12.7-millimetre weapon, the Degtyarov-Shpagin Krupnokaliberny 1938 (DShK-38), was similar, but it was gas-operated. It went into wide use in Soviet-supplied countries. Both of these weapons, as well as their successors (such as the Soviets' Nikitin-Sokolov-Volkov, or NSV, machine gun), were used by infantry units on wheeled or tripod mounts, but they were also mounted on tanks to provide defensive fire against ground vehicles or aircraft (see Figures 24 and 25).

Heavy antiaircraft guns

After 1945, several superheavy machine guns (more than .50 inch) were developed, mostly for antiaircraft use. The single most important was a 14.5-millimetre weapon first introduced by the Soviets for use in armoured vehicles. It was recoil-operated and belt-fed and had a barrel that

could be changed quickly. Later it was fielded on a variety of wheeled carriages and was known as the Zenitnaya Protivovozdushnaya Ustanovka. The ZPU-4, a four-barreled version towed on a trailer, shot down many U.S. aircraft during that nation's involvement in the Vietnam War (1965–73) and remained in service throughout the Third World long afterward. (J.We./J.F.G./E.C.E.)

PISTOLS

Since the 16th century, soldiers have carried handguns to supplement their basic shoulder weapons. However, because the firepower of pistols must be kept low in order to reduce them to manageable weight, and because only skilled soldiers can shoot them accurately beyond 10 yards, they have never been satisfactory military weapons. By World War II, pistols were issued principally to officers as a badge of rank and as a defensive weapon of last resort. Currently, they are most frequently carried by military police and other security personnel.

Revolvers. Until the mid-1840s most pistols were single-shot muzzle-loaders fired by wheel lock, flintlock, and percussion ignition systems. In 1835 Samuel Colt patented the first successful percussion revolver. In the frame of this weapon was a revolving cylinder drilled with several chambers (usually five or six), into which powder and ball (or combustible paper cartridges containing powder and ball) were loaded from the front. In the rear of each chamber, a percussion cap was placed over a hollow nipple that directed the jet of flame to the powder when the cap was struck by the hammer. This type of revolver was eventually called "cap-and-ball." Where earlier revolvers required the shooter to line up a chamber with the barrel and cock the hammer in separate steps, Colt devised a single-action mechanical linkage that rotated the cylinder as the hammer was cocked with the thumb.

Colt dominated the manufacture of revolvers until the expiration of his U.S. patent in 1857. At that time two other Americans, Horace Smith and Daniel B. Wesson, produced the first cartridge revolver, based on a design purchased from Rollin White. Using rim-fire copper cartridges and eliminating the percussion-cap nipple, this weapon could be quickly loaded from the rear.

Double action

When the Smith & Wesson patent expired in 1872, a host of new revolver designs appeared in the United States and Europe. The most important innovations were quick ejection of spent cartridges and double-action cocking. By linking the trigger to the hammer-cocking and cylinder-revolving mechanisms, double action permitted a pistol to be fired with a simple pull of the trigger. This mechanism was first introduced on a cap-and-ball revolver, the English Beaumont-Adams of 1855, but it was quickly adapted to cartridge revolvers. There were several mechanisms for removing spent cartridge cases. In the 1870s Smith & Wesson produced revolvers with hinged frames. When such a revolver was "broken open"—that is, when the barrel and cylinder were tipped on the hinge away from the hammer and handgrip—an ejector rod, located in the middle of the cylinder but having a star-shaped head that radiated into each chamber, pushed out all of the cartridges simultaneously. In the 1890s some Colt revolvers were made with solid frames but with cylinders that swung out to the side, where pushing an ejector rod forced out the cartridges.

By the end of the 19th century the revolver had reached its definitive form and its highest possible effectiveness as a military weapon. Indeed, from the 1880s through World War II, British officers carried such revolvers as the .45-inch Webley and the .38-inch Enfield, both of hinged-frame design. The U.S. military adopted various revolvers, usually Colts or Smith & Wessons of .38-inch or .45-inch calibre, until 1911, when it switched to autoloading pistols.

Self-loaders. A high rate of fire was especially crucial to last-ditch, close-quarters defense, and, with handguns as well as shoulder arms, this meant automatic loading. Following Hiram Maxim's experiments with self-loading weapons (see above *Machine guns*), automatic-pistol designs appeared in the last years of the 19th century.

In 1893 Ludwig Loewe & Company (later known as Deutsches Waffen- und Munitionsfabriken) introduced the first commercially viable self-loading pistol. Designed by an American, Hugo Borchardt, this 7.63-millimetre weapon operated on the principle of recoil. When the gun was fired, the barrel and breechblock, locked together by a "toggle-link" mechanism, slid back together along the top of the frame. The toggle, essentially a two-piece arm hinged in the middle but lying flat behind the breechblock, also recoiled for a short distance before it was forced to buckle upward at its hinge. This unlocked the breechblock from the barrel and allowed it to slide back on its own, extracting and ejecting the spent case, cocking the hammer, and compressing a coiled spring in the rear of the gun. The spring then pushed the breechblock forward, stripping a fresh cartridge from a magazine in the handgrip, and the toggle locked the breechblock once more against the barrel. Borchardt's toggle and spring mechanisms were improved by a German, Georg Luger, who came up with the 7.65-millimetre (later 9-millimetre) Parabellum pistol. This was adopted by the German army in 1908.

Browning's Colt .45

In the United States and many parts of Europe, John M. Browning's handgun designs dominated the first half of the 20th century. In his .45-inch pistol, manufactured by Colt and adopted by the U.S. military in 1911, the barrel and breechblock were covered and locked together by a housing called the slide. When the gun was fired, the recoiling slide pulled the barrel back a short distance until the barrel was disengaged and returned to its forward position by a spring. The unlocked slide and breechblock continued back, ejecting the spent case and cocking the hammer, until a spring forced them forward while a fresh cartridge was picked up from a seven-round magazine in the grip. The M1911 Colt did not begin being replaced until 1987. Its successor, the nine-millimetre Italian Beretta, given the NATO designation M9, reflected post-1970 trends such as large-capacity magazines (15 shots in the Beretta), double-action triggers (which could snap the hammer without its having to be cocked manually or automatically), and ambidextrous safety levers.

GRENADE LAUNCHERS

Soldiers have always favoured grenades for the killing and stunning effect of their explosive power, but the effectiveness of hand grenades has always been limited to the distance they can be thrown. Extending the range of grenades requires that they be launched by some sort of infantry weapon.

Single shot. During World War I, most armies developed attachments for standard service rifles that permitted the launching of "rifle" grenades. However, although range was increased with these devices, accuracy remained poor. An effective answer was a shoulder-fired grenade launcher developed in the 1950s by the Springfield Armory. Resembling a single-shot, break-open, sawed-off shotgun, the M79 lobbed a 40-millimetre, 6-ounce (176-gram) high-explosive fragmentation grenade at a velocity of 250 feet per second to a maximum range of 400 yards. This covered the area between the longest range of hand-thrown grenades (30 to 40 yards) and the middle range of 60-millimetre mortars (300–400 yards).

The grenade round

The M79 employed a "high-low pressure system" developed by Germany during World War II. This involved an aluminum cartridge case with a sealed propellant chamber in front of the primer. The propellant chamber was perforated by a number of partially completed, carefully sized holes leading into a separate expansion chamber within the cartridge case. Upon firing, the high pressures created inside the propellant chamber flowed into the expansion chamber through the previously prepared holes. The resulting moderated gas pressure produced a low impulse that launched the grenade at an adequate velocity and with an acceptable recoil impulse.

M79 grenade launchers were made from 1961 to 1971 and saw a great deal of action in Vietnam. Production was terminated in favour of a launcher attachment for the M16 rifle. Other 40-millimetre grenade launchers were built for the Soviet AK and West German assault rifles.

Automatic fire. Grenade-launching machine guns also appeared during the Vietnam War. Instead of the thin-walled projectiles fired by the M79, these shot higher-

velocity cartridges. The weapons were first mounted on helicopters but afterward appeared on tripods and armoured vehicles. On these mounts, grenade-launching machine guns such as the U.S. Mark 19, firing 40-millimetre rounds, and the Soviet AGS-17, shooting 30-millimetre projectiles, frequently replaced or supplemented .50-inch heavy machine guns.

ANTITANK WEAPONS

Upon their introduction in World War I, tanks posed a very serious problem for foot soldiers. The Germans quickly reacted by introducing the 13-millimetre Tankgewehr ("Antitank Rifle"), a very large-scale single-shot version of the Mauser bolt-action rifle. British designers created the magazine-fed, bolt-action .55-inch Boys antitank rifle in the late 1930s, and the Soviets introduced 14.5-millimetre bolt-action and self-loading antitank rifles during World War II. The increasing thickness of tank armour soon made all of these infantry weapons obsolete, since kinetic-energy weapons that could penetrate tank armour became too heavy and produced too much recoil to be fired from the shoulder.

The search for a shoulder-fired antitank weapon took another turn with the application of a principle discovered in the 1880s by an American inventor, Charles E. Munroe. Munroe found that a hollow cone of explosive material, when detonated with its open end a few inches from metal plate, produced a jet of white-hot gases and molten steel that could penetrate many inches of the best armour. Utilizing the Munroe principle, various "shaped-charge" projectiles were first delivered during World War II by low-velocity, shoulder-held rocket launchers such as the bazooka (see below *Rockets and missile systems*) or by recoilless devices such as the German Panzerfaust ("Tank Fist," or "Tank Puncher"). Issued in the latter half of the war, the German weapon was a 30-inch-long, 1.75-inch-diameter tube containing a charge of gunpowder. A six-inch-diameter bomb, mounted on a stick with collapsible fins, was inserted into the front end, and the weapon, held over the shoulder or under the arm, was fired by a simple firing pin and percussion cap on the outside of the tube. The propellant gases blew a cap off of the rear of the tube, in effect canceling the recoil forces generated by the launching of the bomb, which could be lobbed to ranges of 30 to 100 yards. Its powerful shaped charge of RDX and TNT could penetrate any tank armour.

Following World War II, the Soviet military perfected the Panzerfaust-type recoilless launch mechanism in their Ruchnoy Protivotankovy Granatomet 2 (RPG-2), a "Light Antitank Grenade Launcher" featuring a reusable launcher that lobbed an 82-millimetre shaped-charge warhead more than 150 yards. After 1962, with their RPG-7, they combined recoilless launch with a rocket sustainer to deliver a five-pound warhead to targets beyond 500 yards. The Soviet RPGs became powerful weapons in the hands of guerrillas and irregular fighters in conflict against more conventionally armed and heavily armoured forces. As such, they were used by the Viet Cong to destroy U.S. armoured vehicles in Vietnam and by militiamen in the protracted conflicts of the Middle East (see Figure 14).

The RPG-7

Other countries also developed small, shoulder-held recoilless launchers firing shaped-charge warheads. Some of them, such as the Swedish Miniman, came preloaded and were designed to be discarded after firing. (E.C.E.)

Artillery

For three centuries after the perfection of cast-bronze cannon in the 16th century, few improvements were made in artillery pieces or their projectiles. Then, in the second half of the 19th century, there occurred a series of advances so brilliant as to render the artillery in use when the century closed probably 10 times as efficient as that which marked its opening. These remarkable developments took place in every aspect of gunnery: in the pieces, with the successful rifling of cannon bores; in the projectiles, with the adoption of more stable elongated shapes; and in the propellants, with the invention of more powerful and manageable gunpowders.

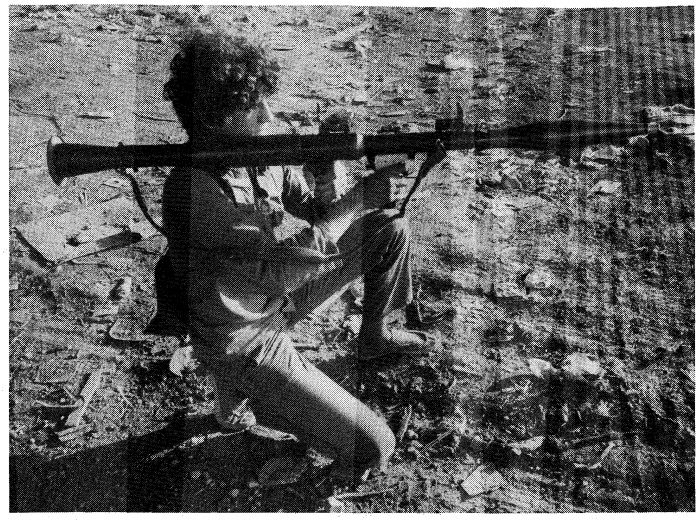


Figure 14: An RPG-7, a Soviet-designed recoilless launcher and rocket-propelled grenade, in the hands of an al-Fatah fighter, Lebanon.
© Reza/Sipa Press

These advances wrought a further transformation in the ever-changing nomenclature and classification of artillery pieces. Until the adoption of elongated projectiles, ordnance was classified according to the weight of the solid cast-iron ball a piece was bored to fire. But, because cylindrical projectiles weighed more than spheres of the same diameter, designation in pounds was abandoned, and the calibre of artillery came to be measured by the diameter of the bore in inches or millimetres. Cannon became the general term for large ordnance. A gun was a cannon designed to fire in a flat trajectory, a howitzer was a shorter piece designed to throw exploding shells in an arcing trajectory, and a mortar was a very short piece for firing at elevations of more than 45°.

Calibre by weight or by diameter

CANNONS

Rifled bores. In the middle years of the 19th century, smoothbore field artillery was placed at a disadvantage by the adoption of rifled small arms, which meant that infantry weapons could now outrange artillery. It therefore became vital to develop rifling for artillery weapons as well. The advantages of rifling were well known, but the technical difficulties of adapting the principle to heavy weapons were considerable. Several systems had been tried; these generally involved lead-coated projectiles that could engage shallow rifling grooves or projectiles fitted with studs that would fit into deeper rifling. None had proved adequate.

Breech loading. In 1854 William Armstrong, an English hydraulic engineer, designed an entirely new type of gun. Instead of simply boring out a solid piece of metal, Armstrong forged his barrel from wrought iron (later from steel). He then forged a succession of tubes and, by heating and shrinking, assembled them over the basic barrel so as to strengthen it in the area where the greatest internal pressure occurred. The barrel was rifled with a number of narrow, spiral grooves, and the projectile was elongated and coated with lead. The gun was loaded from the rear, the breech being closed by a "vent-piece" of steel that was dropped into a vertical slot and secured there by a large-diameter screw. The screw was hollow so as to make it lighter and facilitate loading.

In 1859 the British adopted the Armstrong system for field and naval artillery. During this same period, the Prussians had been testing guns made by Alfred Krupp, and in 1856 they adopted their first Krupp breechloader. This was made of a solid steel forging, bored and then rifled with a few deep grooves, and its breech was closed by a transverse sliding steel wedge. The Krupp projectile had a number of soft metal studs set into its surface, positioned so as to align with the rifling grooves. In both the Armstrong and Krupp guns, obturation—that is, the sealing of the breech against the escape of gas—was per-

formed by a soft metal ring let into the face of the vent piece or wedge. This pressed tightly against the chamber mouth to provide the required seal.

Meanwhile, the French adopted a muzzle-loading system designed by Treuille de Beaulieu, in which the gun had three deep spiral grooves and the projectile had soft metal studs. The gun was loaded from the muzzle by engaging the studs in the grooves before ramming the shell.

Armstrong guns were successful against Maoris in New Zealand and during the Opium Wars in China, but the development of ironclad ships in Europe demanded guns powerful enough to defeat armour, and the Armstrong gun's breech closure was not strong enough to withstand large charges of powder. Therefore, in 1865 the British adopted a muzzle-loading system similar to that of de Beaulieu, since only this would provide the required power and avoid the complications of sealing the breech.

Through the 1870s guns, particularly coastal-defense and naval guns, became longer so as to extract the utmost power from large charges of gunpowder. This made muzzle loading more difficult and gave a greater incentive to the development of an efficient breech-loading system.

The screw-threaded breech

Various mechanisms were tried, but the one that supplanted all others was the interrupted screw, devised in France. In this system the rear end of the bore was screw-threaded, and a similarly screwed plug was used to close the gun. In order to avoid having to turn the plug several times before closure was effected, the plug had segments of its thread removed, while the gun breech had matching segments cut away. In this way the screwed segments of the plug could be slipped past smooth segments of the breech, and the plug slid to its full depth. Then the plug could be revolved part of a turn, sufficient for the remaining threads to engage with those in the breech.

In the earliest applications of this system, obturation was provided by a thin metal cup on the face of the breechblock; this entered the gun chamber and was expanded tightly against the walls by the explosion of the charge. In practice, the cup tended to become damaged, leading to leakage of gas and erosion of the chamber. Eventually a system devised by another French officer, Charles Ragon de Bange, became standard. Here the breechblock was in two pieces—a plug screwed with interrupted threads and having a central hole, and a "vent bolt" shaped like a mushroom. The stem of the bolt passed through the centre of the breechblock, and the "mushroom head" sat in front of the block. Between the mushroom head and the block was a pad of resilient material shaped to conform to the chamber mouth. On firing, the mushroom head was forced back, squeezing the pad outward so as to provide a gas-tight seal. This system, refined by a century of experience, became the principal method of obturation used with major-calibre artillery.

The alternative to this system was the sliding breechblock and metallic cartridge case pioneered by Krupp. Here the case expanded under the charge pressure and sealed against the chamber walls. As the pressure dropped, the case contracted slightly and could be withdrawn when the breechblock was opened. This system was embraced first by German gunmakers and later was widely used in all calibres up to 800 millimetres (about 31 inches). However, during World War II (1939–45), when the Germans were faced with metal shortages that threatened cartridge-case production, they developed a form of "ring obturation" so that bagged charges could be used. In this system an expandable metal ring was set into the face of the sliding breechblock, and its seating was vented in such a manner that some of the propellant gas was able to increase the pressure behind the ring and so force it into tighter contact. As improved in the postwar years, this system was adopted on a number of tank and artillery guns.

Gun construction. The lasting legacy of the Armstrong gun was the system of building up the gun from successive tubes, or "hoops"; this was retained in the rifled muzzle-loading system of the 1870s and was gradually adopted by other countries. Armstrong's method not only economized on material, by distributing metal in accordance with the pressures to be resisted, but it also strengthened the gun.

An exception to the built-up system was practiced by Krupp. He bored guns from solid steel billets, making the barrels in one piece for all but the very largest calibres. In the mid-19th century it was difficult to produce a flawless billet of steel, and a flawed gun would burst explosively, endangering the gunners. A wrought-iron gun, on the other hand, tended to split progressively, giving the gunners warning of an impending failure. This was enough to warrant the use of wrought iron for many years, until steel production became more reliable.

Early steel barrels

The next major advance in gun construction came in the 1890s with wire-winding, in which one or more hoops were replaced by steel wire wound tightly around the tube. This gave good compressive strength but no longitudinal strength, and the guns frequently bent. Beginning in the 1920s, wire-winding was abandoned in favour of "auto-frettaging," in which the gun tube was formed from a billet of steel and then subjected to intense internal pressure. This expanded the interior layers beyond their elastic limit, so that the outer layers of metal compressed the inner in a manner analogous to Armstrong's hoops but in a homogenous piece of metal.

Recoil control. Until the 1860s guns were simply allowed to recoil along with their carriages until they stopped moving, and they were then manhandled back into firing position. The first attempt at controlling recoil came with the development of traversing carriages for coastal defenses and fortress guns. These consisted of a platform, pivoted at the front and sometimes carried on wheels at the rear, upon which a wooden gun carriage rested. The surface of the platform sloped upward to the rear, so that when the gun was fired and the carriage slid backward up the platform, the slope and friction absorbed the recoil. After reloading, the carriage was manhandled down the sliding platform, assisted by gravity, until the gun was once more in firing position, or "in battery." To compensate for varying charges and, hence, varying recoil forces, the surface of the slide could be greased or sanded.

Control was improved by an American invention, the "compressor." This consisted of loose plates, fitted at the sides of the carriage and overlapping the sides of the slide, which were tightened against the slide by means of screws. Another arrangement was the placing of a number of metal plates vertically between the sides of the slide and a similar set of plates hanging from the carriage, so that one set interleaved the other. By placing screw pressure on the slide plates, the carriage plates were squeezed between them and thus acted as a brake on the carriage movement.

American designers added to this by adopting a hydraulic buffer, consisting of a cylinder and piston attached to the rear of the slide. The fired gun recoiled until it struck the piston rod, driving the piston into the cylinder against a body of water to absorb the shock. British designers then adapted this by attaching the buffer to the slide and the piston rod to the carriage. As the gun recoiled, it drove the piston through water inside the cylinder; meanwhile, a hole in the piston head permitted the water to flow from one side of the piston to the other, giving controlled resistance to the movement. Return to battery was still performed by manpower and gravity.

The final improvement came with the development of mechanical methods of returning the gun to battery, generally by the use of a spring. When the gun recoiled, it was braked by a hydraulic cylinder and at the same time compressed a spring. As recoil stopped, the spring reasserted itself, and the gun was propelled back into battery. From there it was a short step to using compressed air or nitrogen instead of a spring, and such "hydropneumatic" recoil-control systems became standard after their introduction by the French in 1897.

Hydro-pneumatic recoil control

Carriages and mountings. In 1850 carriages were broadly of two types. Field pieces were mounted on two-wheeled carriages with solid trails, while fortress artillery was mounted either on the "garrison standing carriage," a box-like structure on four small wheels, or on the platform-and-slide mounting previously described.

Coast guns. Coastal-defense artillery was the focus of most design attention in the 1870–95 period, since rapidly improving warships appeared to constitute the principal threat. The first major advance was a "disappearing car-

riage," in which the gun was mounted at the end of two arms that were hinged to a rotating base. In the firing position, a counterweight or hydraulic press held the arms vertical, so that the gun pointed over the edge of the pit in which the mounting was built. On firing, recoil drove the gun back, causing the arms to pivot and sink the gun into the pit out of sight of the enemy, where it could be reloaded in safety. This type of mounting, in various forms, was widely adopted, but it was gradually realized to be excessively complicated in view of the practical difficulty of a ship's gun being able to hit such a small target at long range. In most countries the disappearing mounting ceased to be built in the 1890s, though many of those already in position continued in use into the 1920s in Europe and into the 1940s in the United States.

In the 1890s the "barbette" mounting for coastal-defense guns became the preferred pattern. Here the mounting was in a shallow pit, protected from enemy fire, but the muzzle and upper shield were permanently in view, firing across a parapet that helped protect the gunners. This type of mounting was made practical by the development of hydraulic recoil control systems, which permitted the mounting to remain stationary while the gun, carried in a cradle, was allowed to recoil under control and then return to battery by spring or pneumatic power. The barbette remained the standard mounting for coastal-defense guns until their virtual disappearance after 1945.

The French 75

Field artillery. Field carriage design entered a new era with the French 75-millimetre gun of 1897 (see Figure 15). This introduced an on-carriage hydropneumatic recoil-control system, a shield to protect the gunners, modern sighting, fixed ammunition, and a quick-acting breech mechanism—thus forming the prototype of what became known as the "quick-firing gun." The idea was quickly taken up in other countries, and, by the outbreak of World War I (1914–18), such weapons were standard in all armies. Mountings for larger guns—up to about 155 millimetres, or 6 inches, in calibre—simply enlarged this basic design.

Up to World War I, with horses providing the standard motive power, it was necessary to design heavy field artillery so that gun and mounting could be dismantled into components, each of which would be within the hauling capacity of a horse team. The gun then traveled in its various pieces until it was reassembled at the firing point. Steam traction was attempted by the British during the South African War (1899–1902), but it was found that tractors could not take guns into firing position, as their smoke and steam was visible to the enemy. The gradual improvement of the internal combustion engine promised a replacement for the horse, but it saw relatively little application until the middle of World War I—and then only for heavier types of artillery.

The type of carriage developed for very heavy weapons was exemplified by that used for the German 420-millimetre howitzers—collectively known as "Big Bertha"—that were used to reduce the fortresses of Liège, Belg., in 1914. The equipment was split into four units—barrel, mounting with recoil system, carriage, and ground platform—which were carried on four wagons pulled by Daimler-Benz tractors. A fifth wagon carried a simple hoist, which, erected over the gun position, was used to lift the components from their wagons and fit them together. As the Great War continued, heavier howitzers and longer-ranging guns were made so large that they could not be split into convenient loads for road movement. Thus, the railway mounting became a major type for guns and howitzers up to 520-millimetre calibre. The heaviest guns could be assembled on large mountings, which in turn could be carried on a number of wheels so as to distribute the load evenly onto a railway track. The most impressive railway gun built during the war was the German 210-millimetre "Paris Gun," which bombarded Paris from a range of 68 miles (109 kilometres) in 1918. Like many other railway guns, the Paris Gun was moved to its firing position by rail but, once in place, was lowered to a prepared ground platform.

Ian V. Hogg

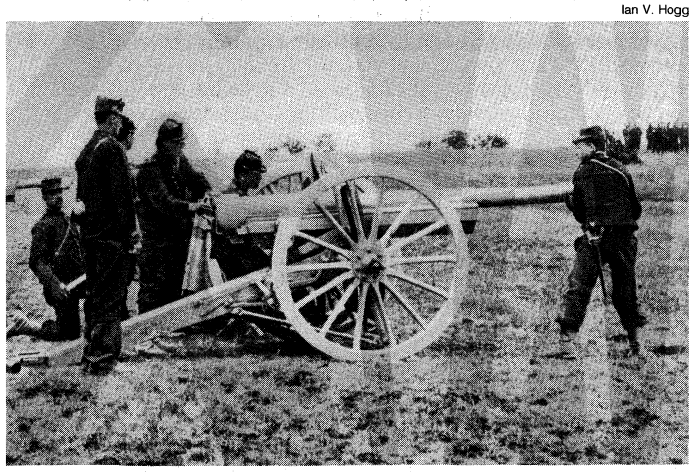


Figure 15: *The French 75-millimetre cannon of 1897.* With screw-threaded breech, hydropneumatic recoil control, and metallic-cartridge ammunition, this was the archtypal quick-firing gun of the late 19th and early 20th centuries.

Advances in carriage design after 1918 were relatively minor. The first was the general adoption of the split trail, in which two trail legs, opened to roughly 45°, were able to support a gun through a wider angle of traverse. Beginning in the 1960s came the gradual adoption of lightweight materials, culminating in the introduction by the British Vickers firm of a carriage built of titanium, which allowed a 155-millimetre howitzer to be helicopter-lifted. The 1960s also saw the introduction of auxiliary propulsion. Consisting of small motors that drove the wheels of towed guns, this permitted the gun to be moved from its firing position to a concealed or alternative position without calling up the towing vehicle. Propulsion motors also allowed the adoption of powered loading and ramming devices and powered assistance in opening trail legs and lowering platforms, thereby allowing the size of the crew to be reduced (see Figure 16).

Auxiliary propulsion

Fire control. *Azimuth and range.* In the 1850s the tactics of artillery were simple: the gun was positioned well to the front and fired across open sights straight at the enemy. The general adoption after the 1880s of long-range rifles firing smokeless-powder rounds rendered this tactic hazardous, and the South African War and Russo-Japanese War (1904–05) brought about a change in policy. Guns had to be concealed from the enemy's view, and a system had to be found that allowed them to be aimed without a direct view of the target. The solution was the adoption of the "goniometric," or "panoramic," sight, which could be revolved in any direction and which was graduated in degrees relative to the axis of the gun bore. The gun's position and that of the target were marked on a map, and the azimuth (the number of degrees clockwise from due north) between the two was measured. A prominent local feature, or a marker placed some distance from the gun, was then selected as an aiming point, and the azimuth between this and the gun's position was also measured. Subtraction of one from the other produced the angle between a line to the aiming point and a line to the target. If this angle was then set on the goniometric sight and the gun shifted until the sight was laid on the aiming point, then the bore of the gun would be pointed at the target.

Once the azimuth was calculated, the range was arrived at by measuring off the map. This was then converted into an angle by consulting a table, calculated during development of the gun, on which ranges were tabulated against angles of elevation. The angle was then set on an adjustable spirit-level (a clinometer) attached to the elevating portion of the gun. Setting the elevation angle displaced a bubble from the level position, and elevating the gun until the bubble returned to the level position brought the gun bore to the correct elevation angle.

The combination of these two techniques was sufficient to fire a shell that would land close to the target. From there, a forward observer would instruct the gunner to change the azimuth and elevation until the shells struck the target. At this point the remaining guns of the battery, which would have followed the corrections and set

Figure 16: *The 155-millimetre South African G5 of 1983.*
With a computer terminal for displaying gun-laying data and a diesel-fueled auxiliary power unit for opening the trail legs and raising the firing platform, this gun-howitzer epitomized artillery of the late 20th century.
Ian V. Hogg

them on their own sights, would join in to carry out the bombardment.

Predicted fire. During World War I it became tactically desirable to bombard an enemy position without alerting him by ranging shots. This brought about the development of "predicted fire."

Influences on the performance of a shell

While it is possible to determine azimuth and range from a map with accuracy, it is difficult to predict the actual performance of a fired shell. The density and temperature of the air through which the shell passes, the temperature of the propelling charge, any variation in weight of the shell from standard, any variation in the velocity of the shell owing to gradual wear on the gun—these and similar environmental changes can alter the performance of the shell from its theoretical values. Beginning in the 1914–18 period, these phenomena were studied and tables of correction were developed, together with a meteorological service that produced information upon which to base the corrections. This technique of predicted fire was slowly improved and was widely used during World War II, but the corrections were an approximation at best, owing to the simple tabular methods of applying the corrections. It was not until the introduction of computers in the 1960s that it became possible to apply corrections more accurately and more rapidly.

Target acquisition. Until the second half of the 20th century, target acquisition—a vital part of fire control—was almost entirely visual, relying upon ground observers. This was augmented first by observation balloons and then, in World War II, by light aircraft, the object of both being to obtain better visual command over the battlefield.

In World War I two technical methods of targeting enemy gun positions were adopted—sound ranging and flash spotting. In sound ranging, a number of microphones were used to detect the sound waves of a gun being fired; by measuring the time interval between the passing of sound waves across different microphones, it was possible to determine a number of rays of direction that, when plotted on a map, would intersect at the position of the enemy's gun. Flash spotting relied upon observers noting the azimuth of gun flashes and plotting these so as to obtain intersections. Both methods were highly effective, and sound ranging remained a major means of target acquisition for the rest of the century. Flash spotting fell into disuse after 1945, owing to the general adoption of flashless propellants, but in the late 1970s a new system of flash spotting became possible, using infrared sensors to detect the position of a fired gun.

Ammunition. *Projectile, powder, and fuze.* In 1850, round solid shot and black powder were standard ammunition for guns, while howitzers fired hollow powder-filled shells ignited by wooden fuzes filled with slow-burning powder. The introduction of rifled ordnance allowed the adoption of elongated projectiles, which, because of their streamlined forms, were much less affected by wind than round balls and, being decidedly heavier than balls of like diameter, ranged much farther. Yet the changing shape of projectiles did not at first affect their nature. For example, the shrapnel shell, as introduced in the 1790s by the Englishman Henry Shrapnel, was a spherical shell packed with a small charge of black powder and a number of musket balls. The powder, ignited by a simple fuze, opened the shell over concentrations of enemy troops, and the balls, with velocity imparted by the flying shell, had the effect of musket fire delivered at long range. When rifled artillery came into use, the original Shrapnel design was simply modified to suit the new elongated shells and remained the standard field-artillery projectile, since it was devastating against troops in the open.

Early shrapnel

Owing to the stabilizing spin imparted them by rifling grooves, elongated projectiles flew much straighter than balls, and they were virtually guaranteed to land point-first. Utilizing this principle, elongated powder-filled shells were fitted at the head with impact fuzes, which ignited the powder charge on striking the target. This in turn led to the adoption of powder-filled shells as antipersonnel projectiles. In naval gunnery, elongated armour-piercing projectiles initially were made of solid cast iron, with the heads chilled during the casting process to make them harder. Eventually, shells were made with a small charge of powder, which exploded by friction at the sudden deceleration of the shell upon impact. This was not an entirely satisfactory arrangement, since the shells generally exploded during their passage through the armour and not after they had penetrated to the vulnerable workings of the ship, but it was even less satisfactory to fit the shells with impact fuzes, which were simply crushed upon impact.

Between 1870 and 1890 much work was done on the development of propellants and explosives. Smokeless powders based on nitrocellulose (called ballistite in France and cordite in Britain) became the standard propellant, and compounds based on picric acid (under various names such as lyddite in Britain, melinite in France, and shimose in Japan) introduced modern high-explosive filling for shells. These more stable compounds demanded the development of fuzes adequate for armour-piercing shells, since friction was no longer a reliable method of igniting them. This was accomplished by fitting fuzes at the base of the shells, where impact against armour would not damage them but the shock of arrival would initiate them.

Time fuzes, designed to burst shrapnel shell over ground forces at a particular point in the shell's trajectory, were gradually refined. These usually consisted of a fixed ring carrying a train of gunpowder, together with a similar but moveable ring. The moveable ring allowed the time of burning to be set by varying the point at which the fixed ring ignited the moveable train and the point at which the moveable train ignited the explosive.

During World War I these fuzes were fitted into antiaircraft shells, but it was discovered that they burned unpredictably at high altitudes. Powder-filled fuzes that worked under these conditions were eventually developed, but the Krupp firm set about developing clockwork fuzes that were not susceptible to atmospheric variations. These clockwork fuzes were also used for long-range shrapnel firing; inevitably, an undamaged specimen was recovered by the British, and the secret was out. By 1939 clockwork fuzes of various patterns, some using spring drive and some centrifugal drive, were in general use.

World War I also saw the development of specialized projectiles to meet various tactical demands. Smoke shells, filled with white phosphorus, were adopted for screening the activities of troops; illuminating shells, containing magnesium flares suspended by parachutes, illuminated the battlefield at night; gas shells, filled with various chemicals such as chlorine or mustard gas, were used against troops; incendiary shells were developed for setting fire to hydrogen-filled zeppelins. High explosives were improved, with TNT (trinitrotoluene) and amatol (a mixture of TNT and ammonium nitrate) becoming standard shell fillings.

World War II saw the general improvement of these shell types, though the same basic features were used and flashless propellants, using nitroguanidine and other organic compounds, gradually took over from the earlier simple nitrocellulose types. The proximity fuze was developed by joint British–American research and was adopted first for air defense and later for ground bombardment. Inside the proximity fuze was a small radio transmitter that sent out a continuous signal; when the signal struck a solid object, it was reflected and detected by the fuze, and the interaction between transmitted and received signals was used to trigger the detonation of the shell. This type of fuze increased the chances of inflicting damage on aircraft targets, and it also allowed field artillery to burst shells in the air at a lethal distance above ground targets without having to establish the exact range for the fuze setting.

The proximity fuze

After 1945 the proximity fuze was improved by the transistor and the integrated circuit. These allowed fuzes to be considerably reduced in size, and they also allowed the cost to be reduced, making it economically possible to have a combination proximity/impact fuze that would cater to almost all artillery requirements. Modern electronics also made possible the development of electronic time fuzes, which, replacing the mechanical clockwork type, could be more easily set and were much more accurate.

Nuclear shells, guided projectiles, and rocket assistance. Nuclear explosive was adapted to artillery by the United States' "Atomic Annie," a 280-millimetre gun introduced in 1953. This fired a 15-kiloton atomic projectile to a range of 17 miles, but, weighing 85 tons, it proved too cumbersome for use in the field and was soon obsolete. In its place, nuclear projectiles with yields ranging from 0.1 to 12 kilotons were developed for conventional 203-millimetre howitzers. Soviet major-calibre artillery was also provided with nuclear ammunition.

The 1970s saw the first moves toward "improved conventional munitions." These were artillery projectiles carrying a number of subprojectiles—antipersonnel bombs or mines or antitank mines—that could be fired from a gun and would be opened, by a time fuze, over the target area to distribute the submunitions. This increased the destructive power of an artillery shell by a large amount and allowed field artillery to place obstacles in the path of enemy tanks at a range of several miles. A further step was the development of guided projectiles. With the 155-millimetre Copperhead, a U.S. system, a forward observer could "illuminate" a target with laser light, a portion of which would be reflected and picked up by sensors in the approaching shell. The greater part of the shell's flight would be entirely ballistic, but in the last few hundred yards it would be controlled by fins or other means, which, guided by the laser detection system, would "home" the shell onto the target.

In order to improve the range of guns, rocket-assisted projectiles were developed, with moderate success, by the Germans during World War II, and they were the subject of further development in succeeding years. Rocket assistance had certain drawbacks—notably, the loss of payload space in the shell to the rocket motor. A system designed to solve this problem was "base bleed," in which a small compartment in the base of the shell was filled with a piece of smokeless propellant. This would burn during flight, and the emergent gases would fill the vacuum left behind the shell in its passage through the air, reducing aerodynamic drag on the shell and improving the range by about 25 to 30 percent.

MORTARS

The mortar declined in importance during the 19th century but was restored by World War I, when short-range, high-trajectory weapons were developed to drop bombs into enemy trenches. Early designs in that conflict ranged from the 170-millimetre German Minenwerfer ("mine thrower"), which was almost a scaled-down howitzer, to primitive muzzle-loading devices manufactured from rejected artillery shells. The prototype of the modern mortar was a three-inch weapon developed by the Englishman Wilfred Stokes in 1915. This consisted of a smooth-bored tube, resting upon a baseplate and supported by a bipod, that had a fixed firing pin at its breech end. The bomb was a simple cylinder packed with explosive and fitted with a shotgun cartridge at the rear; its fuze was adapted from a hand grenade. When the bomb was dropped down the barrel of the mortar, it fired automatically as the shotgun cartridge struck the fixed firing pin. The bomb was unstable in flight but sufficiently accurate for its purpose, and it was soon replaced by a teardrop-shaped bomb with fins at the rear, which lent greater stability and accuracy. The Stokes mortar was rapidly adopted or copied by all belligerents.

The Stokes mortar

Some later mortars were built with rifled barrels, since these provided better sealing of the propellant gas and greater stability and accuracy owing to the spin imparted to the bomb. The difficulty here was to arrange for the bomb to be drop-loaded freely and yet engage the rifling once the propelling charge exploded. The U.S.-made M30, a 107-millimetre rifled mortar, used a saucer-shaped copper disk behind the bomb that flattened out into the rifling under gas pressure and provided obturation. In the 120-millimetre French Hotchkiss-Brandt type, a prerifled copper driving band, wrapped around the bomb, expanded under gas pressure and engaged the grooves in the barrel.

ANTIAIRCRAFT ARTILLERY

Heavy weapons and the problem of fire control. The development of antiaircraft guns began in 1909. The manufacture of suitable guns and mountings was not difficult at that time, but the fire-control problem, involving a target moving in three planes at high speed, was almost insoluble. The first fire-control system used complex gun sights that aimed the gun well in front of the target in order to give the shell time to reach it. The first projectiles were shrapnel, since scattered lead balls were sufficient to damage the aircraft of the day.

During World War I, attacks by German zeppelins led the British to produce incendiary shells. Forced to correct fire by visual methods, they fitted the shells with tracer devices, which, by leaving a trail of flame and smoke, indicated the shell's trajectory in the air. The French invented the "central post" system of fire control, in which an observing instrument in the centre of the battery calculated the aiming information, which was then passed on to the guns. This removed complex sights from the weapons and reduced the number of skilled operators required in a battery. Early warning of approaching aircraft was by visual means and acoustic devices.

In the 1920s work began on the design of "predictors," mechanical computers that could be given the course, height, and speed of the aircraft as well as the ballistic constants of the gun and could then calculate the gun data necessary to place the shell in the future position of the aircraft. These represented a significant advance in antiaircraft fire, but they still relied upon raw data provided by visual acquisition and tracking. In World War II, radar brought more accurate and timely acquisition and tracking, and the gradual adoption of electrical, rather

Predicting the position of aircraft

than mechanical, predictors produced more accurate fire control. Also, rapid-loading and fuze-setting devices were incorporated into gun mountings so that a high rate of fire could be achieved.

The proximity fuze removed the need for fuze setting and thus speeded up the rate of fire, until it was possible for guns of 90- to 100-millimetre calibre to fire at rates up to 60 rounds per minute. However, in the 1950s, when all these techniques were perfected, guided surface-to-air missiles became practical, and, in all major countries except for the Soviet Union, the use of medium and heavy air-defense guns ceased. (For a description of surface-to-air missiles, see below *Rockets and missile systems: Tactical guided missiles.*)

Light weapons. Light air-defense guns, of calibres from 20 to 40 millimetres, were developed in the 1930s for protection against dive bombers and low-level attack. The most famous of these was a 40-millimetre gun sold by the Swedish firm of Bofors. Virtually an enlarged machine gun, this fired small exploding shells at a rate of about 120 rounds per minute—fast enough to provide a dense screen of fragments through which the aircraft would have to fly. Fire control was largely visual, though some guns were equipped with predictors and power control.

The advent of lightweight missiles also threatened to render the light gun obsolete in the 1950s, but two decades later the development of electro-optical sights, using television and thermal-imaging technology and allied to computers and powered mountings, led to a resurgence of this class of weapon. In Egyptian hands in October 1973, the Soviet ZSU-23-4, consisting of four 23-millimetre guns mounted on a tracked vehicle, shot down many Israeli fighters over the Sinai Peninsula. The Bofors firm mounted its guns on wheeled vehicles, and the United States fielded a mobile system called Vulcan, which consisted of a six-barreled, Gatling-type gun firing 20-millimetre ammunition.

ANTITANK GUNS

The development of dedicated weapons for attacking tanks began in earnest in the 1930s. These were all in the 20- to 40-millimetre class, were mounted on light, two-wheeled, split-trail carriages, and were adequate against the tanks of the day. As tanks acquired heavier armour during World War II, so the guns became larger, eventually reaching 128 millimetres in calibre. The guns themselves did not generally demand new technology, but the development of ammunition had to break new ground.

The initial antitank projectile was a solid shot of hardened steel, and, in order to penetrate thicker tank armour, it was fired at higher and higher velocities. However, at a striking velocity of about 2,600 feet (800 metres) per second, steel shot shatters upon impact instead of penetrating. In order to overcome this, projectiles of tungsten carbide were used. The Germans designed a gun with a bore actually tapering in diameter from breech to muzzle, and for ammunition they constructed a projectile with a tungsten core and a soft metal body that would deform and squeeze in the reducing bore. The combination of reduced base area and constant gas pressure increased the projectile's velocity, and the "taper-bore" or "squeeze-bore" gun proved formidable. Guns with tapering calibres of 28/20, 41/29, and 75/55 millimetres were developed, but wartime shortages of tungsten led to their abandonment after 1942. In 1944 Britain perfected "discarding-sabot" projectiles, in which a tungsten core was supported in a conventional gun by a light metal sabot that split and fell free after leaving the muzzle, allowing the core to fly on at extremely high velocity.

High-velocity taper-bore guns

An alternative method was to use high explosives in the form of shaped-charge or squash-head projectiles. The shaped charge was an explosive formed into a hollow cone and lined with heavy metal; upon detonation, the explosive gases and molten metal formed a high-velocity jet capable of punching through armour. The squash-head shell used a plastic explosive filling, which, deposited on the armour and then detonated, drove a shock wave through the plate. This resulted in the failure of the inner face and the ejection of a massive slab of metal into the tank.

Heavy antitank guns relying upon high-velocity projectiles largely fell into disuse after 1945, but the technology was perpetuated in the main armament mounted on tanks (see below *Tanks and armoured vehicles*). Explosive-energy projectiles were also used on tanks as well as on recoilless guns.

RECOILLESS GUNS

Military inventors were long attracted by the prospect of abolishing recoil, since achieving this meant doing away with the gun's heavy recoil system and lightening the carriage. The first to succeed was Commander Cleland Davis of the U.S. Navy, who in 1912 developed a gun with a single chamber and two opposite barrels. One barrel carried the projectile, the other an equal weight of grease and lead shot. The explosion of the central cartridge ejected both loads, and, since the recoils had the same weight and velocity, they canceled each other out and the gun remained stationary. Davis' idea was adopted in 1915 by the Royal Naval Air Service, which ordered guns of 40, 57, and 75 millimetres for arming aircraft against airships and submarines. Few were made, however, and there appears to be no record of their use in combat.

If the Davis principle were taken to its logical ends, the countershot could be half the weight and twice the velocity of the principal projectile or any other combination giving the same momentum; at its ultimate, the countershot could simply be a cloud of high-velocity gas. This was the system upon which recoilless guns of up to 105 millimetres were developed during World War II. The cartridge cases of these weapons had a weakened section that ruptured on firing, allowing about four-fifths of the propellant gas to be exhausted to the rear of the gun. There they passed through a venturi, a nozzle with a constricted portion that increased the gas velocity and so balanced the recoil generated by the projectile. The back-blast caused by the exhausted gases made these weapons difficult to emplace and conceal, but after 1945 they were universally adopted as light antitank weapons. (I.V.H.)

The recoilless cartridge case

Rockets and missile systems

Rocket is a general term used broadly to describe a variety of jet-propelled missiles in which forward motion results from reaction to the rearward ejection of matter (usually hot gases) at high velocity. The propulsive jet of gases usually consists of the combustion products of solid or liquid propellants.

In a more restrictive sense, rocket propulsion is a unique member of the family of jet-propulsion engines that includes turbojet, pulse-jet, and ramjet systems. The rocket engine is different from these in that the elements of its propulsive jet (that is, the fuel and oxidizer) are self-contained within the vehicle. Therefore, the thrust produced is independent of the medium through which the vehicle travels, making the rocket engine capable of flight beyond the atmosphere or propulsion underwater. The turbojet, pulse-jet, and ramjet engines, on the other hand, carry only their fuel and depend on the oxygen content of the air for burning. For this reason, these varieties of jet engine are called air-breathing and are limited to operation within the Earth's atmosphere.

For the purposes of this article, a rocket engine is a self-contained (*i.e.*, non-air-breathing) propulsion system of the type described above, while the term rocket refers to any free-flight (unguided) missile of the types used since the beginning of rocketry. A guided missile is broadly any military missile that is capable of being guided or directed to a target after having been launched. Tactical guided missiles are shorter-ranged weapons designed for use in the immediate combat area. Long-range, or strategic, guided missiles are of two types, cruise and ballistic. Cruise missiles are powered by air-breathing engines that provide almost continuous propulsion along a low, level flight path. A ballistic missile is propelled by a rocket engine for only the first part of its flight; for the rest of the flight the unpowered missile follows an arcing trajectory, small adjustments being made by its guidance mechanism. Strategic missiles usually carry nuclear warheads, while tactical missiles usually carry high explosives.

MILITARY ROCKETS

Early history. There is no reliable early history of the "invention" of rockets. Most historians of rocketry trace the development to China, a land noted in ancient times for its fireworks displays. In 1232, when the Mongols laid siege to the city of K'ai-feng, capital of Honan province, the Chinese defenders used weapons that were described as "arrows of flying fire." There is no explicit statement that these arrows were rockets, but some students have concluded that they were because the record does not mention bows or other means of shooting the arrows. In the same battle, it is reported, the defenders dropped from the walls of the city a kind of bomb described as "heaven-shaking thunder." From these meagre references some students have concluded that by 1232 the Chinese had discovered black powder (gunpowder) and had learned to use it to make explosive bombs as well as propulsive charges for rockets. Drawings made in military documents much later show powder rockets tied to arrows and spears. The propulsive jet evidently added to the range of these weapons and acted as an incendiary agent against targets.

In the same century rockets appeared in Europe. There is indication that their first use was by the Mongols in the Battle of Legnica in 1241. The Arabs are reported to have used rockets on the Iberian Peninsula in 1249; and in 1288 Valencia was attacked by rockets. In Italy, rockets are said to have been used by the Paduans (1379) and by the Venetians (1380).

Construction of early rockets

There are no details of the construction of these rockets, but they were presumably quite crude. The tubular rocket cases were probably many layers of tightly wrapped paper, coated with shellac. The propulsive charge was the basic black powder mixture of finely ground carbon (charcoal), potassium nitrate (saltpetre), and sulfur. The English scientist Roger Bacon wrote formulas for black powder about 1248 in his *Epistola.* In Germany a contemporary of Bacon, Albertus Magnus, described powder charge formulas for rockets in his book *De mirabilibus mundi.* The first firearms appeared about 1325; they used a closed tube and black powder (now referred to as gunpowder) to propel a ball, somewhat erratically, over varying distances. Military engineers then began to invent and refine designs for both guns and rockets.

By 1668, military rockets had increased in size and performance. In that year, a German colonel designed a rocket weighing 132 pounds (60 kilograms); it was constructed of wood and wrapped in glue-soaked sailcloth. It carried a gunpowder charge weighing 16 pounds. Nevertheless, the use of rockets seems to have waned, and for the next 100 years their employment in military campaigns appears to have been sporadic.

The 19th century. A revival commenced late in the 18th century in India. There Hyder Ali, prince of Mysore, developed war rockets with an important change: the use of metal cylinders to contain the combustion powder. Although the hammered soft iron he used was crude, the bursting strength of the container of black powder was much higher than the earlier paper construction. Thus a greater internal pressure was possible, with a resultant greater thrust of the propulsive jet. The rocket body was lashed with leather thongs to a long bamboo stick. Range was perhaps up to three-quarters of a mile (more than a kilometre). Although individually these rockets were not accurate, dispersion error became less important when large numbers were fired rapidly in mass attacks. They were particularly effective against cavalry and were hurled into the air, after lighting, or skimmed along the hard dry ground. Hyder Ali's son, Tippu Sultan, continued to develop and expand the use of rocket weapons, reportedly increasing the number of rocket troops from 1,200 to a corps of 5,000. In battles at Seringapatam in 1792 and 1799 these rockets were used with considerable effect against the British.

Congreve rockets

The news of the successful use of rockets spread through Europe. In England Sir William Congreve began to experiment privately. First, he experimented with a number of black-powder formulas and set down standard specifications of composition. He also standardized construction details and used improved production techniques. Also, his designs made it possible to choose either an explosive (ball charge) or incendiary warhead. The explosive warhead was separately ignited and could be timed by trimming the fuse length before launching. Thus, air bursts of the warheads were feasible at different ranges.

Congreve's metal rocket bodies were equipped on one side with two or three thin metal loops into which a long guide stick was inserted and crimped firm. Weights of eight different sizes of these rockets ranged up to 60 pounds. Launching was from collapsible A-frame ladders. In addition to aerial bombardment, Congreve's rockets were often fired horizontally along the ground.

These side-stick-mounted rockets were employed in a successful naval bombardment of the French coastal city of Boulogne in 1806. The next year a massed attack, using hundreds of rockets, burned most of Copenhagen to the ground. During the War of 1812 between the United States and the British, rockets were employed on numerous occasions. The two best-known engagements occurred in 1814. At the Battle of Bladensburg (August 24) the use of rockets assisted British forces to turn the flank of the American troops defending Washington, D.C. As a result, the British were able to capture the city. In September the British forces attempted to capture Fort McHenry, which guarded Baltimore harbour. Rockets were fired from a specially designed ship, the Erebus, and from small boats. The British were unsuccessful in their bombardment, but on that occasion Francis Scott Key, inspired by the sight of the night engagement, wrote "The Star Spangled Banner," later adopted as the United States national anthem. "The rockets' red glare" has continued to memorialize Congreve's rockets ever since.

In 1815 Congreve further improved his designs by mounting his guide stick along the central axis. The rocket's propulsive jet issued through five equally spaced holes rather than a single orifice. The forward portion of the guide stick, which screwed into the rocket, was sheathed with brass to prevent burning. The centre-stick-mounted rockets were significantly more accurate. Also, their design permitted launching from thin copper tubes.

Maximum ranges of Congreve rockets were from one-half mile to two miles (0.8 to 3.2 kilometres), depending upon size. They were competitive in performance and cost with the ponderous 10-inch mortar and were vastly more mobile.

Hale rockets

The next significant development in rocketry occurred about the middle of the 19th century. William Hale, a British engineer, invented a method of successfully eliminating the deadweight of the flight-stabilizing guide stick. By designing jet vents at an angle, he was able to spin the rocket. He developed various designs, including curved vanes that were acted upon by the rocket jet. These rockets, stabilized by means of spin, represented a major improvement in performance and ease of handling.

Even the new rockets, however, could not compete with the greatly improved artillery with rifled bores. The rocket corps of most European armies were dissolved, though rockets were still used in swampy or mountainous areas that were difficult for the much heavier mortars and guns. The Austrian Rocket Corps, using Hale rockets, won a number of engagements in mountainous terrain in Hungary and Italy. Other successful uses were by the Dutch colonial services in Celebes and by Russia in a number of engagements in the Turkistan War.

Hale sold his patent rights to the United States in time for some 2,000 rockets to be made for the Mexican War, 1846–48. Although some were fired, they were not particularly successful. Rockets were used in a limited way in the American Civil War (1861–65), but reports are fragmentary, and apparently they were not decisive. The U.S. Ordnance Manual of 1862 lists 16-pound Hale rockets with a range of 1.25 miles.

In Sweden about the turn of the century, Wilhelm Unge invented a device described as an "aerial torpedo." Based upon the stickless Hale rocket, it incorporated a number of design improvements. One of these was a rocket motor nozzle that caused the gas flow to converge and then diverge. Another was the use of smokeless powder based on nitroglycerin. Unge believed that his aerial torpedoes would

be valuable as surface-to-air weapons against dirigibles. Velocity and range were increased, and about 1909 the Krupp armament firm of Germany purchased the patents and a number of rockets for further experimentation.

World War I and after. In the United States, meanwhile, Robert Hutchings Goddard was conducting theoretical and experimental research on rocket motors at Worcester, Mass. Using a steel motor with a tapered nozzle, he achieved greatly improved thrust and efficiency. During World War I Goddard developed a number of designs of small military rockets to be launched from a lightweight hand launcher. By switching from black powder to double-base powder (40 percent nitroglycerin, 60 percent nitrocellulose), a far more potent propulsion charge was obtained. These rockets were proving successful under tests by the U.S. Army when the Armistice was signed; they became the forerunners of the bazooka of World War II.

World War I actually saw little use of rocket weapons, despite successful French incendiary antiballoon rockets and a German trench-war technique by which a grappling hook was thrown over enemy barbed wire by a rocket with a line attached.

Many researchers besides Goddard used the wartime interest in rockets to push experimentation, the most noteworthy being Elmer Sperry and his son, Lawrence, in the United States. The Sperrys worked on a concept of an "aerial torpedo," a pilotless airplane, carrying an explosive charge, that would utilize gyroscopic, automatic control to fly to a preselected target. Numerous flight attempts were made in 1917, some successful. Because of early interest in military use, the U.S. Army Signal Corps organized a separate program under Charles F. Kettering in Dayton, Ohio, late in 1918. The Kettering design used a gyroscope for lateral control to a preset direction and an aneroid barometer for pitch (fore and aft) control to maintain a preset altitude. A high angle of dihedral (upward tilt) in the biplane wings provided stability about the roll axis. The aircraft was rail-launched. Distance to target was determined by the number of revolutions of a propeller. When the predetermined number of revolutions had occurred, the wings of the airplane were dropped off and the aircraft carrying the bomb load dropped on the target.

Pilotless airplanes

The limited time available to attack the formidable design problems of these systems doomed the programs, and they never became operational.

As World War II approached, minor and varied experimental and research activities on rockets and guided missiles were underway in a number of countries. But in Germany, under great secrecy, the effort was concentrated. Successful flights as high as one mile were made in 1931–32 with gasoline–oxygen-powered rockets by the German Rocket Society. Funds for such amateur activities were scarce, and the society sought support from the German army. The work of Wernher von Braun, a member of the society, attracted the attention of Captain Walter R. Dornberger. Von Braun became the technical leader of a small group developing liquid-propellant rockets for the German army. By 1937 the Dornberger–Braun team, expanded to hundreds of scientists, engineers, and technicians, moved its operations from Kummersdorf to Peenemünde, a deserted area on the Baltic coast. Here the technology for a long-range ballistic missile was developed and tested (see below *Strategic missiles*).

World War II. World War II saw the expenditure of immense resources and talent for the development of rocket-propelled weapons.

Barrage rockets. The Germans began the war with a lead in this category of weapon, and their 150-millimetre and 210-millimetre bombardment rockets were highly effective. These were fired from a variety of towed and vehicle-mounted multitube launchers, from launching rails on the sides of armoured personnel carriers, and, for massive bombardments, even from their packing crates. Mobile German rocket batteries were able to lay down heavy and unexpected concentrations of fire on Allied positions. The 150-millimetre Nebelwerfer, a towed, six-tube launcher, was particularly respected by U.S. and British troops, to whom it was known as the "Screaming Meemie" or "Moaning Minnie" for the eerie sound made by the incoming rockets. Maximum range was more than 6,000 yards (5,500 metres).

The Nebelwerfer

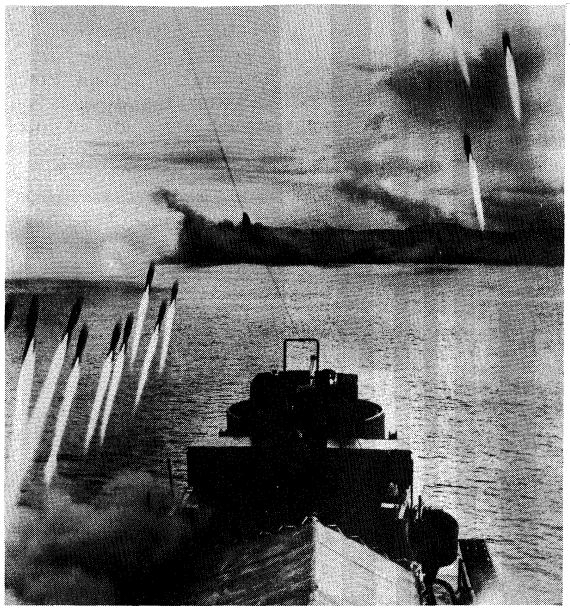


Figure 17: *Barrage rockets during the invasion of Mindoro, Phil., in December 1944.*
Launched in salvoes from landing craft, rockets smothered Japanese beach defenses as U.S. forces began the amphibious assault.
UPI/Bettmann Newsphotos

A five-inch rocket with an explosive warhead was developed in Great Britain. Its range was two to three miles. These rockets, fired from specially equipped naval vessels, were used in heavy coastal bombardment prior to landings in the Mediterranean. Firing rates were 800–1,000 in less than 45 seconds from each ship.

A development of the U.S. Army was the Calliope, a 60-tube launching projector for 4.5-inch rockets mounted on a Sherman tank. The launcher was mounted on the tank's gun turret, and both azimuth (horizontal direction) and elevation were controllable. Rockets were fired in rapid succession (ripple-fired) to keep the rockets from interfering with one another as they would in salvo firing.

Other conventional rockets developed in the United States included a 4.5-inch barrage rocket with a range of 1,100 yards and a five-inch rocket of longer range. The latter was used extensively in the Pacific theatre of war, fired from launching barges against shore installations, particularly just before landing operations (see Figure 17). The firing rate of these flat-bottom boats was 500 per minute. Other rockets were used for smoke laying and demolition. The United States produced more than four million of the 4.5-inch rockets and 15 million of the smaller bazooka rockets during the war.

As far as is known, Soviet rocket development during World War II was limited. Extensive use was made of barrage, ripple-fired rockets. Both A-frame and truck-mounted launchers were used. The Soviets mass-produced a 130-millimetre rocket known as the Katyusha. From 16 to 48 Katyushas were fired from a boxlike launcher known as the Stalin Organ, mounted on a gun carriage.

The bazooka. Beginning in mid-1940, Clarence N. Hickman, who had worked with Robert Goddard during World War I, supervised the development of a refined design of the hand-launched rocket. The new rocket, about 20 inches (50 centimetres) long, 2.36 inches in diameter, and weighing 3.5 pounds, was fired from a steel tube that became popularly known as the bazooka. Designed chiefly for use against tanks and fortified positions at short ranges (up to 600 yards), the bazooka surprised the Germans when it was first used in the North African landings of 1942. Although the rocket traveled slowly, it carried a potent shaped-charge warhead that gave infantrymen the striking power of light artillery.

The German counterpart of the bazooka was a light 88-millimetre rocket launcher known as Panzerschreck ("Tank Terror") or Ofenrohr ("Stovepipe").

Antiaircraft rockets. During World War II high-altitude bombing above the range of antiaircraft guns necessitated the development of rocket-powered weapons.

In Great Britain, initial effort was aimed at achieving the equivalent destructive power of the three-inch and later the 3.7-inch antiaircraft gun. Two important innovations were introduced by the British in connection with the three-inch rocket. One was a rocket-propelled aerial-defense system. A parachute and wire device was rocketed aloft, trailing a wire that unwound at high speed from a bobbin on the ground with the object of snagging the aircraft's propellers or shearing off the wings. Altitudes as high as 20,000 feet were attained. The other device was a type of proximity fuze using a photoelectric cell and thermionic amplifier. A change in light intensity on the photocell caused by light reflected from a nearby airplane (projected on the cell by means of a lens) triggered the explosive shell.

Wire-trailing rockets

The only significant antiaircraft rocket development by the Germans was the Taifun. A slender, six-foot, liquid-propellant rocket of simple concept, the Taifun was intended for altitudes of 50,000 feet. The design embodied coaxial tankage of nitric acid and a mixture of organic fuels, but the weapon never became operational.

Aerial rockets. Britain, Germany, the Soviet Union, Japan, and the United States all developed airborne rockets for use against surface as well as aerial targets. These were almost invariably fin-stabilized because of the effective aerodynamic forces when launched at speeds of 250 miles per hour and more. Tube launchers were used at first, but later straight-rail or zero-length launchers, located under the wings of the airplane, were employed.

One of the most successful of the German rockets was the 50-millimetre R4M. The tail fins remained folded until launch, facilitating close loading arrangements.

The U.S. achieved great success with a 4.5-inch rocket, three or four of which were carried under each wing of Allied fighter planes. These rockets were highly effective against motor columns, tanks, troop and supply trains, fuel and ammunition depots, airfields, and barges.

A variation on the airborne rocket was the addition of rocket motors and fins to conventional bombs. This had the effect of flattening the trajectory, extending the range, and increasing velocity at impact, useful against concrete bunkers and hardened targets. These weapons were called glide bombs, and the Japanese had 100-kilogram and 370-kilogram (225-pound and 815-pound) versions. The Soviet Union employed 25- and 100-kilogram versions, launched from the IL-2 Stormovik attack aircraft.

Postwar. After World War II, unguided, folding-fin rockets fired from multiple-tube pods became a standard air-to-ground munition for ground-attack aircraft and helicopter gunships. Though not as accurate as guided missiles or gun systems, they could saturate concentrations of troops or vehicles with a lethal volume of fire. Many ground forces continued to field truck-mounted, tube-launched rockets that could be fired simultaneously in salvos or ripple-fired in rapid succession (see Figure 18). Such artillery rocket systems, or multiple-launch rocket systems, generally fired rockets of 100 to 150 millimetres in diameter and had ranges of 12 to 18 miles. The rockets carried a variety of warheads, including high explosive, antipersonnel, incendiary, smoke, and chemical.

Ballistic rockets

The Soviet Union and the United States built unguided ballistic rockets for about 30 years after the war. In 1955 the U.S. Army began deployment of the Honest John in western Europe, and from 1957 the Soviet Union built a series of large, spin-stabilized rockets, launched from mobile transporters, given the NATO designation FROG (free rocket over ground). These missiles, from 25 to 30 feet long and two to three feet in diameter, had ranges of 20 to 45 miles and could be nuclear-armed. Egypt and Syria fired many FROG missiles during the opening salvos of the Arab–Israeli War of October 1973, as did Iraq in its war with Iran in the 1980s, but in the 1970s large rockets were phased out of the superpowers' front line in favour of inertially guided missiles such as the U.S. Lance and the Soviet SS-21 Scarab. (F.C.D.III/Ed.)

TACTICAL GUIDED MISSILES

Guided missiles were a product of post-World War II developments in electronics, computers, sensors, avionics, and, to only a slightly lesser degree, rocket and turbojet propulsion and aerodynamics. Although tactical, or battlefield, guided missiles were designed to perform many different roles, they were bound together as a class of weapon by similarities in sensor, guidance, and control systems.

Control methods

Control over a missile's direction was most commonly achieved by the deflection of aerodynamic surfaces such as tail fins; reaction jets or rockets and thrust-vectoring were also employed. But it was in their guidance systems that these missiles gained their distinction, since the ability to make down-course corrections in order to seek or "home" onto a target separated guided missiles from purely ballistic weapons such as free-flight rockets and artillery shells.

Guidance methods. The earliest guided missiles used simple command guidance, but within 20 years of World War II virtually all guidance systems contained autopilots or autostabilization systems, frequently in combination with memory circuits and sophisticated navigation sensors and computers. Five basic guidance methods came to be used, either alone or in combination: command, inertial, active, semiactive, and passive.

Command. Command guidance involved tracking the projectile from the launch site or platform and transmitting commands by radio, radar, or laser impulses or along thin wires or optical fibres. Tracking might be accomplished by radar or optical instruments from the launch site or by radar or television imagery relayed from the missile. The earliest command-guided air-to-surface and antitank munitions were tracked by eye and controlled by hand; later the naked eye gave way to enhanced optics and television tracking, which often operated in the infrared range and issued commands generated automatically by computerized fire-control systems. Another early command guidance method was beam riding, in which the missile sensed a radar beam pointed at the target and automatically corrected back to it. Laser beams were later used for the same purpose. Also using a form of command guidance were television-guided missiles, in which a small television camera mounted in the nose of the weapon beamed a picture of the target back to an operator who sent commands to keep the target centred in the tracking screen until impact. A form of command guidance used from the 1980s by the U.S. Patriot surface-to-air system was called track-via-missile. In this system a radar unit in the missile tracked the target and transmitted relative bearing and velocity information to the launch site, where control systems computed the optimal trajectory for intercepting the target and sent appropriate commands back to the missile.

Inertial. Inertial guidance was installed in long-range ballistic missiles in the 1950s, but, with advances in miniaturized circuitry, microcomputers, and inertial sensors, it became common in tactical weapons after the 1970s. Inertial systems involved the use of small, highly accurate gyroscopic platforms to continuously determine the position of the missile in space. These provided inputs to guidance computers, which used the position information in addition to inputs from accelerometers or integrating circuits to calculate velocity and direction. The guidance computer, which was programmed with the desired flight path, then generated commands to maintain the course.

Tactical uses of inertial guidance

An advantage of inertial guidance was that it required no electronic emissions from the missile or launch platform that could be picked up by the enemy. Many antiship missiles and some long-range air-to-air missiles, therefore, used inertial guidance to reach the general vicinity of their targets and then active radar guidance for terminal homing. Passive-homing antiradiation missiles, designed to destroy radar installations, generally combined inertial guidance with memory-equipped autopilots to maintain their trajectory toward the target in case the radar stopped transmitting.

Active. With active guidance, the missile would track

its target by means of emissions that it generated itself. Active guidance was commonly used for terminal homing. Examples were antiship, surface-to-air, and air-to-air missiles that used self-contained radar systems to track their targets. Active guidance had the disadvantage of depending on emissions that could be tracked, jammed, or tricked by decoys.

Semiactive. Semiactive guidance involved illuminating or designating the target with energy emitted from a source other than the missile; a seeker in the projectile that was sensitive to the reflected energy then homed onto the target. Like active guidance, semiactive guidance was commonly used for terminal homing. In the U.S. Hawk and Soviet SA-6 Gainful antiaircraft systems, for example, the missile homed in on radar emissions transmitted from the launch site and reflected off of the target, measuring the Doppler shift in the reflected emissions to assist in computing the intercept trajectory. (SA-6 Gainful is a designation given by NATO to the Soviet missile system. In this section, all Soviet missile systems and aircraft are referred to by their NATO designations.) The AIM-7 Sparrow air-to-air missile of the U.S. Air Force used a similar semiactive radar guidance method. Laser-guided missiles could also use semiactive methods by illuminating the target with a small spot of laser light and homing onto that precise light frequency through a seeker head in the missile.

With semiactive homing, the designator or illuminator might be remote from the launch platform. The U.S. Hellfire antitank missile, for example, used laser designation by an air or ground observer who could be situated many miles from the launching helicopter.

Passive. Passive guidance systems neither emitted energy nor received commands from an external source; rather, they "locked" onto an electronic emission coming from the target itself. The earliest successful passive homing munitions were "heat-seeking" air-to-air missiles that homed onto the infrared emissions of jet engine exhausts. The first such missile to achieve widespread success was the AIM-9 Sidewinder developed by the U.S. Navy in the early 1950s. Many later passive-homing air-to-air missiles homed onto ultraviolet radiation as well, using onboard guidance computers and accelerometers to compute optimal intercept trajectories. Among the most advanced passive-homing systems were optically tracking munitions that could "see" a visual or infrared image in much the same way as the human eye, memorize it by means of computer logic, and home onto it. Many passive homing systems required target identification and lock-on by a human operator prior to launch. With infrared antiaircraft missiles, a successful lock-on was indicated by an audible tone in the pilot's or operator's headset; with television or imaging infrared systems, the operator or pilot acquired the target on a video screen, which relayed information from the missile's seeker head, and then locked on manually.

"Heat-seeking" missiles

Passive guidance systems benefited enormously from a miniaturization of electronic components and from advances in seeker-head technology. Small, heat-seeking, shoulder-fired antiaircraft missiles first became a major factor in land warfare during the final stages of the Vietnam War, with the Soviet SA-7 Grail playing a major role in neutralizing the South Vietnamese Air Force in the final communist offensive in 1975. Ten years later the U.S. Stinger and British Blowpipe proved effective against Soviet aircraft and helicopters in Afghanistan, as did the U.S. Redeye in Central America.

Guided-missile systems. The principal categories of tactical guided missiles are antitank and assault, air-to-surface, air-to-air, antiship, and surface-to-air. Distinctions between these categories were not always clear, the launching of both antitank and infantry antiaircraft missiles from helicopters being a case in point.

Antitank and guided assault. One of the most important categories of guided missile to emerge after World War II was the antitank, or antiarmour, missile. The guided assault missile, for use against bunkers and structures, was closely related. A logical extension of unguided infantry antitank weapons carrying shaped-charge warheads for penetrating armour, guided antitank missiles acquired considerably more range and power than their shoulder-fired predecessors. While originally intended for issue to infantry formations for self-protection, the tactical flexibility and utility of guided antitank missiles led to their installation on light trucks, on armoured personnel carriers, and, most important, on antitank helicopters.

The first guided antitank missiles were controlled by electronic commands transmitted along extremely thin wires played out from a spool on the rear of the missile. Propelled by solid-fuel sustainer rockets, these missiles used aerodynamic fins for lift and control. Tracking was visual, by means of a flare in the missile's tail, and guidance commands were generated by a hand-operated joystick. In operating these missiles, the gunner simply superimposed the tracking flare on the target and waited for impact. The missiles were typically designed to be fired from their carrying containers, with the total package small enough to be carried by one or two men. Germany was developing weapons of this kind at the end of World War II and may have fired some in battle.

Early wire-guided missiles

After the war, French engineers adapted the German technology and developed the SS-10/SS-11 family of missiles. The SS-11 was adopted by the United States as an interim helicopter-fired antitank missile pending the development of the TOW (for tube-launched, optically tracked, wire-guided) missile. Because it was designed for greater range and hitting power, TOW was mounted primarily on vehicles and, particularly, on attack helicopters. Helicopter-fired antitank missiles were first used in combat when the U.S. Army deployed several TOW-equipped UH-1 "Hueys" to Vietnam in response to the 1972 communist Easter offensive. TOW was the principal U.S. antiarmour munition until Hellfire, a more sophisticated helicopter-fired missile with semiactive laser and passive infrared homing, was mounted on the Hughes AH-64 Apache attack helicopter in the 1980s.

The British Swingfire and the French-designed, internationally marketed MILAN (*missile d'infanterie léger antichar,* or "light infantry antitank missile") and HOT (*haut subsonique optiquement téléguidé tiré d'un tube,* or "high-subsonic, optically teleguided, tube-fired") were similar in concept and capability to TOW.

The Soviets developed an entire family of antitank guided missiles beginning with the AT-1 Snapper, the AT-2 Swatter, and the AT-3 Sagger. The Sagger, a relatively small missile designed for infantry use on the lines of the original German concept, saw use in Vietnam and was used with conspicuous success by Egyptian infantry in the Suez Canal crossing of the 1973 Arab-Israeli War. The AT-6 Spiral, a Soviet version of TOW and Hellfire, became the principal antiarmour munition of Soviet attack helicopters.

Many antitank missile systems of later generations transmitted guidance commands by radio rather than by wire, and semiactive laser designation and passive infrared homing also became common. Guidance and control methods were more sophisticated than the original visual tracking and manual commands. TOW, for example, required the gunner simply to centre the reticle of his optical sight on the target, and the missile was tracked and guided automatically. Extremely thin optical fibres began to replace wires as a guidance link in the 1980s.

Air-to-surface. The United States began to deploy tactical air-to-surface guided missiles as a standard aerial munition in the late 1950s. The first of these was the AGM-12 (for aerial guided munition) Bullpup, a rocket-powered weapon that employed visual tracking and radio-transmitted command guidance. The pilot controlled the missile by means of a small side-mounted joystick and guided it toward the target by observing a small flare in its tail. Though Bullpup was simple and accurate, the delivery aircraft had to continue flying toward the target until the weapon struck—a vulnerable maneuver. The 250-pound (115-kilogram) warhead on the initial version of Bullpup proved inadequate for "hard" targets such as reinforced concrete bridges in Vietnam, and later versions had a 1,000-pound warhead. The rocket-powered AGM-45 Shrike antiradiation missile was used in Vietnam to attack enemy radar and surface-to-air sites by passively

Antiradar missiles

homing onto their radar emissions. The first missile of its kind used in combat, the Shrike had to be tuned to the desired radar frequency before flight. Because it had no memory circuits and required continuous emissions for homing, it could be defeated by simply turning off the target radar. Following the Shrike was the AGM-78 Standard ARM (antiradiation munition), a larger and more expensive weapon that incorporated memory circuits and could be tuned to any of several frequencies in flight. Also rocket-propelled, it had a range of about 35 miles (55 kilometres). Faster and more sophisticated still was the AGM-88 HARM (high-speed antiradiation missile), introduced into service in 1983.

Replacing the Bullpup as an optically tracked missile was the AGM-64/65 Maverick family of rocket-powered missiles. Early versions used television tracking, while later versions employed infrared, permitting the fixing of targets at longer ranges and at night. The self-contained guidance system incorporated computer logic that enabled the missile to lock onto an image of the target once the operator had identified it on his cockpit television monitor. Warheads varied from a 125-pound shaped charge for use against armour to high-explosive blast charges of 300 pounds.

Though less was known about them, the Soviets fielded an extensive array of air-to-surface missiles equivalent to the Bullpup and Maverick and to the Hellfire antitank missile. Notable among these was the radio-command-guided AS-7 Kerry, the antiradar AS-8 and AS-9, and the television-guided AS-10 Karen and AS-14 Kedge (the last with a range of about 25 miles). These missiles were fired from tactical fighters such as the MiG-27 Flogger and attack helicopters such as the Mi-24 Hind and Mi-28 Havoc.

Air-to-air. Developed in 1947, the radar-guided, subsonic Firebird was the first U.S. guided air-to-air missile. It was rendered obsolete within a few years by supersonic missiles such as the AIM-4 (for air-intercept missile) Falcon, the AIM-9 Sidewinder, and the AIM-7 Sparrow. The widely imitated Sidewinder was particularly influential. Early versions, which homed onto the infrared emissions from jet engine tailpipes, could approach only from the target's rear quadrants. Later versions, beginning with the AIM-9L, were fitted with more sophisticated seekers sensitive to a broader spectrum of radiation. These gave the missile the capability of sensing exhaust emissions from the side or front of the target aircraft (see Figure 41). Driven by the requirements of supersonic combat during the 1960s, the ranges of such missiles as the Sidewinder increased from about two miles to 10–15 miles. The AIM-54 Phoenix, a semiactive radar missile with active radar terminal homing introduced by the U.S. Navy in 1974, was capable of ranges in excess of 100 miles. Fired from the F-14 Tomcat, it was controlled by an acquisition, tracking, and guidance system that could engage up to six targets simultaneously. Combat experience in Southeast Asia and the Middle East produced increased tactical sophistication, so that fighter aircraft were routinely armed with several kinds of missile to deal with a variety of situations. U.S. carrier-based fighters, for instance, carried both heat-seeking Sidewinders and radar-homing Sparrows. Meanwhile, the Europeans developed such infrared-homing missiles as the British Red Top and the French Magic, the latter being a short-range (one-quarter to four miles) highly maneuverable equivalent of the Sidewinder.

The Sidewinder

The Soviets fielded an extended series of air-to-air missiles, beginning in the 1960s with the AA-1 Alkali, a relatively primitive semiactive radar missile, the AA-2 Atoll, an infrared missile closely modeled after the Sidewinder, and the AA-3 Anab, a long-range, semiactive radar-homing missile carried by air-defense fighters. The AA-5 Ash was a large, medium-range radar-guided missile, while the AA-6 Acrid was similar to the Anab but larger and with greater range. The AA-7 Apex, a Sparrow equivalent, and the AA-8 Aphid, a relatively small missile for close-in use, were introduced during the 1970s. Both used semiactive radar guidance, though the Aphid was apparently produced in an infrared-homing version as well. The long-range, semiactive radar-guided AA-9 Amos appeared in the mid-1980s; it was associated with the MiG-31 Foxhound interceptor, much as the U.S. Phoenix was associated with the F-14. The Foxhound/Amos combination may have been fitted with a look-down/shoot-down capability, enabling it to engage low-flying targets while looking downward against a cluttered radar background. The AA-10 Alamo, a medium-range missile similar to the Amos, apparently had passive radar guidance designed to home onto carrier-wave emissions from U.S. aircraft firing the semiactive radar-homing Sparrow. The AA-11 Archer was a short-range missile used in combination with the Amos and Alamo.

Improvements in air-to-air missiles included the combined use of several methods of guidance for greater flexibility and lethality. Active radar or infrared terminal homing, for example, were often used with semiactive radar guidance in midcourse. Also, passive radar homing, which became an important means of air-to-air guidance, was backed up by inertial guidance for mid-course and by an alternate terminal homing method in case the target aircraft shut off its radar. Sophisticated optical and laser proximity fuzes became common; these were used with directional warheads that focused their blast effects toward the target. Tactical demands combined with advancing technology to channel the development of air-to-air missiles into three increasingly specialized categories: large, highly sophisticated long-range air-intercept missiles, such as the Phoenix and Amos, capable of ranges from 40 to 125 miles; short-range, highly maneuverable (and less expensive) "dogfighter" missiles with maximum ranges of six to nine miles; and medium-range missiles, mostly using semiactive radar homing, with maximum ranges of 20 to 25 miles. Representative of the third category was the AIM-120 AMRAAM (for advanced medium-range air-to-air missile), jointly developed by the U.S. Air Force and Navy for use with NATO aircraft. AMRAAM combined inertial mid-course guidance with active radar homing (see Figure 41).

Increasing combat range

Antiship. Despite their different methods of delivery, antiship missiles formed a coherent class largely because they were designed to penetrate the heavy defenses of warships.

The Hs-293 missiles developed by Germany during World War II were the first guided antiship missiles. Though accurate, they required the delivery aircraft to stay on the same line of sight as the weapon and target; the resultant flight paths were predictable and highly vulnerable, and the Allies quickly developed effective defenses.

Partly because Britain and the United States relied on carrier-based aircraft armed with conventional torpedoes, bombs, and unguided rockets to attack naval targets, antiship missiles at first received little emphasis in the West after the war. The Soviets, however, saw antiship missiles as a counter to Western naval superiority and developed an extensive range of air- and surface-launched antiship missiles, beginning with the AS-1 Kennel. The destruction of an Israeli destroyer by two SS-N-2 Styx missiles fired by Soviet-supplied Egyptian missile boats in October 1967 demonstrated the effectiveness of the Soviet systems, and the Western powers developed their own guided missiles. The resultant systems began entering service in the 1970s and first saw combat in 1982, during the Falkland Islands War. In that conflict the British Sea Skua, a small, rocket-powered, sea-skimming missile with semiactive radar homing, weighing about 325 pounds, was fired successfully from helicopters, while the Argentines sank a destroyer and a containership and damaged another destroyer with the solid-rocket-powered, active radar-homing French Exocet, fired from both aircraft and ground launchers. The Exocet weighed about 1,500 pounds and had an effective range of 35 to 40 miles.

The Exocet was one of a number of Western antiship missiles of the same general kind. Guidance was mostly by active radar, often supplemented in mid-course by inertial autopilots and in terminal flight by passive radar and infrared homing. Although designed for use from carrier-based attack aircraft, missiles of this sort were also carried by bombers and coastal patrol aircraft and were mounted on ship- and land-based launchers. The most important U.S. antiship missile was the turbojet-powered Harpoon,

The Harpoon and Sea Eagle

which weighed about 1,200 pounds in its air-launched version and had a 420-pound warhead. Employing both active and passive radar homing, this missile could be programmed for sea-skimming attack or a "pop-up and dive" maneuver to evade a ship's close-in defense systems. The turbojet-powered British Sea Eagle weighed somewhat more than the Harpoon and employed active radar homing. The West German Kormoran, somewhat smaller than the Exocet, was also an air-launched missile. The Norwegian Penguin, a rocket-powered missile weighing between 700 and 820 pounds and employing technology derived from the U.S. Maverick air-to-surface missile, had a range of about 17 miles and supplemented its active radar guidance with passive infrared homing. The Penguin was exported widely for fighter-bomber, attack boat, and helicopter use. The Israeli Gabriel, a 1,325-pound missile with a 330-pound warhead launched from both aircraft and ships, employed active radar homing and had a range of 20 miles.

The U.S. Navy Tomahawk defined a separate category of antiship missile: it was a long-range, turbofan-powered cruise missile first developed as a strategic nuclear delivery system (see below *Strategic missiles*). Tomahawk was carried by surface vessels and submarines in both ground-attack and antiship versions. The antiship version, equipped with a modified Harpoon guidance system, had a range of 275 miles. Only 20 feet long and 20.5 inches (53 centimetres) in diameter, the Tomahawk was fired from its launch tubes by a solid-fueled booster and cruised at subsonic speeds on flip-out wings.

For short-range antiship warfare, the Soviet Union deployed its AS series, 7, 8, 9, 10, and 14 air-to-surface missiles. Long-range antiship missiles designed for use from bomber and patrol aircraft included the 50-foot, swept-wing AS-3 Kangaroo, introduced in 1961 with a range exceeding 400 miles. The AS-4 Kitchen, a Mach-2 (twice the speed of sound) rocket-powered missile with a range of about 250 miles, was also introduced in 1961, and the liquid-fuel rocket-powered, Mach-1.5 AS-5 Kelt was first deployed in 1966. The Mach-3 AS-6 Kingfish, introduced in 1970, could travel 250 miles.

Ship-based Soviet systems included the SS-N-2 Styx, a subsonic aerodynamic missile first deployed in 1959–60 with a range of 25 miles, and the SS-N-3 Shaddock, a much larger system resembling a swept-wing fighter aircraft with a range of 280 miles. The SS-N-12 Sandbox, introduced in the 1970s on the Kiev-class antisubmarine carriers, was apparently an improved Shaddock. The SS-N-19 Shipwreck, a small, vertically launched, flip-out wing supersonic missile with a range of about 390 miles, appeared in the 1980s.

Antimissile defense

To defend against antiship missiles, navies employed towed or helicopter-borne decoys. Sometimes chaff (strips of foil or clusters of fine glass or wire) would be released in the air to create false radar targets. Defenses included long-range chaff rockets to mask a vessel from the radar of distant ships, close-in quick-blooming chaff flares to confuse active radar homers on missiles, and radar jamming to defeat acquisition and tracking radars and confuse missile seeker systems. For close-in defense, combatant ships were fitted with high-performance, short-range missiles such as the British Seawolf and automatic gun systems such as the U.S. 20-millimetre Phalanx. Advances in missile-defense systems had to keep up with the natural affinity of antiship missiles for stealth technology: the visual and infrared signatures and radar cross sections of Western antiship missiles, in particular, became so small that relatively minor modifications in shape and modest applications of radar-absorptive materials could make them very difficult to detect with radar and electro-optical systems, except at very short ranges.

Surface-to-air. Guided surface-to-air missiles, or SAMs, were under development when World War II ended, notably by the Germans, but were not sufficiently perfected to be used in combat. This changed in the 1950s and '60s with the rapid development of sophisticated SAM systems

By courtesy of (top right) British Aerospace (Dynamics) Limited; photographs, (left) Tass from Sovfoto, (bottom right) John F. Burns/NYT Pictures

Figure 18: *Three mobile surface-to-air missiles, for low- and medium-level antiaircraft defense.* (Left) The Soviet SA-6 Gainful, tracked and guided by radar, on maneuvers with armoured personnel carriers and truck-mounted artillery rockets. (Top right) The optically tracked and radio-guided version of the British Rapier, as deployed in the Falkland Islands. (Bottom right) The infrared-homing U.S. Stinger, in the hands of an Afghan guerrilla fighter.

in the Soviet Union, the United States, Great Britain, and France. With other industrialized nations following suit, surface-to-air missiles of indigenous design, particularly in the smaller categories, were fielded by many armies and navies.

The Soviet Union committed more technical and fiscal resources to the development of guided-missile air-defense systems than any other nation. Beginning with the SA-1 Guild, developed in the immediate postwar period, the Soviets steadily fielded SAMs of growing sophistication. These fell into two categories: systems such as the Guild, the SA-3 Goa, the SA-5 Gammon, and the SA-10 Grumble, which were deployed in defense of fixed installations; and mobile tactical systems capable of accompanying land forces. Most of the tactical systems had naval versions. The SA-2 Guideline, introduced in 1958, was the most widely deployed of the early SAMs and was the first surface-to-air guided-missile system used in combat. This two-stage missile with a solid booster and a liquid-propellant (kerosene and nitric acid) sustainer, could engage targets at ranges of 28 miles and as high as 60,000 feet. Equipped with an array of van-mounted radars for target acquisition and tracking and for missile tracking and command guidance, Guideline proved effective in Vietnam. With adequate warning, U.S. fighters could outmaneuver the relatively large missiles, called "flying telephone poles" by pilots, and electronic countermeasures (ECM) reduced the effectiveness of the tracking radars; but, while these SAMs inflicted relatively few losses, they forced U.S. aircraft down to low altitudes, where antiaircraft artillery and small arms exacted a heavy toll. Later versions of the SA-2 were equipped with optical tracking to counter the effects of ECM; this became a standard feature on SAM systems. After retirement from first-line Soviet service, the SA-2 remained in use in the Third World.

The SA-3 Goa, derived from the Guideline but modified for use against low-altitude targets, was first deployed in 1963—primarily in defense of fixed installations. The SA-N-1 was a similar naval missile.

The SA-4 Ganef was a long-range mobile system first deployed in the mid-1960s; the missiles, carried in pairs on a tracked launcher, used drop-off solid-fuel boosters and a ramjet sustainer motor. Employing a combination of radar command guidance and active radar homing, and supported by an array of mobile radars for target acquisition, tracking, and guidance, they could engage targets over the horizon. (Because the SA-4 strongly resembled the earlier British Bloodhound, NATO assigned it the code name Ganef, meaning "Thief" in Hebrew.) Beginning in the late 1980s, the SA-4 was replaced by the SA-12 Gladiator, a more compact and capable system.

The SA-5 Gammon was a high- and medium-altitude strategic missile system with a range of 185 miles; it was exported to Syria and Libya.

The SA-6 Gainful

The SA-6 Gainful was a mobile tactical system with a range of two to 35 miles and a ceiling of 50,000 feet. Three 19-foot missiles were carried in canisters atop a tracked transporter-erector-launcher, or TEL (see Figure 18), and the radar and fire-control systems were mounted on a similar vehicle, each of which supported four TELs. The missiles used semiactive radar homing and were powered by a combination of solid-rocket and ramjet propulsion. (The SA-N-3 Goblet was a similar naval system.) Gainful, the first truly mobile land-based SAM system, was first used in combat during the 1973 Arab-Israeli War and was highly effective at first against Israeli fighters. The Mach-3 missile proved virtually impossible to outmaneuver, forcing the fighters to descend below effective radar coverage, where antiaircraft guns such as the ZSU 23-4 mobile system were particularly lethal. (Similar factors prevailed in the 1982 Falklands conflict, where long-range British Sea Dart missiles achieved relatively few kills but forced Argentine aircraft down to wave-top level.) The SA-6 was replaced by the SA-11 Gadfly beginning in the 1980s.

The SA-8 Gecko, first deployed in the mid-1970s, was a fully mobile system mounted on a novel six-wheeled amphibious vehicle. Each vehicle carried four canister-launched, semiactive radar homing missiles, with a range of about 7.5 miles, plus guidance and tracking equipment in a rotating turret. It had excellent performance but, in Syrian hands during the 1982 conflict in Lebanon, proved vulnerable to Israeli electronic countermeasures. The equivalent naval system was the widely deployed SA-N-4 Goblet.

The SA-7 Grail shoulder-fired, infrared-homing missile was first deployed outside the Soviet Union in the final stages of the Vietnam War; it also saw extensive action in the Middle East. The SA-9 Gaskin carried four infrared-homing missiles on a turreted mount atop a four-wheeled vehicle. Its missiles were larger than the SA-7 and had more sophisticated seeker and guidance systems.

The first generation of American SAMs included the Army Nike Ajax, a two-stage, liquid-fueled missile that became operational in 1953, and the rocket-boosted, ramjet-powered Navy Talos. Both used radar tracking and target acquisition and radio command guidance. The later Nike Hercules, also command-guided, had a range of 85 miles. After 1956 the Talos was supplemented by the Terrier, a radar-beam rider, and the Tartar, a semiactive radar homing missile. These were replaced in the late 1960s by the Standard semiactive radar homing system.

The Standard

The solid-fueled, Mach-2 Standard missiles were deployed in medium-range (MR) and two-stage extended-range (ER) versions capable, respectively, of about 15 miles and 35 miles. Within 10 years a second generation of Standard missiles doubled the range of both versions. These newer missiles contained an inertial-guidance system that, by electronically communicating with the Aegis radar fire-control system, allowed corrections to be made in midcourse before the semiactive terminal homing took over (see Figure 32).

For 20 years, the most important land-based American SAM was the Hawk, a sophisticated system employing semiactive radar guidance. From the mid-1960s the Hawk provided the backbone of U.S. surface-based air defenses in Europe and South Korea and was exported to many allies. In Israeli use, Hawk missiles proved highly effective against low-flying aircraft. The longer-ranged Patriot missile system began entering service in 1985 as a partial replacement for the Hawk. Like the Hawk, the Patriot was semimobile; that is, the system components were not mounted permanently on vehicles and so had to be removed from their transport for firing. For target acquisition and identification, as well as for tracking and guidance, the Patriot system used a single phased-array radar, which controlled the direction of the beam by electronically varying the signals at several antennas rather than pivoting a single large antenna. The single-stage, solid-fueled Patriot missile was controlled by command guidance and employed track-via-missile homing, in which information from the radar in the missile itself was used by the launch site fire-control system.

The shoulder-fired Redeye, an infrared-homing missile that was also deployed on truck-mounted launchers, was fielded in the 1960s to provide U.S. Army units close-in protection against air attack. After 1980 the Redeye was replaced by the Stinger (see Figure 18), a lighter system whose missile accelerated faster and whose more advanced seeker head could detect the hot exhaust of approaching aircraft even four miles away and up to 5,000 feet in altitude.

Western European mobile SAM systems include the German-designed Roland, an SA-8 equivalent fired from a variety of tracked and wheeled vehicles, and the French Crotale, an SA-6 equivalent that used a combination of radar command guidance and infrared terminal homing. Both systems were widely exported. Less directly comparable to Soviet systems was the British Rapier, a short-range, semimobile system intended primarily for airfield defense.

The Rapier

The Rapier missile was fired from a small, rotating launcher that was transported by trailer. In the initial version, deployed in the early 1970s and used with some success in 1982 in the Falklands conflict (see Figure 18), the target aircraft was tracked by a gunner using an optical sight. A television camera in the tracker measured differences between the missile's flight path and the path to the target, and microwave radio signals issued guidance corrections. The Rapier had a combat range of one-

quarter to four miles and a ceiling of 10,000 feet. Later versions used radar tracking and guidance for all-weather engagements.

A new generation of Soviet SAM systems entered service in the 1980s. These included the SA-10 Grumble, a Mach-6 mobile system with a 60-mile range deployed in both strategic and tactical versions; the SA-11 Gadfly, a Mach-3 semiactive radar homing system with a range of 17 miles; the SA-12 Gladiator, a track-mobile replacement of Ganef; the SA-13 Gopher, a replacement for Gaskin; and the SA-14, a shoulder-fired Grail replacement. Both Grumble and Gadfly had naval equivalents, the SA-N-6 and SA-N-7. The Gladiator might have been designed with an antimissile capability, making it an element of the antiballistic missile defense around Moscow. (J.F.G.)

STRATEGIC MISSILES

Strategic missiles represent a logical step in the attempt to attack enemy forces at a distance. As such, they can be seen as extensions of either artillery (in the case of ballistic missiles) or manned aircraft (in the case of cruise missiles). Ballistic missiles are rocket-propelled weapons that travel by momentum in a high, arcing trajectory after they have been launched into flight by a brief burst of power. Cruise missiles, on the other hand, are powered continuously by air-breathing jet engines and are sustained along a low, level flight path by aerodynamic lift.

Although experiments were undertaken before World War II on crude prototypes of the cruise and ballistic missile, the modern weapons are generally considered to have their true origins in the V-1 and V-2 missiles launched by Germany in 1944–45. Both of those Vergeltungswaffen, or "Vengeance Weapons," defined the problems of propulsion and guidance that have continued ever since to shape cruise and ballistic missile development.

Long-range missiles and nuclear warheads

Given the extremely long ranges required of strategic weapons, even the most modern guidance systems cannot deliver a missile's warhead to the target with consistent, pinpoint accuracy. For this reason, strategic missiles have almost exclusively carried nuclear warheads, which need not strike a target directly in order to destroy it. By contrast, missiles of shorter range (often called tactical- or battlefield-range) have been fitted with both nuclear and conventional warheads. For example, the SS-1 Scud, a ballistic missile with ranges of up to 185 miles (300 kilometres), was fielded with nuclear warheads by Soviet troops in eastern Europe from the 1950s through the 1980s; but in the "war of the cities" during the Iran–Iraq conflict of the 1980s, many SS-1s armed with conventional warheads were launched by both sides, killing thousands of civilians. Other "dual-capable" short-range ballistic missiles are the U.S. Lance, with a range of about 80 miles, and the Soviet SS-21 Scarab, with a range of 75 miles. (In this section, Soviet missile systems are referred to by their NATO designations, such as SS-21 Scarab.)

The exclusively nuclear capacity of strategic-range weapons has confined serious development of cruise and ballistic missile technology to the world's nuclear powers—particularly the United States and the Soviet Union. These two countries have taken different paths in exploiting missile technology. Soviet cruise missiles, for instance, have been designed mostly for tactical antiship use rather than for threatening strategic land targets (as has been the U.S. emphasis). Throughout the ballistic missile arms race, the United States has tended to streamline its weapons, seeking greater accuracy and lower explosive power, or yield. Meanwhile, the Soviet Union, perhaps to make up for its difficulties in solving guidance problems, has concentrated on larger missiles and higher yields. Most U.S. systems have carried warheads of less than one megaton, with the largest being the nine-megaton Titan II, in service from 1963 through 1987. The Soviet warheads have often exceeded five megatons, with the largest being a 20- to 25-megaton warhead deployed on the SS-7 Saddler from 1961 to 1980 and a 25-megaton warhead on the SS-9 Scarp, deployed from 1967 to 1982. (For the development of nuclear weapons, see below *Nuclear weapons.*)

Most other countries pursuing missile technology have not developed strategic weapons to the extent of the United States and the Soviet Union. Nonetheless, several other nations have produced them; their emphasis, however, has been on ballistic rather than cruise missiles, because of the extremely sophisticated guidance systems required of cruise missiles. Also, as with any technology, there has occurred a transfer of ballistic missile technology to less-developed countries. Combined with the widespread capacity to produce chemical warheads, such weapons represent a potent addition to the arsenals of emerging powers of the Third World.

Ballistic missiles. Strategic ballistic missiles can be divided into two general categories according to their basing mode: those that are launched from land and those launched at sea (from submarines beneath the surface). They can also be divided according to their range into intermediate-range ballistic missiles (IRBMs) and intercontinental ballistic missiles (ICBMs). IRBMs have ranges of about 600 to 3,500 miles, while ICBMs have ranges exceeding 3,500 miles. Modern land-based strategic missiles are almost all of ICBM range, whereas all but the most modern submarine-launched ballistic missiles (SLBMs) have been of intermediate range.

IRBMs and ICBMs

Prelaunch survivability (that is, the ability to survive an enemy attack) has been a long-standing problem with land-based ICBMs. (SLBMs achieve survivability by being based on relatively undetectable submarines.) At first, they were considered safe from attack because neither U.S. nor Soviet missiles were sufficiently accurate to strike the other's launch sites; hence, early systems were launched from above ground. However, as missile accuracies improved, above-ground missiles became vulnerable, and in the 1960s both countries began to base their ICBMs below ground in concrete tubes called silos, some of which were hardened against nuclear blast. Later, even greater improvements in accuracy brought ICBM basing strategy back to above-ground systems. This time, prelaunch survivability was to be achieved by mobile ICBMs that would confound an attacker with multiple moving targets.

Most U.S. silos are designed for one-time "hot-launch" use, the rocket engines igniting within the silo and essentially destroying it as the missile departs. The Soviets pioneered the "cold-launch" method, in which the missile is expelled by gas and the rocket engine ignited after the missile clears the silo. This method, essentially the same system used with SLBMs, allows silos to be reused after minor repair.

In order to increase their range and throw weight, ballistic missiles are usually multistaged. By shedding weight as the flight progresses (that is, by burning the fuel and then discarding the pumps, flight controls, and associated equipment of the previous stage), each successive stage has less mass to accelerate. This permits a missile to fly farther and carry a larger payload.

Boost, midcourse, and reentry phases

The flight path of a ballistic missile has three successive phases. In the first, called the boost phase, the rocket engine (or engines, if the missile contains two or three stages) provides the precise amount of propulsion required to place the missile on a specific ballistic trajectory. Then the engine quits, and the final stage of the missile (called the payload) coasts in the mid-course phase, usually beyond the Earth's atmosphere. The payload contains the warhead (or warheads), the guidance system, and such penetration aids as decoys, electronic jammers, and chaff to help elude enemy defenses. The weight of this payload constitutes the missile's throw weight, that is, the total weight that the missile is capable of placing on a ballistic trajectory toward a target. By mid-course, the warheads have detached from the remainder of the payload, and all elements are on a ballistic path. The terminal phase of flight occurs when gravity pulls the warheads (now referred to as the reentry vehicles, or RVs) back into the atmosphere and down to the target area.

Most ballistic missiles use inertial guidance to arrive at the vicinity of their targets. This technology, based on Newtonian physics, involves measuring disturbances to the missile in three axes. The device used to measure these disturbances is usually composed of three gyroscopically stabilized accelerometers mounted at right angles to one another. By calculating the acceleration imparted by ex-

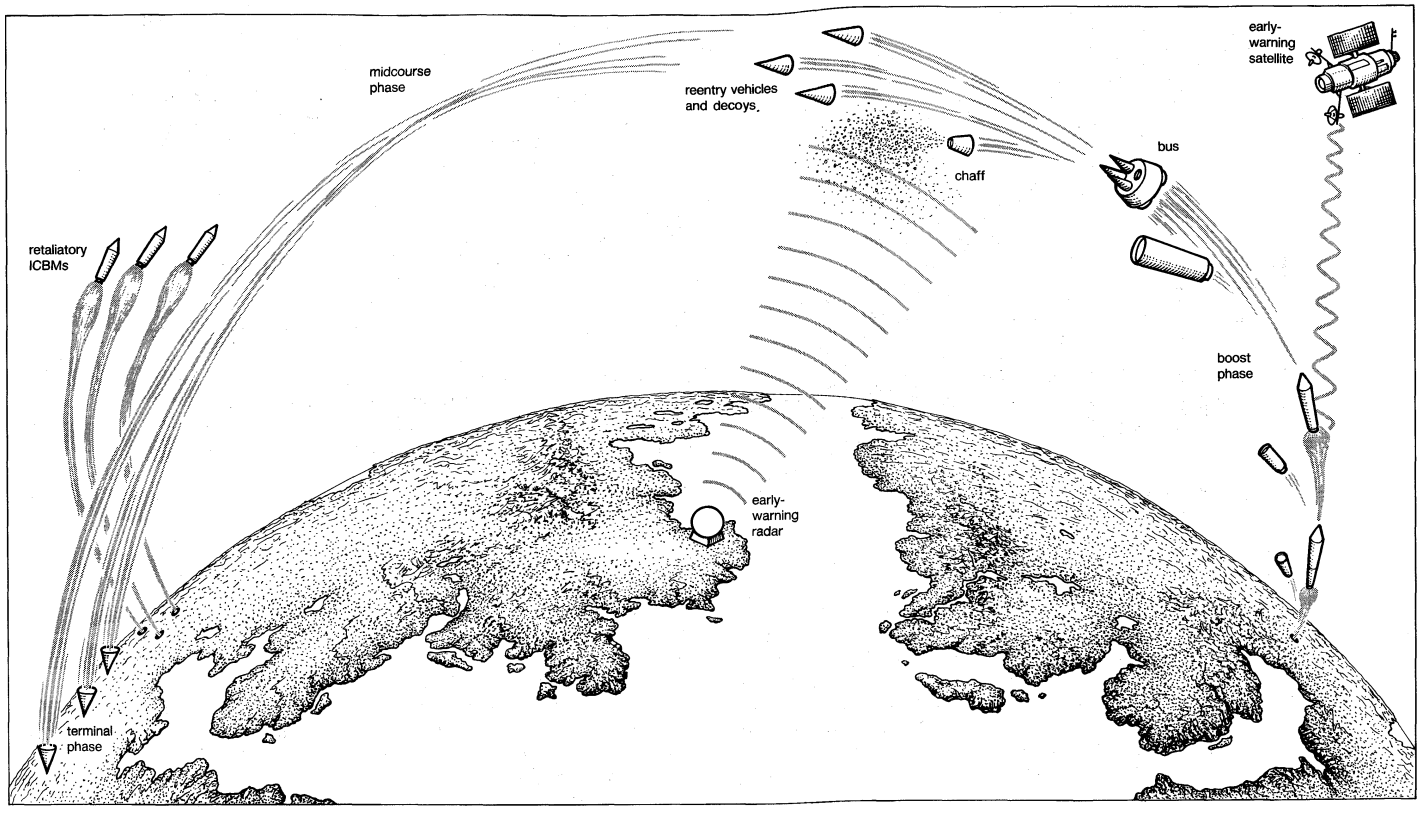


Figure 19: *Flight path of an intercontinental ballistic missile (ICBM).* While being boosted into a ballistic trajectory beyond the Earth's atmosphere, the missile's final stage, called the bus, would dispense its nuclear warheads, or reentry vehicles. Decoys and a cloud of reflective chaff would confuse the enemy's radar-guided defenses, and the reentry vehicles would strike their targets within 30 minutes of launch. By that time, retaliatory ICBMs would probably be on their way.

From L. Freedman, *Atlas of Global Strategy*, copyright © Equinox (Oxford) Ltd 1985

ternal forces (including the rocket engine's thrust), and by comparing these forces to the launch position, the guidance system can determine the missile's position, velocity, and heading. Then the guidance computer, predicting the gravitational forces that will act on the reentry vehicle, can calculate the velocity and heading required to reach a predetermined point on the ground. Given these calculations, the guidance system can issue a command to the missile thrust system during boost phase to place the payload at a specific point in space, on a specific heading, and at a specific velocity—at which point thrust is shut off and a purely ballistic flight path begins.

Ballistic missile guidance is complicated by two factors. First, during the latter stages of the powered boost phase, the atmosphere is so thin that aerodynamic flight controls such as fins cannot work and the only corrections that can be made to the flight path must come from the rocket engines themselves. But, because the engines only provide a force vector roughly parallel to the missile's fuselage, they cannot be used to provide major course corrections; making major corrections would create large gravitational forces perpendicular to the fuselage that could destroy the missile. Nevertheless, small corrections can be made by slightly gimballing the main engines so that they swivel, by placing deflective surfaces called vanes within the rocket exhaust, or, in some instances, by fitting small rocket engines known as thrust-vector motors or thrusters. This technique of introducing small corrections into a missile's flight path by slightly altering the force vector of its engines is known as thrust-vector control.

Thrust-vector control

A second complication occurs during reentry to the atmosphere, when the unpowered RV is subject to relatively unpredictable forces such as wind. Guidance systems have had to be designed to accommodate these difficulties.

Errors in accuracy for ballistic missiles (and for cruise missiles as well) are generally expressed as launch-point errors, guidance/en-route errors, or aim-point errors. Both launch- and aim-point errors can be corrected by surveying the launch and target areas more accurately. Guidance/en-route errors, on the other hand, must be corrected by improving the missile's design—particularly its guidance. Guidance/en-route errors are usually measured by a missile's circular error of probability (CEP) and bias. CEP uses the mean point of impact of missile test firings, usually taken at maximum range, to calculate the radius of a circle that would take in 50 percent of the impact points. Bias measures the deviation of the mean impact point from the actual aim point. An accurate missile has both a low CEP and low bias.

The V-2. The precursor of modern ballistic missiles was the German V-2, a single-stage, fin-stabilized missile propelled by liquid oxygen and ethyl alcohol to a maximum range of about 200 miles (see Figure 20). The V-2 was officially designated the A-4, being derived from the fourth of the *Aggregat* series of experiments conducted at Kummersdorf and Peenemunde under General Walter Dornberger and the civilian scientist Wernher von Braun.

The most difficult technical problem facing the V-2 was achieving maximum range. An inclined launch ramp was normally used to give missiles maximum range, but this could not be used with the V-2 because the missile was quite heavy at lift-off (more than 12 tons) and would not be traveling fast enough to sustain anything approaching horizontal flight. Also, as the rocket used up its fuel its weight (and velocity) would change, and this had to be allowed for in the aiming. For these reasons the V-2 had to be launched straight up and then had to change to the flight angle that would give it maximum range. The Germans calculated this angle to be slightly less than 50°.

Early problems in flight control

The change in direction mandated some sort of pitch control during flight, and, because a change in pitch would induce yaw, control was needed on the yaw axis too. Added to these problems was the natural tendency of a cylinder to rotate. Thus, the V-2 (and every ballistic missile afterward) needed a guidance and control system to deal with in-flight rolling, pitching, and yawing. Using

Figure 20: *German V-2 missile being erected for launch.* This "Vengeance Weapon" of World War II, which struck cities 200 miles away, was the precursor of modern ballistic missiles.
Bilderdienst im Suddeutschen Verlag

three-axis autopilots adapted from German aircraft, the V-2 was controlled by large vertical fins and smaller stabilizing surfaces to dampen roll and by vanes attached to the horizontal fins to modify pitch and yaw. Vanes were also installed in the exhaust nozzle for thrust vector control.

A combination of in-flight weight changes and changes in atmospheric conditions presented additional problems. Even over the fairly limited course of a V-2 trajectory (with a range of approximately 200 miles and an altitude of roughly 50 miles), changes in missile velocity and air density produced drastic shifts in the distance between the centre of gravity and the centre of aerodynamic pressure. This meant the guidance system had to adjust its input to the control surfaces as the flight proceeded. As a result, V-2 accuracy never ceased to be a problem for the Germans.

Still, the missile caused a great deal of damage. The first V-2 used in combat was fired against Paris on Sept. 6, 1944. Two days later the first of more than 1,000 missiles was fired against London. By the end of the war 4,000 of these missiles had been launched from mobile bases against Allied targets. During February and March 1945, only weeks before the war in Europe ended, an average of 60 missiles was launched weekly. The V-2 killed an estimated five persons per launch (versus slightly more than two per launch for the V-1). Three major factors contributed to this difference. First, the V-2 warhead weighed more than 1,600 pounds (725 kilograms). Second, several V-2 attacks killed more than 100 people. Finally, there was no known defense against the V-2; it could not be intercepted and, traveling faster than sound, it arrived unexpectedly. The V-2 threat was eliminated only by bombing the launch sites and forcing the German army to retreat beyond missile range.

The V-2 obviously ushered in a new age of military technology. After the war there was intense competition between the United States and the Soviet Union to obtain these new missiles, as well as to obtain the German scientists who had developed them. The United States succeeded in capturing both Dornberger and von Braun as well as more than 60 V-2s; it was not revealed precisely what (or whom) the Soviets captured. However, given the relative immaturity of ballistic missile technology at that time, neither country achieved usable ballistic missiles for some time. During the late 1940s and early 1950s most of the nuclear competition between the two countries dealt with strategic bombers. Events in 1957 reshaped this contest.

The first ICBMs. In 1957 the Soviets launched a multistage ballistic missile (later given the NATO designation SS-6 Sapwood) as well as the first man-made satellite, Sputnik. This prompted the "missile gap" debate in the United States and resulted in higher priorities for the U.S. Thor and Jupiter IRBMs. Although originally scheduled for deployment in the early 1960s, these programs were accelerated, with Thor being deployed to England and Jupiter to Italy and Turkey in 1958. Thor and Jupiter were both single-stage, liquid-fueled missiles with inertial guidance systems and warheads of 1.5 megatons. Political difficulties in deploying these missiles on foreign soil prompted the United States to develop ICBMs, so that by late 1963 Thor and Jupiter had been terminated. (The missiles themselves were used extensively in the space program.)

Thor and Jupiter

The Soviet SS-6 system was an apparent failure. Given its limited range (less than 3,500 miles), it had to be launched from northern latitudes in order to reach the United States. The severe weather conditions at these launch facilities (Novaya Zemlya and the Arctic mainland bases of Norilsk and Vorkuta) seriously degraded operational effectiveness; pumps for liquid propellants froze, metal fatigue was extreme, and lubrication of moving parts was nearly impossible. In 1960 a missile engine exploded during a test, killing Mitrofan Ivanovich Nedelin, chief of the Strategic Rocket Forces, and several hundred observers.

Possibly as a result of these technical failures (and possibly in response to the deployment of Thor and Jupiter), the Soviets attempted to base the SS-4 Sandal, an IRBM with a one-megaton warhead and a range of 900–1,000 miles, closer to the United States and in a warmer climate. This precipitated the Cuban missile crisis of 1962, after which the SS-4 was withdrawn to Central Asia. (It was unclear whether the United States' deactivation of Thor and Jupiter was a condition of this withdrawal.)

In the meantime, the United States was developing operational ICBMs to be based on U.S. territory. The first versions were the Atlas and the Titan I. The Atlas-D (the first version deployed) had a liquid-fueled engine that generated 360,000 pounds of thrust. The missile was radio-inertial guided, launched above ground, and had a range of 7,500 miles. The follow-on Atlas-E/F increased thrust to 390,000 pounds, used all-inertial guidance, and moved from an aboveground to horizontal canister launch in the E and, finally, to silo-stored vertical launch in the F. The Atlas E carried a two-megaton, and the Atlas F a four-megaton, warhead. The Titan I was a two-stage, liquid-fueled, radio-inertial guided, silo-launched ICBM carrying a four-megaton warhead and capable of traveling 6,300 miles. Both systems became operational in 1959.

From liquid to solid fuel. This first generation of missiles was typified by its liquid fuel, which required both a propellant and an oxidizer for ignition as well as a complex (and heavy) system of pumps. The early liquid fuels were quite dangerous, difficult to store, and time-consuming to load. For example, Atlas and Titan used so-called cryogenic (hypercold) fuels that had to be stored and handled at very low temperatures (−422° F [−252° C] for liquid hydrogen). These propellants had to be stored outside the rocket and pumped aboard just before launch, consuming more than an hour.

As each superpower produced, or was thought to produce, more ICBMs, military commanders became concerned about the relatively slow reaction times of their own ICBMs. The first step toward "rapid reaction" was the rapid loading of liquid fuels. Using improved pumps, the reaction time of the Titan I was reduced from over one hour to less than 20 minutes. Then, with a second generation of storable liquids that could be kept loaded in the missile, reaction time was reduced to approximately one minute. Examples of second-generation storable-liquid missiles were the Soviet SS-7 Saddler and SS-8 Sasin (the latter deployed in 1963) and the U.S. Titan II. The Titan II was the largest ballistic missile ever developed by the United States. This two-stage ICBM was more than 100 feet long and 10 feet in diameter. Weighing more than 325,000 pounds at launch, it delivered its single warhead (with a throw weight of about 8,000 pounds) to a range of 9,000 miles and with a CEP of about one mile.

Storable liquids

In about 1964 China began developing a series of liquid-fueled IRBMs given the NATO designation CSS, for Chinese surface-to-surface missile. (The Chinese named the series Dong Feng, meaning "East Wind.") The CSS-1 carried a 20-kiloton warhead to a range of 600 miles. The CSS-2, entering service in 1970, was fueled by storable

liquids; it had a range of 1,500 miles and carried a one- to two-megaton warhead. With the two-stage CSS-3 (active from 1978) and the CSS-4 (active from 1980), the Chinese reached ICBM ranges of over 4,000 and 7,000 miles, respectively. The CSS-4 carried a warhead of four to five megatons.

Because storable liquids did not alleviate the dangers inherent in liquid fuels, and because the flight times of missiles flying between the United States and the Soviet Union shrank to less than 35 minutes from launch to impact, still faster reactions were sought with even safer fuels. This led to a third generation of missiles, powered by solid propellants. Solid propellants were, eventually, easier to make, safer to store, lighter in weight (because they did not require on-board pumps), and more reliable than their liquid predecessors. Here the oxidizer and propellant were mixed into a canister and kept loaded aboard the missile, so that reaction times were reduced to seconds. However, solid fuels were not without their complications. First, while it was possible with liquid fuels to adjust in flight the amount of thrust provided by the engine, rocket engines using solid fuel could not be throttled. Also, some early solid fuels had uneven ignition, producing surges or abrupt velocity changes that could disrupt or severely confound guidance systems.

The first solid-fueled U.S. system was the Minuteman I. This ICBM, conceived originally as a rail-mobile system, was deployed in silos in 1962, became operational the following year, and was phased out by 1973. The first Soviet solid-fueled ICBM was the SS-13 Savage, which became operational in 1969. This missile could carry a 750-kiloton warhead more than 5,000 miles. Because the Soviet Union deployed several other liquid-fueled ICBMs between 1962 and 1969, Western specialists speculated that the Soviets experienced engineering difficulties in producing solid propellants.

The French deployed the first of their solid-fueled S-2 missiles in 1971. These two-stage IRBMs carried a 150-kiloton warhead and had a range of 1,800 miles. The S-3, deployed in 1980, could carry a one-megaton warhead to a range of 2,100 miles.

The first SLBMs. Simultaneous with the early Soviet and U.S. efforts to produce land-based ICBMs, both countries were developing SLBMs. In 1955 the Soviets launched the first SLBM, the one- to two-megaton SS-N-4 Sark. This missile, deployed in 1958 aboard diesel-electric submarines and later aboard nuclear-powered vessels, had to be launched from the surface and had a range of only 350 miles. Partly in response to this deployment, the United States gave priority to its Polaris program, which became operational in 1960. Each Polaris A-1 carried a warhead of one megaton and had a range of 1,400 miles. The Polaris A-2, deployed in 1962, had a range of 1,700 miles and also carried a one-megaton warhead. The U.S. systems were solid-fueled, whereas the Soviets initially used storable liquids. The first Soviet solid-fueled SLBM was the SS-N-17 Snipe, deployed in 1978 with a range of 2,400 miles and a 500-kiloton warhead.

Polaris

Beginning in 1971, France deployed a series of solid-fueled SLBMs comprising the M-1, M-2 (1974), and M-20 (1977). The M-20, with a range of 1,800 miles, carried a one-megaton warhead. In the 1980s the Chinese fielded the two-stage, solid-fueled CSS-N-3 SLBM, which had a range of 1,700 miles and carried a two-megaton warhead.

Multiple warheads. By the early 1970s, several technologies were maturing that would produce a new wave of ICBMs. First, thermonuclear warheads, much lighter than the earlier atomic devices, had been incorporated into ICBMs by 1970. Second, the ability to launch larger throw weights, achieved especially by the Soviets, allowed designers to contemplate adding multiple warheads to each ballistic missile. Finally, improved and much lighter electronics translated into more accurate guidance.

The first steps toward incorporating these technologies came with multiple warheads, or multiple reentry vehicles (MRVs), and the Fractional Orbital Bombardment System (FOBS). The Soviets introduced both of these capabilities with the SS-9 Scarp, the first "heavy" missile, beginning in 1967. FOBS was based on a low-trajectory launch that would be fired in the opposite direction from the target and would achieve only partial earth orbit. With this method of delivery, it would be quite difficult to determine which target was being threatened. However, given the shallow reentry angles associated with a low trajectory and partial earth orbit, the accuracy of FOBS missiles was questionable. A missile carrying MRVs, on the other hand, would be launched toward the target in a high ballistic trajectory. Several warheads from the same missile would strike the same target, increasing the probability of killing that target, or individual warheads would strike separate targets within a very narrow ballistic "footprint." (The footprint of a missile is that area which is feasible for targeting, given the characteristics of the reentry vehicle.) The SS-9, model 4, and the SS-11 Sego, model 3, both had three MRVs and ballistic footprints equal to the dimensions of a U.S. Minuteman complex. The only instance in which the United States incorporated MRVs was with the Polaris A-3, which, after deployment in 1964, carried three 200-kiloton warheads a distance of 2,800 miles. In 1967 the British adapted their own warheads to the A-3, and beginning in 1982 they upgraded the system to the A3TK, which contained penetration aids (chaff, decoys, and jammers) designed to foil ballistic missile defenses around Moscow.

Soon after adopting MRVs the United States took the next technological step, introducing multiple independently targetable reentry vehicles (MIRVs). Unlike MRVs, independently targeted RVs could be released to strike widely separated targets, essentially expanding the footprint established by a missile's original ballistic trajectory. This demanded the capacity to maneuver before releasing the warheads, and maneuvering was provided by a structure in the front end of the missile called the "bus," which contained the RVs. The bus was essentially a final, guided stage of the missile (usually the fourth), that now had to be considered part of the missile's payload. Since any bus capable of maneuvering would take up weight, MIRVed systems would have to carry warheads of lower yield. This in turn meant that the RVs would have to be released on their ballistic paths with great accuracy. As stated above, solid-fueled motors could be neither throttled nor shut down and restarted; for this reason, liquid-fueled buses were developed for making the necessary course corrections. The typical flight profile for a MIRVed ICBM then became approximately 300 seconds of solid-rocket boost and 200 seconds of bus maneuvering to place the warheads on independent ballistic trajectories.

MIRVs

The first MIRVed system was the U.S. Minuteman III. Deployed in 1970, this three-stage, solid-fueled ICBM carried three MIRVs of an estimated 170 to 335 kilotons. The warheads had a range of 8,000 miles with CEPs of 725–925 feet. Beginning in 1970 the United States also MIRVed its SLBM force with the Poseidon C-3, which could deliver up to 14 50-kiloton RVs to a range of 2,800 miles and with a CEP of about 1,450 feet. After 1979 this force was upgraded with the Trident C-4, or Trident I, which could deliver eight 100-kiloton MIRVs with the same accuracy as the Poseidon, but to a distance of 4,600 miles. Much longer range was made possible in the Trident by adding a third stage, by replacing aluminum with lighter graphite epoxies, and by adding an "aerospike" to the nose cone that, extending after launch, produced the streamlining effect of a pointed design while allowing the larger volume of a blunt design. Accuracy was maintained by updating the missile's inertial guidance during bus maneuvering with stellar navigation.

By 1978 the Soviet Union had fielded its first MIRVed SLBM, the SS-N-18 Stingray. This liquid-fueled missile could deliver three or five 500-kiloton warheads to a distance of 4,000 miles, with a CEP of about 3,000 feet. On land in the mid-1970s, the Soviets deployed three MIRVed, liquid-fueled ICBM systems, all with ranges exceeding 6,000 miles and with CEPs of 1,000 to 1,500 feet: the SS-17 Spanker, with four 750-kiloton warheads; the SS-18 Satan, with up to 10 500-kiloton warheads; and the SS-19 Stiletto, with six 550-kiloton warheads. Each of these Soviet systems had several versions that traded multiple warheads for higher yield. For instance, the SS-18,

model 3, carried a single 20-megaton warhead. This giant missile, which replaced the SS-9 in the latter's silos, had about the same dimensions as the Titan II, but its throw weight of more than 16,000 pounds was twice that of the U.S. system.

Beginning in 1985, France upgraded its SLBM force with the M-4, a three-stage MIRVed missile capable of carrying six 150-kiloton warheads to ranges of 3,600 miles.

MX Peacekeeper

A second generation of MIRVed U.S. systems was represented by the Peacekeeper. Known as the MX during its 15-year development phase before entering service in 1986, this three-stage ICBM carried 10 300-kiloton warheads and had a range of 7,000 miles. Originally designed to be based on mobile railroad or wheeled launchers, the Peacekeeper was eventually housed in Minuteman silos. A second-generation MIRVed SLBM of the 1990s was the Trident D-5, or Trident II. Even though it was one-third again as long as its predecessor and had twice the throw weight, the D-5 could deliver 10 475-kiloton warheads to a range of 7,000 miles. Both the Trident D-5 and Peacekeeper represented a radical advance in accuracy, having CEPs of only 400 feet. The improved accuracy of the Peacekeeper was due to a refinement in the inertial guidance system, which housed the gyros and accelerometers in a floating-ball device, and to the use of an exterior celestial navigation system that updated the missile's position by reference to stars or satellites. The Trident D-5 also contained a star sensor and satellite navigator. This gave it several times the accuracy of the C-4 at more than twice the range.

Within the generally less-advanced guidance technology of the Soviet Union, an equally radical advance came with the solid-fueled SS-24 Scalpel and SS-25 Sickle ICBMs, deployed in 1987 and 1985, respectively. The SS-24 could carry eight or 10 MIRVed warheads of 100 kilotons, and the SS-25 was fitted with a single 550-kiloton RV. Both missiles had a CEP of 650 feet. In addition to their accuracy, these ICBMs represented a new generation in basing mode. The SS-24 was launched from railroad cars, while the SS-25 was carried on wheeled launchers that shuttled between concealed launch sites. As mobile-based systems, they were long-range descendants of the SS-20 Saber, an IRBM carried on mobile launchers that entered service in 1977, partly along the border with China and partly facing western Europe. That two-stage, solid-fueled missile could deliver three 150-kiloton warheads a distance of 3,000 miles with a CEP of 1,300 feet. It was phased out after the signing of the Intermediate-Range Nuclear Forces (INF) Treaty in 1987.

Ballistic missile defense. Although ballistic missiles followed a predictable flight path, defense against them was long thought to be technically impossible because their RVs were small and traveled at great speeds. Nevertheless, in the late 1960s the United States and Soviet Union pursued layered antiballistic missile (ABM) systems that combined a high-altitude interceptor missile (the U.S. Spartan and Soviet Galosh) with a terminal-phase interceptor (the U.S. Sprint and Soviet Gazelle). All systems were nuclear-armed. Such systems were subsequently limited by the Treaty on Anti-Ballistic Missile Systems of 1972, under a protocol in which each side was allowed one ABM location with 100 interceptor missiles each. The Soviet system, around Moscow, remained active and was upgraded in the 1980s, whereas the U.S. system was deactivated in 1976. Still, given the potential for renewed or surreptitious ballistic missile defenses, all countries incorporated penetration aids along with warheads in their missiles' payloads. MIRVs were also used to overcome missile defenses.

Maneuverable warheads. Even after updating a missile's guidance with stellar or satellite references, disturbances in final descent could throw a warhead off course. Also, given the advances in ballistic missile defenses that were achieved even after the ABM treaty was signed, RVs remained vulnerable. Two technologies offered possible means of overcoming these difficulties. Maneuvering warheads, or MaRVs, were first integrated into the U.S. Pershing II IRBMs deployed in Europe from 1984 until they were dismantled under the terms of the INF Treaty.

Pershing II and Radag guidance

The warhead of the Pershing II contained a radar area guidance (Radag) system that compared the terrain toward which it descended with information stored in a self-contained computer. The Radag system then issued commands to control fins that adjusted the glide of the warhead. Such terminal-phase corrections gave the Pershing II, with a range of 1,100 miles, a CEP of 150 feet. The improved accuracy allowed the missile to carry a low-yield 15-kiloton warhead.

MaRVs would present ABM systems with a shifting, rather than ballistic, path, making interception quite difficult. Another technology, precision-guided warheads, or PGRVs, would actively seek a target, then, using flight controls, actually "fly out" reentry errors. This could yield such accuracy that nuclear warheads could be replaced by conventional explosives.

Cruise missiles. The single most important difference between ballistic missiles and cruise missiles is that the latter operate within the atmosphere. This presents both advantages and disadvantages. One advantage of atmospheric flight is that traditional methods of flight control (*e.g.,* airfoil wings for aerodynamic lift, rudder and elevator flaps for directional and vertical control) are readily available from the technologies of manned aircraft. Also, while strategic early-warning systems can immediately detect the launch of ballistic missiles, low-flying cruise missiles presenting small radar and infrared cross sections offer a means of slipping past these air-defense screens.

The principal disadvantage of atmospheric flight centres around the fuel requirements of a missile that must be powered continuously for strategic distances. Some tactical-range antiship cruise missiles such as the U.S. Harpoon and the Soviet SS-N-19 Shipwreck have been powered by turbojet engines, and even some non-cruise missiles such as the Soviet SA-6 Gainful surface-to-air missile have used ramjets to reach supersonic speed, but at ranges of 1,000 miles or more these engines would require huge amounts of fuel. This in turn would require a larger missile, which would approach a manned jet aircraft in size and would thereby lose the unique ability to evade enemy defenses. This problem of maintaining balance between range, size, and fuel consumption was not solved until reliable, fuel-efficient turbofan engines were made small enough to propel a missile of radar-evading size.

The problem of guidance

As with ballistic missiles, guidance has been a long-standing problem in cruise missile development. Tactical cruise missiles generally use radio or inertial guidance to reach the general vicinity of their targets and then home onto the targets with various radar or infrared mechanisms. Radio guidance, however, is subject to line-of-sight range limitations, and inaccuracies tend to arise in inertial systems over the long flight times required of strategic cruise missiles. Radar and infrared homing devices, moreover, can be jammed or spoofed. Adequate long-range guidance for cruise missiles was not available until inertial systems were designed that could be updated periodically by self-contained electronic map-matching devices.

Beginning in the 1950s, the Soviet Union pioneered the development of tactical air- and sea-launched cruise missiles, and in 1984 a strategic cruise missile given the NATO designation AS-15 Kent became operational aboard Tu-95 bombers. But Soviet programs have been so cloaked in secrecy that the following account of the development of cruise missiles focuses by necessity on U.S. programs.

The V-1. The first practical cruise missile was the German V-1 of World War II, which was powered by a pulse jet that used a cycling flutter valve to regulate the air and fuel mixture. Since the pulse jet required airflow for ignition, it could not operate below 150 miles per hour. Therefore, a ground catapult boosted the V-1 to 200 miles per hour, at which time the pulse-jet engine was ignited. Once ignited, it could attain speeds of 400 miles per hour and ranges exceeding 150 miles. Course control was accomplished by a combined air-driven gyroscope and magnetic compass, and altitude was controlled by a simple barometric altimeter; as a consequence, the V-1 was subject to heading, or azimuth, errors resulting from gyro drift, and it had to be operated at fairly high altitudes (usually above 2,000 feet) to compensate for altitude er-

rors caused by differences in atmospheric pressure along the route of flight.

The missile was armed in flight by a small propeller that, after a specified number of turns, activated the warhead at a safe distance from the launch. As the V-1 approached its target, the control vanes were inactivated and a rear-mounted spoiler, or drag device, deployed, pitching the missile nose-down toward the target. This usually interrupted the fuel supply, causing the engine to quit, and the weapon detonated upon impact.

Because of the rather crude method of calculating the impact point by the number of revolutions of a small propeller, the Germans could not use the V-1 as a precision weapon, nor could they determine the actual impact point in order to make course corrections for subsequent flights. In fact, the British publicized inaccurate information on impact points, causing the Germans to adjust their preflight calculations erroneously. As a result, V-1s often fell well short of their intended targets.

Following the war there was considerable interest in cruise missiles. Between 1945 and 1948, the United States began approximately 50 independent cruise missile projects, but lack of funding gradually reduced that number to three by 1948. These three—Snark, Navaho, and Matador—provided the necessary technical groundwork for the first truly successful strategic cruise missiles, which entered service in the 1980s.

Snark. The Snark was an air force program begun in 1945 to produce a subsonic (600-mile-per-hour) cruise missile capable of delivering a 2,000-pound atomic or conventional warhead to a range of 5,000 miles, with a CEP of less than 1.75 miles. Initially, the Snark used a turbojet engine and an inertial navigation system, with a complementary stellar navigation monitor to provide intercontinental range. By 1950, due to the yield requirements of atomic warheads, the design payload had changed to 5,000 pounds, accuracy requirements shrank the CEP to 1,500 feet, and range increased to more than 6,200 miles. These design changes forced the military to cancel the first Snark program in favour of a "Super Snark," or Snark II.

Control and propulsion problems with the Snark

The Snark II incorporated a new jet engine that was later used in the B-52 bomber and KC-135A aerial tanker operated by the Strategic Air Command. Although this engine design was to prove quite reliable in manned aircraft, other problems—in particular, those associated with flight dynamics—continued to plague the missile. The Snark lacked a horizontal tail surface, it used elevons instead of ailerons and elevators for attitude and directional control, and it had an extremely small vertical tail surface. These inadequate control surfaces, and the relatively slow (or sometimes nonexistent) ignition of the jet engine, contributed significantly to the missile's difficulties in flight tests—to a point where the coastal waters off the test site at Cape Canaveral, Fla., were often referred to as "Snark-infested waters." Flight control was not the least of the Snark's problems: unpredictable fuel consumption also resulted in embarrassing moments. One 1956 flight test appeared amazingly successful at the outset, but the engine failed to shut off and the missile was last seen "heading toward the Amazon." (The vehicle was found in 1982 by a Brazilian farmer.)

Considering the less than dramatic successes in the test program, the Snark, as well as other cruise missile programs, probably would have been destined for cancellation had it not been for two developments. First, antiaircraft defenses had improved to a point where bombers could no longer reach their targets with the usual high-altitude flight paths. Second, thermonuclear weapons were beginning to arrive in military inventories, and these lighter, higher-yield devices allowed designers to relax CEP constraints. As a result, an improved Snark was deployed in the late 1950s at two bases in Maine and Florida.

The new missile, however, continued to exhibit the unreliabilities and inaccuracies typical of earlier models. On a series of flight tests, the Snark's CEP was estimated to average 20 miles, with the most accurate flight striking 4.2 miles left and 1,600 feet short. This "successful" flight was the only one to reach the target area at all and was one of only two to go beyond 4,400 miles. Accumulated test data showed that the Snark had a 33-percent chance of successful launch and a 10-percent chance of achieving the required distance. As a consequence, the two Snark units were deactivated in 1961.

Navaho. The second postwar U.S. cruise missile effort was the Navaho, an intercontinental supersonic design. Unlike earlier efforts, which were extrapolated from V-1 engineering, the Navaho was based on the V-2; the basic V-2 structure was fitted with new control surfaces, and the rocket engine was replaced by a turbojet/ramjet combination. Known by a variety of names, the Navaho emerged into a missile more than 70 feet long, with canard fins (*i.e.,* control surfaces set forward of the wing), a V tail, and a large delta wing. (These flight control designs would eventually make their way onto other supersonic aircraft, such as the experimental XB-70 Valkyrie bomber, several fighter planes, and the supersonic transport.)

With the exception of technologies associated with supersonic lift and control, few other aspects of the Navaho met designers' expectations. Most frustrating were difficulties with the ramjet engine, which was necessary for sustained supersonic flight. For a variety of reasons, including interrupted fuel flow, turbulence in the ramjet cavity, and clogging of the ramjet fire-ring, few of the engines ignited. This led engineers to label the project "Never Go, Navaho"—a name that stuck until the program was cancelled in 1958 after achieving only 1½ hours airborne. No missile was ever deployed.

Technologies explored in the Navaho program, besides those of flight dynamics, were used in other areas. Derivatives of the missile's titanium alloys, which were developed to accommodate surface temperatures at supersonic speed, came to be used on most high-performance aircraft. The rocket booster (which launched the missile until the ramjet ignited) eventually became the Redstone engine, which powered the Mercury manned spacecraft series, and the same basic design was used in the Thor and Atlas ballistic missiles. The guidance system, an inertial autonavigation design, was incorporated into a later cruise missile (Hound Dog) and was used by the nuclear submarine USS *Nautilus* for its under-the-ice passage of the North Pole in 1958.

Matador and other programs. The third postwar U.S. cruise missile effort was the Matador, a ground-launched, subsonic missile designed to carry a 3,000-pound warhead to a range of more than 600 miles. In its early development, Matador's radio-controlled guidance, which was limited essentially to the line of sight between the ground controller and the missile, covered less than the missile's potential range. However, in 1954 an automatic terrain recognition and guidance (Atran) system was added (and the missile system was subsequently designated Mace). Atran, which used radar map-matching for both en-route and terminal guidance, represented a major breakthrough in accuracy, a problem long associated with cruise missiles. The low availability of radar maps, especially of areas in the Soviet Union (the logical target area), limited operational use, however. Nonetheless, operational deployments began in 1954 to Europe and in 1959 to Korea. The missile was phased out in 1962, its most serious problems being associated with guidance.

Atran guidance: a major breakthrough

While the U.S. Air Force was exploring the Snark, Navaho, and Matador programs, the navy was pursuing related technologies. The Regulus, which was closely akin to the Matador (having the same engine and roughly the same configuration), became operational in 1955 as a subsonic missile launched from both submarines and surface vessels, carrying a 3.8-megaton warhead. Decommissioned in 1959, the Regulus did not represent much of an improvement over the V-1.

A follow-on design, Regulus II, was pursued briefly, striving for supersonic speed. However, the navy's preference for the new large, angle-deck nuclear aircraft carriers and for ballistic missile submarines relegated sea-launched cruise missiles to relative obscurity. Another project, the Triton, was similarly bypassed due to design difficulties and lack of funding. The Triton was to have had a range of 12,000 miles and a payload of 1,500 pounds. Radar map-matching guidance was to have given it a CEP of 1,800 feet.

In the early 1960s the Air Force produced and deployed the Hound Dog cruise missile on B-52 bombers. This supersonic missile was powered by a turbojet engine to a range of 400–450 miles. It used the guidance system of the earlier Navaho. The missile was so large, however, that only two could be carried on the outside of the aircraft. This external carriage allowed B-52 crew members to use the Hound Dog engines for extra thrust on takeoff, but the extra drag associated with the carriage, as well as the additional weight (20,000 pounds), meant a net loss of range for the aircraft. By 1976 the Hound Dog had given way to the short-range attack missile, or SRAM, essentially an internally carried, air-launched ballistic missile.

ALCM, SLCM, and GLCM. By 1972, constraints placed on ballistic missiles by the SALT I treaty prompted U.S. nuclear strategists to think again about using cruise missiles. There was also concern over Soviet advances in antiship cruise missile technology, and in Vietnam remotely piloted vehicles had demonstrated considerable reliability in gathering intelligence information over previously inaccessible, highly defended areas. Improvements in electronics—in particular, microcircuits, solid-state memory, and computer processing—presented inexpensive, lightweight, and highly reliable methods of solving the persistent problems of guidance and control. Perhaps most important, terrain contour mapping, or Tercom, techniques, derived from the earlier Atran, offered excellent en route and terminal-area accuracy.

Tercom guidance: the problem solved

Tercom used a radar or photographic image from which a digitalized contour map was produced. At selected points in the flight known as Tercom checkpoints, the guidance system would match a radar image of the missile's current position with the programmed digital image, making corrections to the missile's flight path in order to place it on the correct course. Between Tercom checkpoints, the missile would be guided by an advanced inertial system; this would eliminate the need for constant radar emissions, which would make electronic detection extremely difficult. As the flight progressed, the size of the radar map would be reduced, improving accuracy. In practice, Tercom brought the CEP of modern cruise missiles down to less than 150 feet (see Figure 21).

Improvements in engine design also made cruise missiles more practical. In 1967 the Williams International Corporation produced a small turbofan engine (12 inches in diameter, 24 inches long) that weighed less than 70 pounds and produced more than 400 pounds of thrust. New fuel mixtures offered more than 30-percent increases in fuel energy, which translated directly into extended range.

By the end of the Vietnam War, both the U.S. Navy and Air Force had cruise missile projects under way. At 19 feet three inches, the navy's sea-launched cruise missile (SLCM; eventually designated the Tomahawk) was 30 inches shorter than the air force's air-launched cruise missile (ALCM), but system components were quite similar and often from the same manufacturer (both missiles used the Williams engine and the McDonnell Douglas Corporation's Tercom). The Boeing Company produced the ALCM, while the General Dynamics Corporation produced the SLCM as well as the ground-launched cruise missile, or GLCM. The SLCM and GLCM were essentially the same configuration, differing only in their basing mode. The GLCM was designed to be launched from wheeled transporter-erector-launchers, while the SLCM was expelled from submarine tubes to the ocean surface in steel canisters or launched directly from armoured box launchers aboard surface ships. Both the SLCM and GLCM were propelled from their launchers or canisters by a solid-rocket booster, which dropped off after the wings and tail fins flipped out and the jet engine ignited. The ALCM, being dropped from a bomb-bay dispenser or wing pylon of a flying B-52 or B-1 bomber, did not require rocket boosting.

As finally deployed, the U.S. cruise missiles were intermediate-range weapons that flew at an altitude of 100 feet to a range of 1,500 miles. The SLCM was produced in three versions: a tactical-range (275-mile) antiship missile, with a combination of inertial guidance and active radar homing and with a high-explosive warhead; and two intermediate-range land-attack versions, with combined inertial and Tercom guidance and with either a high-explosive or a 200-kiloton nuclear warhead. The ALCM carried the same nuclear warhead as the SLCM, while the GLCM carried a low-yield warhead of 10 to 50 kilotons.

The ALCM entered service in 1982 and the SLCM in 1984. The GLCM was first deployed to Europe in 1983, but all GLCMs were dismantled after the signing of the INF Treaty.

Although their small size and low flight paths made the ALCM and SLCM difficult to detect by radar (the ALCM presented a radar cross section only one one-thousandth that of the B-52 bomber), their subsonic speed of about 500 miles per hour made them vulnerable to air defenses once they were detected. For this reason, the U.S.

From L. Freedman, *Atlas of Global Strategy*, copyright © Equinox (Oxford) Ltd 1985

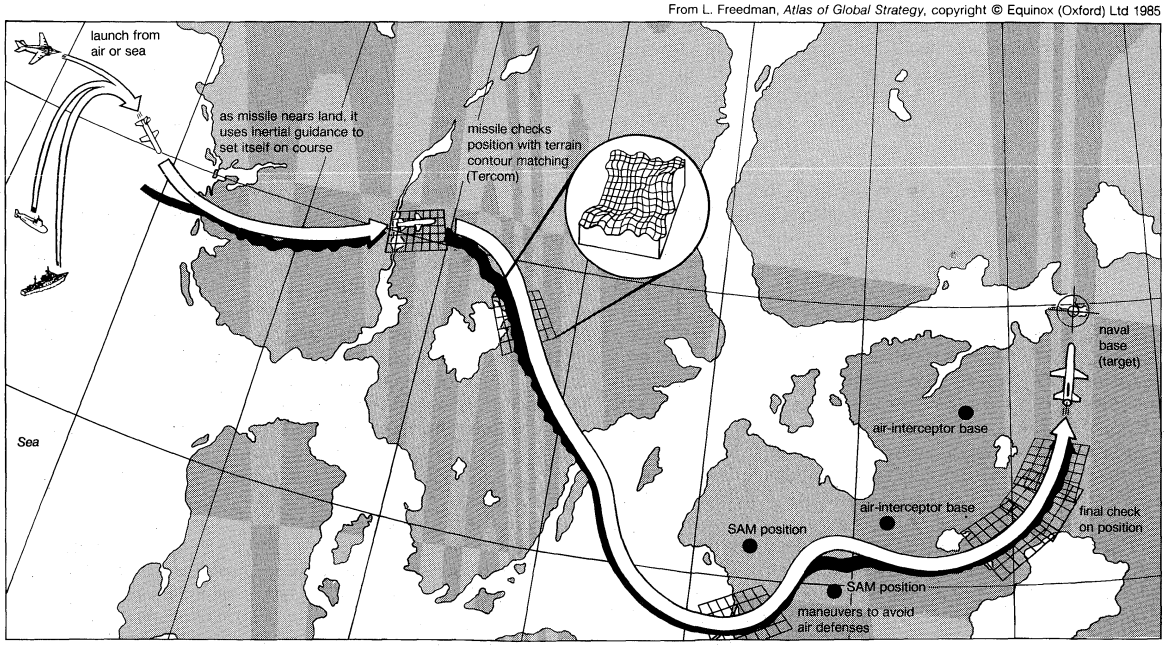


Figure 21: *Flight path of a cruise missile.*
Limited to subsonic speeds, the missile would rely on its small size and ground-hugging flight profile to evade detection. U.S. cruise missiles, guided by a digitalized map of the terrain along their routes, could strike their targets with a nuclear warhead after being launched from 1,500 miles away.

Air Force began production of an advanced cruise missile, which would incorporate stealth technologies such as radar-absorbent materials and smooth, nonreflective surface shapes. The advanced cruise missile would have a range of over 1,800 miles. (S.O.F.)

Nuclear weapons

Nuclear weapons derive their enormous explosive force from either the fission or fusion of atomic nuclei. Their significance may best be appreciated by the coining of the words kiloton (1,000 tons) and megaton (one million tons) to describe their blast effect in equivalent weights of TNT. For example, the first nuclear fission bomb, the one dropped on Hiroshima, Japan, in 1945, released energy equaling 15,000 tons (15 kilotons) of chemical explosive from less than 130 pounds (60 kilograms) of uranium. Fusion bombs, on the other hand, have given yields up to almost 60 megatons. The first nuclear weapons were bombs delivered by aircraft; warheads for strategic ballistic missiles, however, have become by far the most important nuclear weapons (see above *Strategic missiles*). There are also smaller tactical nuclear weapons that include artillery projectiles, demolition munitions (land mines), antisubmarine depth bombs, torpedoes, and short-range ballistic and cruise missiles. The U.S. stockpile of nuclear weapons reached its peak in 1967 with more than 32,000 warheads of 30 different types; the Soviet stockpile reached its peak of about 33,000 warheads in 1988.

Principles of fission and fusion

The basic principle of nuclear fission weapons (also called atomic bombs) involves the assembly of a sufficient amount of fissile material (*e.g.*, the uranium isotope uranium-235 or the plutonium isotope plutonium-239) to "go supercritical"—that is, for neutrons (which cause fission and are in turn released during fission) to be produced at a much faster rate than they can escape from the assembly. There are two ways in which a subcritical assembly of fissionable material can be rendered supercritical and made to explode. The subcritical assembly may consist of two parts, each of which is too small to have a positive multiplication rate; the two parts can be shot together by a gun-type device. Alternatively, a subcritical assembly surrounded by a chemical high explosive may be compressed into a supercritical one by detonating the explosive.

The basic principle of the fusion weapon (also called the thermonuclear or hydrogen bomb) is to produce ignition conditions in a thermonuclear fuel such as deuterium, an isotope of hydrogen with double the weight of normal hydrogen, or lithium deuteride. The Sun may be considered a thermonuclear device; its main fuel is deuterium, which it consumes in its core at temperatures of 18,000,000° to 36,000,000° F (10,000,000° to 20,000,000° C). To achieve comparable temperatures in a weapon, a fission triggering device is used.

THE DEVELOPMENT OF FISSION WEAPONS

Following the discovery of artificial radioactivity in the 1930s, the Italian physicist Enrico Fermi performed a series of experiments in which he exposed many elements to low-velocity neutrons. When he exposed thorium and uranium, chemically different radioactive products resulted, indicating that new elements had been formed, rather than merely isotopes of the original elements. Fermi concluded that he had produced elements beyond uranium (element 92), then the last element in the periodic table; he called them transuranic elements and named two of them ausenium (element 93) and hesperium (element 94). During the autumn of 1938, however, when Fermi was receiving the Nobel Prize for his work, Otto Hahn and Fritz Strassmann of Germany discovered that one of the "new" elements was actually barium (element 56).

Discovery of uranium fission

The Danish scientist Niels Bohr visited the United States in January 1939, carrying with him an explanation, devised by the Austrian refugee scientist Lise Meitner and her nephew Otto Frisch, of the process behind Hahn's surprising data. Low-velocity neutrons caused the uranium nucleus to fission, or break apart, into two smaller pieces; the combined atomic numbers of the two pieces—for example, barium and krypton—equalled that of the uranium nucleus. Much energy was released in the process. This news set off experiments at many laboratories. Bohr worked with John Wheeler at Princeton; they postulated that the uranium isotope uranium-235 was the one undergoing fission; the other isotope, uranium-238, merely absorbed the neutrons. It was discovered that neutrons were produced during the fission process; on the average, each fissioning atom produced more than two neutrons. If the proper amount of material were assembled, these free neutrons might create a chain reaction. Under special conditions, a very fast chain reaction might produce a very large release of energy; in short, a weapon of fantastic power might be feasible.

The first atomic bomb. The possibility that such a weapon might first be developed by Nazi Germany alarmed many scientists and was drawn to the attention of President Franklin D. Roosevelt by Albert Einstein, then living in the United States. The president appointed an Advisory Committee on Uranium; it reported that a chain reaction in uranium was possible, though unproved. Chain-reaction experiments with carbon and uranium were started in New York City at Columbia University, and in March 1940 it was confirmed that the isotope uranium-235 was responsible for low-velocity neutron fission in uranium. The Advisory Committee on Uranium increased its support of the Columbia experiments and arranged for a study of possible methods for separating the uranium-235 isotope from the much more abundant uranium-238. (Normal uranium contains approximately 0.7 percent uranium-235, most of the remainder being uranium-238.) The centrifuge process, in which the heavier isotope is spun to the outside, as in a cream separator, at first seemed the most useful, but at Columbia a rival process was proposed. In that process, gaseous uranium hexafluoride is diffused through barriers, or filters; more molecules containing the lighter isotope, uranium-235, would pass through the filter than those containing the heavier isotope, slightly enriching the mixture on the far side. A sequence of several thousand stages would be needed to enrich the mixture to 90 percent uranium-235; the total barrier area would be many acres.

During the summer of 1940, Edwin McMillan and Philip Abelson of the University of California at Berkeley discovered element 93, named neptunium; they inferred that this element would decay into element 94. The Bohr and Wheeler fission theory suggested that one of the isotopes, mass number 239, of this new element might also fission under low-velocity neutron bombardment. The cyclotron at the University of California at Berkeley was put to work to make enough element 94 for experiments; by mid-1941, element 94 had been firmly identified and named plutonium, and its fission characteristics had been established. Low-velocity neutrons did indeed cause it to undergo fission, and at a rate much higher than that of uranium-235. The Berkeley group, under Ernest Lawrence, was also considering producing large quantities of uranium-235 by turning one of their cyclotrons into a super mass spectrograph. A mass spectrograph employs a magnetic field to bend a current of uranium ions; the heavier ions (uranium-238) bend at a larger radius than the lighter ions (uranium-235), allowing the two separated currents to be collected in separate receivers.

Discovery of plutonium fission

In May 1941 a review committee reported that a nuclear explosive probably could not be available before 1945, that a chain reaction in natural uranium was probably 18 months off, and that it would take at least an additional year to produce enough plutonium for a bomb and three to five years to separate enough uranium-235. Further, it was held that all of these estimates were optimistic. In late June 1941 President Roosevelt established the Office of Scientific Research and Development under the direction of the scientist Vannevar Bush.

In the fall of 1941 the Columbia chain-reaction experiment with natural uranium and carbon yielded negative results. A review committee concluded that boron impurities might be poisoning it by absorbing neutrons. It was decided to transfer all such work to the University of Chicago and repeat the experiment there with high-purity

carbon. At Berkeley, the cyclotron, converted into a mass spectrograph (later called a calutron), was exceeding expectations in separating uranium-235, and it was enlarged to a 10-calutron system capable of producing one-tenth of an ounce (about three grams) of uranium-235 per day.

The Manhattan Project

The U.S. entry into World War II in December 1941 was decisive in providing funds for a massive research and production effort for obtaining fissionable materials, and in May 1942 the momentous decision was made to proceed simultaneously on all promising production methods. Bush decided that the army should be brought into the production plant construction activities. The Corps of Engineers opened an office in New York City and named it the Manhattan Engineer District Office. After considerable argument over priorities, a workable arrangement was achieved with the formation of a three-man policy board chaired by Bush and the appointment on September 17 of Colonel Leslie Groves as head of the Manhattan Engineer District. Groves arranged contracts for a gaseous diffusion separation plant, a plutonium production facility, and a calutron pilot plant, which might be expanded later. The day before the success of Fermi's chain-reaction experiment at the University of Chicago on Dec. 2, 1942, Groves (now a brigadier general) signed the construction contract for the plutonium production reactors. Many problems were still unsolved, however. First, the gaseous diffusion barrier had not yet been demonstrated as practical. Second, Berkeley had been successful with its empirically designed calutron, but the Oak Ridge pilot plant contractors were understandably uneasy about the rough specifications available for the massive separation of uranium-235, which was designated the Y-12 effort. Third, plutonium chemistry was almost unknown; in fact, it was not known whether or not plutonium gave off neutrons during fission, or, if so, how many.

Meantime, as part of the June 1942 reorganization, J. Robert Oppenheimer became, in October, the director of Project Y, the group that was to design the actual weapon. This effort was spread over several locations. On November 16 Groves and Oppenheimer visited the former Los Alamos Ranch School, some 60 miles (100 kilometres) north of Albuquerque, N.M., and on November 25 Groves approved it as the site for the Los Alamos Scientific Laboratory. By July two essential and encouraging pieces of experimental data had been obtained—plutonium did give off neutrons in fission, more than uranium-235; and the neutrons were emitted in a short time compared to that needed to bring the weapon materials into a supercritical assembly. The theorists contributed one discouraging note: their estimate of the critical mass for uranium-235 had risen over threefold, to something between 50 and 100 pounds.

The triggering methods

The emphasis during the summer and fall of 1943 was on the gun method of assembly, in which the projectile, a subcritical piece of uranium-235 (or plutonium-239), would be placed in a gun barrel and fired into the target, another subcritical piece of uranium-235. After the mass was joined (and now supercritical), a neutron source would be used to start the chain reaction. A problem developed with applying the gun method to plutonium, however. In manufacturing plutonium-239 from uranium-238 in a reactor, some of the plutonium-239 absorbs a neutron and becomes plutonium-240. This material undergoes spontaneous fission, producing neutrons. Some neutrons will always be present in a plutonium assembly and cause it to begin multiplying as soon as it goes critical, before it reaches supercriticality; it will then explode prematurely and produce comparatively little energy. The gun designers tried to beat this problem by achieving higher projectile speeds, but they lost out in the end to a better idea—the implosion method.

In April 1943 a Project Y physicist, Seth Neddermeyer, proposed to assemble a supercritical mass from many directions, instead of just two as in the gun. In particular, a number of shaped charges placed on the surface of a sphere would fire many subcritical pieces into one common ball at the centre of the sphere. John von Neumann, a mathematician who had had experience in shaped-charge, armour-piercing work, supported the implosion method enthusiastically and pointed out that the greater speed of assembly might solve the plutonium-240 problem. The physicist Edward Teller suggested that the converging material might also become compressed, offering the possibility that less material would be needed. By late 1943 the implosion method was being given an increasingly higher priority; by July 1944 it had become clear that the plutonium gun could not be built. The only way to use plutonium in a weapon was by the implosion method.

By 1944 the Manhattan Project was spending money at a rate of more than $1 billion per year. The situation was likened to a nightmarish horse race; no one could say which of the horses (the calutron plant, the diffusion plant, or the plutonium reactors) was likely to win or whether any of them would even finish the race. In July 1944 the first Y-12 calutrons had been running for three months but were operating at less than 50 percent efficiency; the main problem was in recovering the large amounts of material that reached neither the uranium-235 nor uranium-238 boxes and, thus, had to be rerun through the system. The gaseous diffusion plant was far from completion, the production of satisfactory barriers remaining the major problem. And the first plutonium reactor at Hanford, Wash., had been turned on in September, but it had promptly turned itself off. Solving this problem, which proved to be caused by absorption of neutrons by one of the fission products, took several months. These delays meant almost certainly that the war in Europe would be over before the weapon could be ready. The ultimate target was slowly changing from Germany to Japan.

Within 24 hours of Roosevelt's death on April 12, 1945, President Harry S. Truman was told briefly about the atomic bomb by Secretary of War Henry Stimson. On April 25 Stimson, with Groves's assistance, gave Truman a more extensive briefing on the status of the project: the uranium-235 gun design had been frozen, but sufficient uranium-235 would not be accumulated until around August 1. Enough plutonium-239 would be available for an implosion assembly to be tested in early July; a second would be ready in August. Several B-29s had been modified to carry the weapons, and support construction was under way at Tinian, in the Mariana Islands, 1,500 miles south of Japan.

The test of the plutonium weapon was named Trinity; it was fired at 5:29:45 AM (local time) on July 16, 1945, at the Alamogordo Bombing Range in south central New Mexico. The theorists' predictions of the energy release ranged from the equivalent of less than 1,000 tons of TNT to 45,000 tons. The test produced an energy, or yield, equivalent to 21,000 tons of TNT.

The dropping of the bomb

A single B-29 bomber, named the *Enola Gay,* flew over Hiroshima, Japan, on Monday, Aug. 6, 1945, at 8:15 in the morning, local time. The untested uranium-235 gun-assembly bomb, nicknamed Little Boy, was air-burst 1,900 feet (680 metres) above the city to maximize destruction. Two-thirds of the city area was destroyed. The population actually present at the time was estimated at 350,000; of these, 140,000 died by the end of the year. The second weapon, a duplicate of the plutonium-239 implosion assembly tested in Trinity and nicknamed Fat Man, was to be dropped on Kokura on August 11; a third was being prepared in the United States for possible use in late August or early September. To avoid bad weather, the schedule was moved up two days to August 9. The B-29, named *Bock's Car,* spent 10 minutes over Kokura without sighting its aim point; it then proceeded to the secondary target of Nagasaki, where, at 11:02 AM local time, the weapon was air-burst at 1,650 feet with a force later estimated at 21 kilotons. About half the city was destroyed, and, of the estimated 270,000 people present at the time, about 70,000 died by the end of the year.

The spread of atomic weapons. Scientists in several countries performed experiments in connection with nuclear reactors and fission weapons during World War II, but no country other than the United States carried its projects as far as separating uranium-235 or manufacturing plutonium-239.

The Axis powers. By the time the war began on Sept. 1, 1939, Germany had a special office for the military

application of nuclear fission; chain-reaction experiments with uranium and carbon were being planned, and ways of separating the uranium isotopes were under study. Some measurements on carbon, later shown to be in error, led the physicist Werner Heisenberg to recommend that heavy water be used, instead, for the moderator. This dependence on scarce heavy water was a major reason the German experiments never reached a successful conclusion. The isotope separation studies were oriented toward low enrichments (about 1 percent uranium-235) for the chain reaction experiments; they never got past the laboratory apparatus stage, and several times these prototypes were destroyed in bombing attacks. As for the fission weapon itself, it was a rather distant goal, and practically nothing but "back-of-the-envelope" studies were done on it.

German difficulties

Like their counterparts elsewhere, Japanese scientists initiated research on an atomic bomb. In December 1940, Japan's leading scientist, Nishina Yoshio, undertook a small-scale research effort supported by the armed forces. It did not progress beyond the laboratory owing to lack of government support, resources, and uranium.

Great Britain. The British weapon project started informally, as in the United States, among university physicists. In April 1940 a short paper by Otto Frisch and Rudolf Peierls, expanding on the idea of critical mass, estimated that a superweapon could be built using several pounds of pure uranium-235 and that this amount of material might be obtainable from a chain of diffusion tubes. This three-page memorandum was the first report to foretell with scientific conviction the practical possibility of making a bomb and the horrors it would bring. A group of scientists known as the MAUD committee was set up in the Ministry of Aircraft Production in April 1940 to decide if a uranium bomb could be made. The committee approved a report on July 15, 1941, concluding that the scheme for a uranium bomb was practicable, that work should continue on the highest priority, and that collaboration with the Americans should be continued and expanded. As the war took its toll on the economy, the British position evolved through 1942 and 1943 to one of full support for the American project with the realization that Britain's major effort would come after the war. While the British program was sharply reduced at home, approximately 50 scientists and engineers went to the United States at the end of 1943 and during 1944 to work on various aspects of the Manhattan Project. The valuable knowledge and experience they acquired sped the development of the British bomb after 1945.

The formal postwar decision to manufacture a British atomic bomb was made by Prime Minister Clement Attlee's government during a meeting of the Defence Subcommittee of the Cabinet in early January 1947. The construction of a first reactor to produce fissile material and associated facilities had gotten under way the year before. William Penney, a member of the British team at Los Alamos during the war, was placed in charge of fabricating and testing the bomb, which was to be of a plutonium type similar to the one dropped on Nagasaki. That Britain was developing nuclear weapons was not made public until Prime Minister Winston Churchill announced on Feb. 17, 1952, plans to test the first British-made atomic bomb at the Monte Bello Islands off the northwest coast of Australia. There, on Oct. 3, 1952, the first British atomic weapons test, called Hurricane, was successfully conducted aboard the frigate HMS *Plym*. By early 1954 Royal Air Force Canberra bombers were armed with atomic bombs.

The British test

The Soviet Union. In the decade before the war, Soviet physicists were actively engaged in nuclear and atomic research. By 1939 they had established that, once uranium has been fissioned, each nucleus emits neutrons and can therefore, at least in theory, begin a chain reaction. The following year, physicists concluded that such a chain reaction could be ignited in either natural uranium or its isotope, uranium-235, and that this reaction could be sustained and controlled with a moderator such as heavy water. In June 1940 the Soviet Academy of Sciences established the Uranium Commission to study the "uranium problem."

In February 1939, news had reached Soviet physicists of the discovery of nuclear fission in the West. The military implications of such a discovery were immediately apparent, but Soviet research was brought to a halt by the German invasion in June 1941. In early 1942 the physicist Georgy N. Flerov noticed that articles on nuclear fission were no longer appearing in western journals; this indicated that research on the subject had become secret. In response, Flerov wrote to, among others, Premier Joseph Stalin, insisting that "we must build the uranium bomb without delay." In 1943 Stalin ordered the commencement of a research project under the supervision of Igor V. Kurchatov, who had been director of the nuclear physics laboratory at the Physico-Technical Institute in Leningrad. Kurchatov initiated work on three fronts: achieving a chain reaction in a uranium pile, designing both uranium-235 and plutonium bombs, and separating isotopes from these materials.

By the end of 1944, 100 scientists were working under Kurchatov, and by the time of the Potsdam Conference, which brought the Allied leaders together the day after the Trinity test, the project on the atomic bomb was seriously under way. During one session at the conference, Truman remarked to Stalin that the United States had built a "new weapon of unusual destructive force." Stalin replied that he would like to see the United States make "good use of it against the Japanese."

Upon his return from Potsdam, Stalin ordered that work on the fission bomb proceed at a faster pace. On Aug. 7, 1945, the day after the bombing of Hiroshima, he placed Lavrenty P. Beria, the chief of secret police, in charge of the Soviet version of the Manhattan Project. The first Soviet chain reaction took place in Moscow on Dec. 25, 1946, using an experimental graphite-moderated natural uranium pile, and the first plutonium production reactor became operational at Kyshtym, in the Ural Mountains, on June 19, 1948. The first Soviet weapon test occurred on Aug. 29, 1949, using plutonium; it had a yield of 10 to 20 kilotons.

The first Soviet device

France. French scientists, such as Henri Becquerel, Marie and Pierre Curie, and Frédéric and Irène Joliot-Curie, made important contributions to 20th-century atomic physics. During World War II several French scientists participated in an Anglo-Canadian project in Canada, where eventually a heavy water reactor was built at Chalk River, Ont., in 1945.

On Oct. 18, 1945, the Atomic Energy Commission (Commissariat à l'Énergie Atomique; CEA) was established by General Charles de Gaulle with the objective of exploiting the scientific, industrial, and military potential of atomic energy. The military application of atomic energy did not begin until 1951. In July 1952 the National Assembly adopted a five-year plan, a primary goal of which was to build plutonium production reactors. Work began on a reactor at Marcoule in the summer of 1954 and on a plutonium separating plant the following year.

On Dec. 26, 1954, the issue of proceeding with a French bomb was raised at Cabinet level. The outcome was that Prime Minister Pierre Mendès-France launched a secret program to develop a bomb. On Nov. 30, 1956, a protocol was signed specifying tasks the CEA and the Defense Ministry would perform. These included providing the plutonium, assembling a device, and preparing a test site. On July 22, 1958, de Gaulle, who had resumed power as prime minister, set the date for the first atomic explosion to occur within the first three months of 1960. On Feb. 13, 1960, the French detonated their first atomic bomb from a 330-foot tower in the Sahara in what was then French Algeria.

China. On Jan. 15, 1955, Mao Zedong (Mao Tse-tung) and the Chinese leadership decided to obtain their own nuclear arsenal. From 1955 to 1958 the Chinese were partially dependent upon the Soviet Union for scientific and technological assistance, but from 1958 until the break in relations in 1960 they became more and more self-sufficient. Facilities were built to produce and process uranium and plutonium at the Lan-chou Gaseous Diffusion Plant and the Chiu-ch'üan Atomic Energy Complex, both in the northwestern province of Kansu. A design laboratory

(called the Ninth Academy) was established at Hai-yen, east of the Koko Nor in Tsinghai province. A test site at Lop Nor, in far northwestern China, was established in October 1959. Overall leadership and direction was provided by Nie Rongzhen (Nieh Jung-chen), director of the Defense Science and Technology Commission.

Unlike the initial U.S. or Soviet tests, the first Chinese detonation—on Oct. 16, 1964—used uranium-235 in an implosion-type configuration. Plutonium was not used until the eighth explosion, on Dec. 27, 1968.

Other countries. On May 18, 1974, India detonated a nuclear device in the Rājasthān desert near Pokaran with a reported yield of 15 kilotons. India characterized the test as being for peaceful purposes and apparently did not stockpile weapons. Pakistan declared its nuclear program to be solely for peaceful purposes, but it acquired the necessary infrastructure of facilities to produce weapons and was generally believed to possess them.

Nuclear proliferation

Several other countries were believed to have built nuclear weapons or to have acquired the capability of assembling them on short notice. Israel was believed to have built an arsenal of more than 200 weapons, including thermonuclear bombs. In August 1988 the South African foreign minister said that South Africa had "the capability to [produce a nuclear bomb] should we want to." Argentina, Brazil, South Korea, and Taiwan also had the scientific and industrial base to develop and produce nuclear weapons, but they did not seem to have active programs. (W.J.F./T.B.C./R.S.N.)

THE DEVELOPMENT OF FUSION WEAPONS

The United States. U.S. research on thermonuclear weapons started from a conversation in September 1941 between Fermi and Teller. Fermi wondered if the explosion of a fission weapon could ignite a mass of deuterium sufficiently to begin thermonuclear fusion. (Deuterium, an isotope of hydrogen with one proton and one neutron in the nucleus—*i.e.*, twice the normal weight—makes up 0.015 percent of natural hydrogen and can be separated in quantity by electrolysis and distillation. It exists in liquid form only below about −417° F, or −250° C.) Teller undertook to analyze the thermonuclear processes in some detail and presented his findings to a group of theoretical physicists convened by Oppenheimer in Berkeley in the summer of 1942. One participant, Emil Konopinski, suggested that the use of tritium be investigated as a thermonuclear fuel, an insight that would later be important to most designs. (Tritium, an isotope of hydrogen with one proton and two neutrons in the nucleus—*i.e.*, three times the normal weight—does not exist in nature except in trace amounts, but it can be made by irradiating lithium in a nuclear reactor. It is radioactive and has a half-life of 12.5 years.)

From A. DeVolpi *et al., Born Secret: The H-Bomb, the Progressive Case and National Security,* copyright © 1981 Pergamon Press Inc.

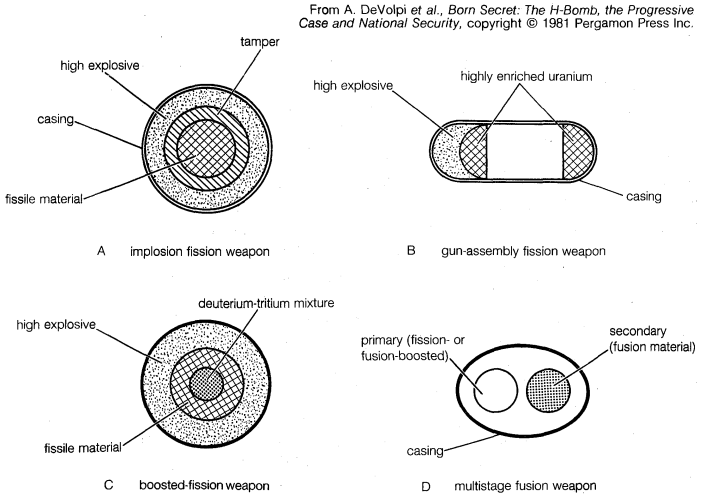


Figure 22: *Idealized representations of four nuclear weapon designs.*
A fission bomb using the implosion method (A) was dropped on Nagasaki; a gun-assembly bomb (B) destroyed Hiroshima. Boosted-fission designs (C) are the principal means of triggering modern thermonuclear weapons. Most modern nuclear weapons are multistage thermonuclear designs (D).

As a result of these discussions the participants concluded that a weapon based on thermonuclear fusion was possible. When the Los Alamos laboratory was being planned, a small research program on the Super, as it came to be known, was included. Several conferences were held at the laboratory in late April 1943 to acquaint the new staff members with the existing state of knowledge and the direction of the research program. The consensus was that modest thermonuclear research should be pursued along theoretical lines. Teller proposed more intensive investigations, and some work did proceed, but the more urgent task of developing a fission weapon always took precedence—a necessary prerequisite for a hydrogen bomb in any event.

In the fall of 1945, after the success of the atomic bomb and the end of World War II, the future of the Manhattan Project, including Los Alamos and the other facilities, was unclear. Government funding was severely reduced, many scientists returned to universities and to their careers, and contractor companies turned to other pursuits. The Atomic Energy Act, signed by President Truman on Aug. 1, 1946, established the Atomic Energy Commission (AEC), replacing the Manhattan Engineer District, and gave it civilian authority over all aspects of atomic energy, including oversight of nuclear warhead research, development, testing, and production.

From April 18 to 20, 1946, a conference led by Teller at Los Alamos reviewed the status of the Super. At that time it was believed that a fission weapon could be used to ignite one end of a cylinder of liquid deuterium and that the resulting thermonuclear reaction would self-propagate to the other end. This conceptual design was known as the "classical Super."

Design problems of a thermonuclear bomb

One of the two central design problems was how to ignite the thermonuclear fuel. It was recognized early on that a mixture of deuterium and tritium theoretically could be ignited at lower temperatures and would have a faster reaction time than deuterium alone, but the question of how to achieve ignition remained unresolved. The other problem, equally difficult, was whether and under what conditions burning might proceed in thermonuclear fuel once ignition had taken place. An exploding thermonuclear weapon involves many extremely complicated, interacting physical and nuclear processes. The speeds of the exploding materials can be up to millions of feet per second, temperatures and pressures are greater than those at the centre of the Sun, and time scales are billionths of a second. To resolve whether the "classical Super" or any other design would work required accurate numerical models of these processes—a formidable task, since the computers that would be needed to perform the calculations were still under development. Also, the requisite fission triggers were not yet ready, and the limited resources of Los Alamos could not support an extensive program.

On Sept. 23, 1949, Truman announced that "we have evidence that within recent weeks an atomic explosion occurred in the U.S.S.R." This first Soviet test stimulated an intense, four-month, secret debate about whether to proceed with the hydrogen bomb project. One of the strongest statements of opposition against proceeding with a hydrogen bomb program came from the General Advisory Committee (GAC) of the AEC, chaired by Oppenheimer. In their report of Oct. 30, 1949, the majority recommended "strongly against" initiating an all-out effort, believing "that extreme dangers to mankind inherent in the proposal wholly outweigh any military advantages that could come from this development." "A super bomb," they went on to say, "might become a weapon of genocide." They believed that "a super bomb should never be produced." Nevertheless, the Joint Chiefs of Staff, the State and Defense departments, the Joint Committee on Atomic Energy, and a special subcommittee of the National Security Council all recommended proceeding with the hydrogen bomb. Truman announced on Jan. 31, 1950, that he had directed the AEC to continue its work on all forms of atomic weapons, including hydrogen bombs. In March, Los Alamos went on a six-day workweek.

In the months that followed Truman's decision, the prospect of actually being able to build a hydrogen bomb became less and less likely. The mathematician Stanislaw M. Ulam, with the assistance of Cornelius J. Everett, had undertaken calculations of the amount of tritium that would be needed for ignition of the classical Super. Their results were spectacular and, to Teller, discouraging: the amount needed was estimated to be enormous. In the summer of 1950 more detailed and thorough calculations by other members of the Los Alamos Theoretical Division confirmed Ulam's estimates. This meant that the cost of the Super program would be prohibitive.

Also in the summer of 1950, Fermi and Ulam calculated that liquid deuterium probably would not burn—that is, there would probably be no self-sustaining and propagating reaction. Barring surprises, therefore, the theoretical work to 1950 indicated that every important assumption regarding the viability of the classical Super was wrong. If success was to come, it would have to be accomplished by other means.

Solving the problem of ignition

The other means became apparent between February and April 1951, following breakthroughs achieved at Los Alamos. One breakthrough was the recognition that the burning of thermonuclear fuel would be more efficient if a high density were achieved throughout the fuel prior to raising its temperature, rather than the classical Super approach of just raising the temperature in one area and then relying on the propagation of thermonuclear reactions to heat the remaining fuel. A second breakthrough was the recognition that these conditions—high compression and high temperature throughout the fuel—could be achieved by containing and converting the radiation from an exploding fission weapon and then using this energy to compress a separate component containing the thermonuclear fuel.

The major figures in these breakthroughs were Ulam and Teller. In December 1950 Ulam had proposed a new fission weapon design, using the mechanical shock of an ordinary fission bomb to compress to a very high density a second fissile core. (This two-stage fission device was conceived entirely independently of the thermonuclear program, its aim being to use fissionable materials more economically.) Early in 1951 Ulam went to see Teller and proposed that the two-stage approach be used to compress and ignite a thermonuclear secondary. Teller suggested radiation implosion, rather than mechanical shock, as the mechanism for compressing the thermonuclear fuel in the second stage. On March 9, 1951, Teller and Ulam presented a report containing both alternatives, entitled "On Heterocatalytic Detonations I. Hydrodynamic Lenses and Radiation Mirrors." A second report, dated April 4, by Teller, included some extensive calculations by Frederic de Hoffmann and elaborated on how a thermonuclear bomb could be constructed. The two-stage radiation implosion design proposed by these reports, which led to the modern concept of thermonuclear weapons, became known as the Teller–Ulam configuration.

It was immediately clear to all scientists concerned that these new ideas—achieving a high density in the thermonuclear fuel by compression using a fission primary—provided for the first time a firm basis for a fusion weapon. Without hesitation, Los Alamos adopted the new program. Gordon Dean, chairman of the AEC, convened a meeting at the Institute for Advanced Study in Princeton, hosted by Oppenheimer, on June 16–18, 1951, where the new idea was discussed. In attendance were the GAC members, AEC commissioners, and key scientists and consultants from Los Alamos and Princeton. The participants were unanimously in favour of active and rapid pursuit of the Teller–Ulam principle.

Just prior to the conference, on May 8 at Enewetak atoll in the western Pacific, a test explosion called George had successfully used a fission bomb to ignite a small quantity of deuterium and tritium. The original purpose of George had been to confirm the burning of these thermonuclear fuels (about which there had never been any doubt), but with the new conceptual understanding contributed by Teller and Ulam, the test provided the bonus of successfully demonstrating radiation implosion.

The first thermonuclear explosion

In September 1951, Los Alamos proposed a test of the Teller–Ulam concept for November 1952. Engineering of the device, nicknamed Mike, began in October 1951, but unforeseen difficulties required a major redesign of the experiment in March 1952. The Mike device weighed 82 tons, owing in part to cryogenic (low-temperature) refrigeration equipment necessary to keep the deuterium in liquid form. It was successfully detonated during Operation Ivy, on Nov. 1, 1952 (local time), at Enewetak. The explosion achieved a yield of 10.4 million tons of TNT, or 500 times larger than the Nagasaki bomb, and it produced a crater 6,240 feet in diameter and 164 feet deep.

With the Teller–Ulam configuration proved, deliverable thermonuclear weapons were designed and initially tested during Operation Castle in 1954. The first test of the series, conducted on March 1, 1954 (local time), was called Bravo. It used solid lithium deuteride rather than liquid deuterium and produced a yield of 15 megatons, 1,000 times as large as the Hiroshima bomb. Here the principal thermonuclear reaction was the fusion of deuterium and tritium. The tritium was produced in the weapon itself by neutron bombardment of the lithium-6 isotope in the course of the fusion reaction. Using lithium deuteride instead of liquid deuterium eliminated the need for cumbersome cryogenic equipment.

With completion of Castle, the feasibility of lightweight, solid-fuel thermonuclear weapons was proved. Vast quantities of tritium would not be needed after all. New possibilities for adaptation of thermonuclear weapons to various kinds of missiles began to be explored.

The Soviet Union. In 1948 Kurchatov organized a theoretical group, under the supervision of physicist Igor Y. Tamm, to begin work on a fusion bomb. (This group included Andrey Sakharov, who, after contributing several important ideas to the effort, later became known as the "father of the Soviet H-bomb.") In general, the Soviet program was two to three years behind that of the United States. The test that took place on Aug. 12, 1953, produced a fusion reaction in lithium deuteride and had a yield of 200 to 400 kilotons. This test, however, was not of a high-yield hydrogen bomb based on the Teller–Ulam configuration or something like it. The first such Soviet test, with a yield in the megaton range, took place on Nov. 22, 1955. On Oct. 30, 1961, the Soviet Union tested the largest known nuclear device, with an explosive force of 58 megatons.

Great Britain. Minister of Defence Harold Macmillan announced in his Statement of Defence, on Feb. 17, 1955, that the United Kingdom planned to develop and produce hydrogen bombs. The formal decision to proceed had been made earlier in secret by a small Defence subcommittee on June 16, 1954, and put to the Cabinet in July. The decision was unaccompanied by the official debate that characterized the American experience five years earlier.

British testing

It remained unclear exactly when the first British thermonuclear test occurred. Three high-yield tests in May and June 1957 near Malden Island in the Pacific Ocean were probably of boosted fission designs (see below). The most likely date for the first two-stage thermonuclear test, using the Teller–Ulam configuration or a variant, was Nov. 8, 1957. This test and three others that followed in April and September 1958 contributed novel ideas to modern thermonuclear designs.

France. Well before their first atomic test, the French assumed they would eventually have to become a thermonuclear power as well. The first French thermonuclear test was conducted on Aug. 24, 1968.

China. Plans to proceed toward a Chinese hydrogen bomb were begun in 1960, with the formation of a group by the Institute of Atomic Energy to do research on thermonuclear materials and reactions. In late 1963, after the design of the fission bomb was complete, the Theoretical Department of the Ninth Academy, under the direction of Deng Jiaxian (Teng Chia-hsien), was ordered to shift to thermonuclear work. By the end of 1965 the theoretical work for a multistage bomb had been accomplished. After testing two boosted fission devices in 1966, the first Chinese multistage fusion device was detonated on June 17, 1967.

REFINEMENTS IN DESIGN

From the late 1940s, U.S. nuclear weapon designers developed and tested warheads to improve their ballistics, to standardize designs for mass production, to increase yields, to improve yield-to-weight and yield-to-volume ratios, and to study their effects. These improvements resulted in the creation of nuclear warheads for a wide variety of strategic and tactical delivery systems.

Fission. The first advances came through the test series Operation Sandstone, conducted in the spring of 1948. These three tests used implosion designs of a second generation, which incorporated composite and levitated cores. A composite core consisted of concentric shells of both uranium-235 and plutonium-239, permitting more efficient use of these fissile materials. Higher compression of the fissile material was achieved by levitating the core—that is, introducing an air gap into the weapon to obtain a higher yield for the same amount of fissile material.

Tests during Operation Ranger in early 1951 included implosion devices with cores containing a fraction of a critical mass—a concept originated in 1944 during the Manhattan Project. Unlike the original Fat Man design, these "fractional crit" weapons relied on compressing the fissile core to a higher density in order to achieve a supercritical mass. These designs could achieve appreciable yields with less material.

Boosted reactions

One technique for enhancing the yield of a fission explosion was called "boosting." Boosting referred to a process whereby thermonuclear reactions were used as a source of neutrons for inducing fissions at a much higher rate than could be achieved with neutrons from fission chain reactions alone. The concept was invented by Teller by the middle of 1943. By incorporating deuterium and tritium into the core of the fissile material, a higher yield could be obtained from a given quantity of fissile material—or, alternatively, the same yield could be achieved with a smaller amount. The fourth test of Operation Greenhouse, on May 24, 1951, was the first proof test of a booster design. In subsequent decades approximately 90 percent of nuclear weapons in the U.S. stockpile relied on boosting.

Fusion. Refinements of the basic two-stage Teller–Ulam configuration resulted in thermonuclear weapons with a wide variety of characteristics and applications. Some high-yield deliverable weapons incorporated additional thermonuclear fuel (lithium deuteride) and fissionable material (uranium-235 and uranium-238) in a third stage. While there was no theoretical limit to the yield that could be achieved from a thermonuclear bomb (for example, by adding more stages), there were practical limits on the size and weight of weapons that could be carried by aircraft or missiles. The largest U.S. bombs had yields of from 10 to 20 megatons and weighed up to 20 tons. Beginning in the early 1960s, however, the United States built a variety of smaller, lighter weapons that exhibited steadily improving yield-to-weight and yield-to-volume ratios.

The neutron bomb

A nuclear explosion releases energy in a variety of forms, including blast, heat, and radiation (X rays, gamma rays, and neutrons). By varying a weapon's design, these effects could be tailored for a specific military purpose. In an enhanced-radiation weapon, more commonly called a neutron bomb, the objective was to minimize the blast by reducing the fission yield and to enhance the neutron radiation. Such a weapon would prove lethal to invading troops without, it was hoped, destroying the defending country's towns and countryside. It was actually a small (on the order of one kiloton), two-stage thermonuclear weapon that utilized deuterium and tritium, rather than lithium deuteride, to maximize the release of fast neutrons. The first U.S. application of this principle was an antiballistic missile warhead in the mid-1970s. Enhanced-radiation warheads were produced for the Lance short-range ballistic missile and for an eight-inch artillery shell.

(T.B.C./R.S.N.)

Chemical and biological weapons

CHEMICAL WARFARE

Chemical warfare agents are substances, whether gaseous, liquid, or solid, intended for use in warfare because of their direct toxic effects on people, animals, or plants. Worldwide revulsion toward chemical weapons is embodied in the Geneva Protocol of 1925, prohibiting "the use in war of asphyxiating, poisonous or other gases, and of all analogous liquids, materials or devices." More than 140 states, including all major nations, are parties to the Protocol of 1925.

History of use. Toxic smokes and other toxic substances were used occasionally in war from ancient times, but the earliest large-scale use of chemical warfare agents was in World War I. Preceded by both sides' sporadic use of various tear gases in artillery and other projectiles starting in 1914, it was the German attack with chlorine released from thousands of cylinders along a four-mile (six-kilometre) front at Ypres on April 22, 1915, that initiated the massive use of chemicals in that conflict. The wind-borne cloud totally broke the lines of the unprepared French Territorial and Algerian units in its path, but the attackers failed to exploit the opportunity. Although numerous chlorine gas-cloud attacks were subsequently carried out by both sides, they accomplished little, owing to the introduction of gas masks and other protective measures.

Chlorine in World War I

As other gases and more effective delivery methods were introduced, so too were improved defenses. Notable among offensive improvements were respiratory agents more poisonous than chlorine, such as phosgene, and chemicals that blistered the skin and attacked the eyes, especially mustard gas. The defense kept pace, with the introduction of better gas masks, protective clothing, and battlefield tactics for minimizing exposure. More than 100,000 tons of various chemical warfare agents were used in World War I; but gas was an unimportant weapon in overall military terms, largely because of the effectiveness of defenses against it.

In World War II, chemical weapons were stockpiled by both sides, but they were not used and were not integrated into military planning. Records indicated various reasons for this: (1) military opinion that chemical weapons would be no more effective than conventional weapons and would complicate and delay operations, (2) fear of retaliation, especially against civilian centres, and (3) aversion to gas warfare by political and military leaders, reflecting the proscriptions of the Geneva Protocol.

Chemical weapons were used in only a few of the more than 200 wars fought after World War I. In each case—as in Ethiopia (1935–36), China (1938–42), the Yemen (1966–67), and Iraq–Iran (1984–88)—chemicals were used against forces initially lacking gas masks.

The weapons. Modern lethal chemical weapons employed the organophosphorus nerve agents first produced but not used by Germany during World War II. Related to certain insecticides but much more toxic to man, they would cause intense sweating, filling of the bronchial passages with mucus, dimming of vision, uncontrollable vomiting and defecation, convulsions, and finally paralysis and respiratory failure. Death would result from asphyxia, generally within a few minutes after respiratory exposure or within hours if exposure was through a liquid nerve agent on the skin.

U.S. and Soviet nerve agents

The U.S. stockpile of chemical warfare agents, loaded into munitions or stored in bulk, included the nerve agents sarin and VX, while the Soviet Union stocked the nerve agents sarin, VX, and soman. Of these three nerve agents (all liquids), sarin would evaporate the most rapidly and would pose mainly a respiratory hazard. VX, the least volatile, would act primarily as a contact poison. Soman, with volatility intermediate between that of sarin and VX, would pose both respiratory and contact hazards.

In addition to nerve agents, both nations stocked mustard gas and the irritant CS, which was also used by police. The Soviets also stocked lewisite, a blister agent developed but not used by the United States during World War I. Mustard gas and lewisite would not be nearly so lethal as the nerve agents, causing casualties principally from incapacitating blisters and temporary blindness. Their full effects would take several hours to develop, although lewisite, in contrast to mustard gas, would cause immediate pain to the skin and eyes.

Liquid chemical warfare agents, such as mustard gas,

lewisite, and the nerve agents, could be loaded into artillery projectiles, bombs, or missile warheads, to be dispersed by an explosive charge as a vapour cloud or a liquid spray. Liquid agents might also be carried in tanks and sprayed from aircraft at low altitude. Greater persistence and more controlled dispersion might be obtained by the addition of thickeners. Solid agents, such as CS, might be dispersed explosively or aerosolized from pyrotechnic mixtures in various munitions.

An innovation put into quantity production by the United States in 1987 was the binary sarin artillery projectile, in which two relatively nontoxic precursors of sarin were held in separate canisters. Upon firing, the two chemicals would mix and react to form sarin. One of the canisters might be stored and shipped separately, to be inserted into the projectile at the ammunition depot or the gun site. This built-in safety feature was intended to provide greater operational flexibility in the storage and transport of the weapon. The binary principle could be applied to other types of chemical warfare agent.

The amount of a chemical warfare agent required to create a hazardous cloud over a target area would be highly dependent on air movements. The weight of sarin, for example, required to produce a lethal respiratory hazard to unprotected persons over most of an open mile-square area could be between 0.3 and 10 tons, depending on atmospheric conditions. As an illustration, the delivery of these amounts by 155-millimetre artillery would require the firing of approximately 100 to 3,000 projectiles.

For causing casualties to unprotected troops, chemicals could be more effective than an equivalent weight of conventional high-explosive fragmentation weapons. For troops with good protection, however, the reverse would be true; soldiers with modern antichemical protection would be far less vulnerable to chemicals than to conventional weapons.

Defense. The first and most important line of defense against chemical warfare agents (also needed for protection against radioactive fallout) was the individual protection provided by masks and protective clothing, and the collective protection of combat vehicles and mobile or fixed shelters. Filters for masks and shelters contained specially treated activated charcoal to remove vapours, and paper membranes or other materials to remove particles. Such filters typically could reduce the concentration of chemical (and biological) warfare agents by a factor of at least 100,000. Masks could be donned in less than 10 seconds and could be worn for long periods, even in sleep. Modern protective overgarments were made of fabric containing activated charcoal or other adsorptive forms of carbon. A complete suit typically weighed about four pounds (two kilograms). The fabric could breathe and pass water-vapour perspiration. In warm weather, periods of heavy exertion in full protective gear would have to be limited in order to avoid heat stress, or else protection would have to be partly relaxed, as by partially opening the protective jacket. Under common European conditions, military units routinely exercised at or near full protection for several days continuously.

Effectiveness of masks and overgarments

Other items for chemical defense were detectors and alarms sensitive to nerve and blister agents, prophylactic and antidote drugs that would provide partial protection against nerve agents, and equipment for decontaminating people and equipment.

The effectiveness of chemical weapons against prepared forces would depend more on the interference with fighting performance imposed by wearing protective equipment and taking other precautions than on direct casualties. The extent of such interference, and hence the military value of chemicals in comparison with other weapons, was difficult to assess. Estimates, based on controlled field exercises, of the reduction in performance in military units under chemical attack ranged from near zero to more than 30 percent, depending on the mission and the conditions of the exercise.

BIOLOGICAL WARFARE

Biological warfare agents are infectious microbes, including viruses, bacteria, and fungi, intended for use in warfare because of their pathogenic effects on people, animals, or plants. The development, production, and stockpiling of weapons based on them were outlawed by the 1972 Biological Weapons Convention, to which more than 100 states were party, including all five permanent members of the United Nations Security Council. The treaty also covered weapons based on naturally occurring poisons, known as toxins, however produced. As with chemical weapons, actual employment of biological weapons was outlawed by the 1925 Geneva Protocol.

Banning of biological weapons

At the time of their destruction in accordance with presidential directives of 1969 and 1970, the biological weapons of the United States (the only country for which authenticated information was available) included dry-powder or liquid-slurry formulations of the microbes that cause tularemia, Q fever, Venezuelan equine encephalitis, rice blast, and stem rust of wheat. They also included a number of toxins, such as paralytic shellfish poison. A variety of dispensers, both large and small, was also on hand.

Biological weapons designed to dispense airborne clouds of pathogenic microbes could in theory kill or incapacitate unprotected populations over very large areas. Such weapons were never used. (Ma.Me./J.P.P.R.)

Fortification

Fortifications are military positions that have been strengthened against attack. They are usually of two types: permanent and field. Permanent fortifications include elaborate forts and troop shelters; they are most often erected in times of peace or upon threat of war. Field fortifications are constructed when in contact with an enemy or when contact is imminent. They consist of entrenched positions for personnel and crew-served weapons, cleared fields of fire, and obstacles such as explosive mines, barbed-wire entanglements, felled trees, and antitank ditches. Both field and permanent fortifications often take advantage of natural obstacles, such as canals and rivers, and they are usually camouflaged or otherwise concealed. Both types are designed to assist the defender to obtain the greatest advantage from his own strength and weapons while preventing the enemy from using his resources to best advantage.

This section discusses military fortification since the introduction of rifled artillery and small arms. For discussions of fortification up to the modern era, see above *Military technology before the modern era: Antiquity and the classical age; The age of cavalry;* and *The gunpowder revolution.*

TRENCH WARFARE, 1860–1918

The American Civil War. In the American Civil War, field fortifications emerged as an essential of warfare, with both armies employing entrenchments to an extent never before seen. Troops learned to fortify newly won positions immediately; employing spades and axes carried in their packs, they first dug rifle pits and then expanded them into trenches. Early in the war, General Robert E. Lee adopted the frontier rifleman's breastwork composed of two logs on the parapet of the entrenchment, and many of Lee's victories were the result of his ability to use hasty entrenchments as a base for aggressive employment of fire and maneuver. Two notable sieges, that of Vicksburg, Miss., in the west, and Petersburg, Va., in the east, were characterized by the construction of extensive and continuous trench lines that foreshadowed those of World War I. In the Cold Harbor, Va., campaign, when General Ulysses S. Grant sent his troops against Confederate earthworks, he lost 14,000 men in 13 days. Field mines and booby traps were used extensively, and trench mortars were developed to lob shells into opposing trenches.

The rifleman's breastwork

World War I. The lesson taught by accurate, long-range fire from cntrenched positions in the American Civil War was lost on European commanders. Even the bitter experiences of appalling losses in the Crimean, Franco-German, and Boer wars failed to lessen an ardour for the theory of the offensive that was so fervent as to leave little concern for defensive tactics in the field. Few took notice of the immense casualties the Turks inflicted from behind

field fortifications in the Russo-Turkish War of 1877–78, and even though the Russo-Japanese War soon after the turn of the century underscored the lethal power of the machine gun and breech-loading rifled artillery, most European commanders saw the increased firepower as more a boon to the offensive than to the defensive.

The fallacy of the faith in offensive firepower was soon convincingly demonstrated. Once the French had checked the German right wing at the Marne River, the fighting degenerated into what was in effect a massive siege. For 600 miles (1,000 kilometres), from Switzerland to the North Sea, the landscape was soon scarred with opposing systems of zigzag, timber-revetted, sandbag-reinforced trenches, fronted by tangles of barbed wire sometimes more than 150 feet (45 metres) deep and featured here and there by covered dugouts providing shelter for troops and horses and by observation posts in log bunkers or concrete turrets. The trench systems consisted of several lines in depth, so that if the first line was penetrated, the assailants were little better off. Rail and motor transport could rush fresh reserves forward to seal off a gap faster than the attackers could continue forward. Out beyond the trenches and the barbed wire was a muddy, virtually impassable desert called no-man's-land, where artillery fire soon eliminated habitation and vegetation alike. The fighting involved masses of men, masses of artillery, and masses of casualties. Toxic gases—asphyxiating, lachrymatory, and vesicant—were introduced in a vain effort to break the dominance of the defense, which was so overpowering that for more than two years the opposing lines varied less than 10 miles in either direction.

The Hindenburg Line

During the winter of 1916–17, the Germans prepared a reserve trench system, the Hindenburg Line, containing deep dugouts where the men could take cover against artillery fire and machine guns emplaced in concrete shelters called pillboxes. Approximately two miles behind the forward line was a second position, almost as strong. The Hindenburg Line resisted all Allied assaults in 1917, including a vast British mining operation under the Messines Ridge in Belgium that literally blew up the ridge, inflicting 17,000 casualties at one blow; the advance failed to carry beyond the ridge.

PERMANENT FORTIFICATION, 1914–45

World War I. Most defensive thinking on the eve of World War I was reserved for the permanent fort, which was designed to canalize enemy advance and to afford time for national mobilization. The leading fortification engineer of the time was a Belgian, Henri Brialmont. He placed his forts, built of concrete, at an average distance of four miles from a city, as with 12 forts at Liège, and at intervals of approximately 2.5 miles. At Antwerp his defense system was even more dense. He protected the big guns of his forts with turrets of steel and developed disappearing cupolas. Some forts were pentagonal, others triangular, with much of the construction underground.

In building defenses along the frontier facing Germany, French engineers emulated Brialmont, with particularly strong clusters of fortresses at Verdun and Belfort. So monstrous were the forts of the time that they were known as "land battleships." But by marching through Belgium with a strong right wing (the Schlieffen plan), the Germans circumvented the powerful French fortresses. Passing between the forts at Liège, which Brialmont had intended to be connected with trenches, they took the city in only three days, then systematically reduced the forts. Namur, also heavily fortified, resisted the powerful Big Bertha guns for only four days. The concrete of the Belgian fortifications crumbled under the pounding, but the French forts at Verdun, of more recent and sturdier construction, later absorbed tremendous punishment and served as focal points for some of the war's bloodiest fighting.

Linear fortifications of World War II. *The Maginot Line and the West Wall.* In the interval between world wars, several European countries built elaborate permanent fortifications. The largest was the French Maginot Line, a system of mammoth, self-contained forts stretching from Switzerland to the vicinity of the Belgian frontier near Montmédy. The reinforced concrete of the forts was thicker than any theretofore used, the disappearing guns bigger and more heavily armoured. Ditches, embedded steel beams, and minefields guarded against tank attack. A large part of the works were completely underground. Outposts were connected to the main forts by concrete tunnels. But, because French and British military leaders were convinced that if war came again with Germany the Allies would fight in Belgium, the French failed to extend the line to the sea, relying instead on an outmoded system of unconnected fortresses left over from before World War I. It was this weakness that the Germans subsequently exploited in executing a modified version of the Schlieffen plan, cutting in behind the permanent defenses and defeating France without having to come to grips with the Maginot Line.

The Germans confronted that portion of the Maginot Line facing the Saar River with fortifications of their own, the West Wall. Later extended northward to the Dutch frontier and southward along the Rhine to Switzerland, the West Wall was not a thin line of big forts but a deep band, up to five miles thick, of more than 3,000 small, mutually supporting pillboxes, observation posts, and troop shelters. For passive antitank defense the line depended upon natural obstacles, such as rivers and lakes, and upon "dragon's teeth," five rows of pyramid-shaped reinforced concrete projections.

Pillboxes and tank obstacles

The Germans did not rely on the West Wall to halt an attack but merely to delay it until counterattacks by mobile reserves could eliminate any penetration. The value of their concept remains undetermined; the line was not attacked until late 1944, after the German armies had incurred severe defeats and lacked adequate reserves. The West Wall nevertheless forced Allied troops into costly attacks to eliminate it.

Other fort series. Elsewhere in World War II many fortifications similar to these two basic types were built. The Italians constructed a series of new fortifications and modernized existing World War I defenses along the country's mountainous northern and northeastern frontiers; the Finns maintained a World War I defense facing the Soviet Union, the Mannerheim Line (named after a Finnish marshal and statesman); the Soviets built the Stalin Line facing Poland; the Czechoslovaks constructed what became known as the Little Maginot Line to oppose Germany; the Greeks built the Metaxas Line facing Bulgaria; and the Belgians erected a series of elaborate forts along the Albert Canal. German capture of the most elaborate and allegedly impregnable of the Belgian forts, Eben Emael, in a matter of hours in the first two days of the campaign against France and the Low Countries in 1940 startled the world. Arriving silently on the night of May 10 in gliders, troops landed atop the fort and began systematically to destroy turrets and casemates. Soon after daylight they were joined by 300 men arriving by parachute. Around noon of May 11 the 1,000-man garrison surrendered.

Despite at least comparable surprise and the same so-called blitzkrieg methods, the Germans required more time to penetrate the more dispersed forts of the Stalin Line in the Soviet Union. The delay gained two months of invaluable time for the Soviet troops, without which they might well have been unable to stop the Germans at the gates of Moscow.

German channel defenses. The Germans employed Fritz Todt, the engineer who had designed the West Wall, and thousands of impressed labourers to construct permanent fortifications along the Belgian and French coasts facing the English Channel; this was the Atlantic Wall. The line consisted primarily of pillboxes and gun emplacements embedded in cliffsides or placed on the waterfronts of seaside resorts and ports. Included were massive blockhouses with disappearing guns, newsreels of which the Germans sent out through neutral sources in an effort to awe their adversaries, but the numbers of big blockhouses actually were few. Behind the line, in likely landing spots for gliders and parachutists, the Germans emplaced slanted poles, which the troops called *Rommelsspargel* (Rommel's asparagus), after their commander Field Marshal Erwin Rommel. Embedded in the sand of the beaches below the high-tide mark were numerous obstacles, varying in shape

Obstacles to air and amphibious assault

and depth, some topped with mines. Barbed wire and antitank and antipersonnel mines interlaced the whole. On the French southwestern and southern coasts similar, though less formidable, defenses were erected.

When the Allies landed in force on the Cotentin Peninsula of Normandy on D-Day—June 6, 1944—they found the defenses far less formidable than they had anticipated. This was attributable to a number of reasons. The Germans had constructed the strongest defenses in the Pas-de-Calais region facing the narrowest part of the English Channel and had stationed their most battleworthy troops there; demands of other fighting fronts had siphoned many of the best German troops from France; the Germans lacked air and naval support; Allied airpower was so strong that movement of German reserves was seriously impeded; landings of Allied airborne troops behind the beaches spread confusion in German ranks; and the Germans were deluded into believing the invasion was a diversion, that a second and larger invasion was to follow in the Pas-de-Calais. Only at one of the two American beaches, given the code name Omaha, was the success of the landing ever in doubt, partly because of rough seas, partly because of the chance presence of an elite German division, and partly because of the presence of high bluffs. Paradoxically, the Allies had less difficulty with the highly publicized beach defenses than they had later with field fortifications based on the Norman hedgerows, earthen embankments several feet thick and five feet high that localfarmers through the centuries had erected around thousands of irregularly shaped little fields to fence their cattle and protect their crops from strong ocean winds.

NUCLEAR FORTIFICATION

At the close of World War II most military theorists considered that permanent fortifications of the type previously employed were economically impracticable in view of their vulnerability to the incredible power of nuclear explosives and the methods, such as vertical envelopment from the air, that might be employed to reduce them. Important exceptions to this generalization were the reinforced concrete and deep tunnels used to protect strategic-missile launch facilities. The United States, the Soviet Union, and (to a lesser degree) France, Great Britain, Israel, and China invested heavily in such defensive works.

The missile silo

Probably the most important and most characteristic of these works was the missile silo, a tubular structure of heavily reinforced concrete sunk into the ground to serve as a protective installation and launch facility for a single intercontinental ballistic missile (ICBM). These silos were "hardened" to resist a calculated amount of blast and shock from a nuclear detonation. Launch crews were protected in similarly constructed underground bunkers nearby. Elaborate calculations on the number of ICBM warheads needed to destroy a hardened silo with a given degree of certainty became an integral part of the strategic calculus in the 1960s. In this way, permanent fortifications resumed their previous place of importance in strategic calculations.

Of particular concern to strategists of the United States and the Soviet Union was the vulnerability of land-based ICBMs to preemptive nuclear attack. Elaborate defensive works were proposed to protect them. One basing scheme involved a network of fortified missile shelters connected by roads or railroad tracks. Huge, closed missile transporters would shuttle the missiles from one shelter to another in such a manner that the enemy would not know which shelters were occupied and which were empty. An even more extreme plan for protecting the U.S. land-based ICBM force was designed around fratricide, the theory that multiple nuclear explosions cannot occur at the same time in close proximity to one another since the first detonated warhead triggers low-yield partial explosions in the others. The proposal, called dense pack, would exploit this phenomenon by packing a large number of super-hardened ICBM silos closely together in a single location.

Other permanent fortifications of the nuclear age were designed as headquarters sites or command and control installations. For example, a joint U.S.-Canadian project, the North American Air Defense Command (Norad), included a series of radar posts across northern Canada and Alaska to provide early warning of the approach of hostile bombers or missiles. The system and the aircraft and missiles supporting it were controlled from a vast underground complex embedded in rock of Cheyenne Mountain near Colorado Springs, Colo.

(W.H.B./C.B.MacD./J.F.G.)

Tanks and armoured vehicles

Tanks are heavily armed and armoured combat vehicles that move on two endless metal chains called tracks. They are the principal type of armoured vehicle. Other major types include tracked and wheeled infantry carriers, which were conceived only for transporting troops into action but which also have been used to some extent for fighting; armoured cars, which often resemble the lighter types of tanks but run on wheels; self-propelled guns, some of which are only partly armoured or wheeled; and such specialized vehicles as armoured recovery vehicles, bridgelayers, armoured bulldozers, and amphibious landing vehicles.

TANKS

Tanks are essentially tracked, protected weapon platforms that make the weapons mounted in them more effective by their cross-country mobility and by the protection they provide for their crews. Weapons mounted in tanks have ranged from single rifle-calibre machine guns to, in recent years, long-barreled guns of 120- or 125-millimetre (4.72- or 4.92-inch) calibre.

Earliest developments. The use of vehicles for fighting dates to the 2nd millennium BC, when horse-drawn war chariots were used in the Middle East by the Egyptians, Hittites, and others as mobile platforms for combat with bows and arrows. The concept of protected vehicles can be traced back through the wheeled siege towers and battering rams of the Middle Ages to similar devices used by the Assyrians in the 9th century BC. The two ideas began to merge in the battle cars proposed in 1335 by Guido da Vigevano, in 1484 by Leonardo da Vinci, and by others, down to James Cowen, who took out a patent in England in 1855 for an armed, wheeled, armoured vehicle based on the steam tractor.

Early armoured fighting vehicles

But it was only at the beginning of the 20th century that armoured fighting vehicles began to take practical form. By then the basis for them had become available with the appearance of the traction engine and the motor car. Thus, the first self-propelled armoured vehicle was built in 1900 in England when John Fowler & Company armoured one of their steam traction engines for hauling supplies in the South African (Boer) War (1899–1902). The first motor vehicle used as a weapon carrier was a powered quadricycle on which F.R. Simms mounted a machine gun in 1899 in England. The inevitable next step was a vehicle that was both armed and armoured. Such a vehicle was constructed to the order of Vickers, Son and Maxim Ltd. and was exhibited in London in 1902. Two years later a fully armoured car with a turret was built in France by the Société Charron, Girardot et Voigt, and another was built concurrently in Austria by the Austro-Daimler Company.

To complete the evolution of the basic elements of the modern armoured fighting vehicle, it only remained to adopt tracks as an alternative to wheels. This became inevitable with the appearance of the tracked agricultural tractor, but there was no incentive for this until after the outbreak of World War I. A tracked armoured vehicle was proposed in France as early as 1903 but failed to arouse the interest of military authorities, as did a similar proposal made in England in 1908. Three years later a design for a tracked armoured vehicle was rejected by the Austro-Hungarian and then by the German general staffs, and in 1912 the British War Office turned down yet another design.

World War I. The outbreak of World War I in 1914 radically changed the situation. Its opening stage of mobile warfare accelerated the development of armoured cars, numbers of which were quickly improvised in Belgium, France, and Britain. The ensuing trench warfare,

which ended the usefulness of armoured cars, brought forth new proposals for tracked armoured vehicles. Most of these resulted from attempts to make armoured cars capable of moving off roads, over broken ground, and through barbed wire. The first tracked armoured vehicle was improvised in July 1915, in Britain, by mounting an armoured car body on a Killen-Strait tractor. The vehicle was constructed by the Armoured Car Division of the Royal Naval Air Service, whose ideas, backed by the First Lord of the Admiralty, Winston S. Churchill, resulted in the formation of an Admiralty Landships Committee. A series of experiments by this committee led in September 1915 to the construction of the first tank, called "Little Willie." A second model, called "Big Willie," quickly followed. Designed to cross wide trenches, it was accepted by the British Army, which ordered 100 tanks of this type (called Mark I) in February 1916 (see Figure 23).

First tracked armoured vehicle

Simultaneously but independently, tanks were also developed in France. Like the very first British tank, the first French tank (the Schneider) amounted to an armoured box on a tractor chassis; 400 were ordered in February 1916. But French tanks were not used until April 1917, whereas British tanks were first sent into action on Sept. 15, 1916. Only 49 were available, and their success was limited, but on Nov. 20, 1917, 474 British tanks were concentrated at the Battle of Cambrai and achieved a spectacular breakthrough. These tanks, however, were too slow and had too short an operating range to exploit the breakthrough. In consequence, demand grew for a lighter, faster type of tank, and in 1918 the 14-ton Medium A appeared with a speed of eight miles (13 kilometres) per hour and a range of 80 miles. After 1918, however, the most widely used tank was the French Renault F.T., a light six-ton vehicle designed for close infantry support (see Figure 23).

When World War I ended in 1918, France had produced 3,870 tanks, and Britain 2,636. Most French tanks survived into the postwar period; these were the Renault F.T., much more serviceable than their heavier British counterparts. Moreover, the Renault F.T. fitted well with traditional ideas about the primacy of the infantry, and the French army adopted the doctrine that tanks were a mere auxiliary to infantry. France's lead was followed in most other countries; the United States and Italy both assigned tanks to infantry support and copied the Renault F.T. The U.S. copy was the M1917 light tank, and the Italian the Fiat 3000. The only other country to produce tanks by the end of the war was Germany, which built about 20.

Light infantry-support tanks

Interwar developments. The Renault F.T. remained the most numerous tank in the world into the early 1930s. Aware of the need for more powerful vehicles, if only for leading infantry assaults, the French army took the lead in developing well-armed tanks. The original 1918 French Schneider and St. Chamond tanks already had 75-millimetre guns while the heavier British tanks were at best armed with 57-millimetre guns. After the war the French built 10 68-ton 2C tanks with the first turret-mounted 75-millimetre guns and continued to develop 75-millimetre-gun tanks, notably the 30-ton Char B of 1936.

In the meantime, Britain took the lead, technically and tactically, in developing the mobility of tanks. Even before World War I had ended, work had started on the Medium D with a maximum speed of 20 miles per hour. Between 1923 and 1928 the British army ordered 160 of the new Vickers Medium tanks. They were virtually the only tanks the British army had until the early 1930s and the only tanks to be produced in quantity anywhere in the world during the mid-1920s. The Vickers Mediums stimulated the Royal Tank Corps to develop mobile tactics, and various experiments during the 1920s and early '30s resulted in the general adoption of two categories of tanks. Mobile tanks were intended for the role performed earlier by horse cavalry, while slower but more heavily armoured tanks provided infantry support.

Before this division into mobile and slow tanks had crystallized, several different designs were tried. The British Independent tank of 1925, with five turrets, started a trend toward multiturreted heavy tanks. Another trend setter was a small turretless tankette, originated in Britain by Major Giffard le Quesne Martel and John Carden in the mid-1920s, and a slightly heavier, turreted two-man light tank. The number of light tanks grew rapidly after 1929, as several countries started to produce armoured vehicles. The Soviet Union was by far the most important producer; on a much smaller scale Poland, Czechoslovakia, and Japan entered the field in 1930–31. Concurrently, tank production started up again in France and Italy. As tank production grew and spread among nations, the value of light tanks armed only with machine guns decreased, and heavier models armed with 37- to 47-millimetre guns for fighting other tanks began to displace them. An early example was the Vickers-Armstrong six-ton model of 1930, copied on a large scale in the Soviet Union (as the T-26). The most successful example was the BT, also built in large numbers in the Soviet Union. The fastest tank of its day, the BT was based on designs evolved in the United States by J.W. Christie, who in 1928 built an experimental model capable of 42.5 miles per hour. Christie's vehicles could run on wheels after the removal of tracks and, far more significant, had road wheels independently suspended. This enabled them to move over broken ground faster than tanks with the earlier types of suspension.

Although they were relatively well armed and mobile, tanks of the T-26 and BT type were lightly armoured (plates 10 to 15 millimetres thick) and were not, therefore, suitable for close infantry support. This was clearly

By courtesy of the Imperial War Museum, London; photographs, Camera Press

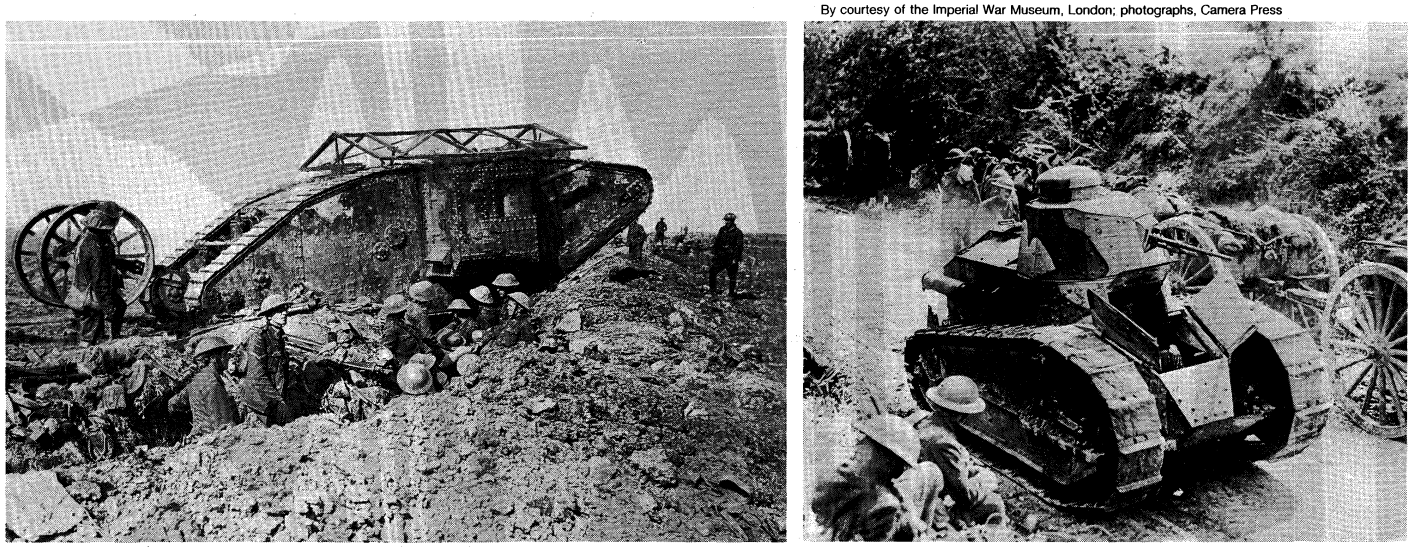


Figure 23: *Tanks of World War I.*
(Left) British Mark I tank with antibomb roof and "tail," 1916. (Right) French Renault F.T. light tank, 1918.

demonstrated in 1937 during the Civil War in Spain, where T-26 and BT tanks were used by the Republican forces. Even before this time, it had become clear that tanks that moved at the slow pace of the infantry and were therefore exposed to the full effect of antitank guns had to be thickly armoured. This realization led in the mid-1930s to such infantry tanks as the French R-35 with 40-millimetre and the British A.11 with up to 60-millimetre armour.

Heavily armed tank-fighting tanks

Apart from being lightly armoured, the Soviet BT, the equivalent British cruiser tanks, and the German Pz. III also required support from more heavily armed tanks if they were to engage in fighting of any intensity. The need for tanks with more powerful 75-millimetre guns was clearly recognized in Germany, leading in 1934 to the design of the Pz. IV. The problem was realized less clearly in the Soviet Union, even though the T-28 and T-35 multi-turret tanks with 76-millimetre guns were first built there in 1932–33. But the Russians recognized more quickly than others the need for the next step, which was to replace all the light-medium tanks armed with 37- to 47-millimetre guns by medium tanks armed with 75- or 76-millimetre guns. Thus, in 1939, while the Germans were still developing the Pz. III from a 37-millimetre to a 50-millimetre version, the Russians were already concentrating on the T-34 medium tank with a 76-millimetre gun.

Other armies were farther behind in producing well-armed tanks on the eve of World War II. All but 80 of the 1,148 tanks that Britain had produced between 1930 and 1939 were still armed only with machine guns. Italy was even worse off, with only 70 M/11 tanks with 37-millimetre guns while the rest of its total of 1,500 were small, machine-gun-armed tankettes. The United States had only about 300 machine-gun-armed light tanks. Most of the 2,000 tanks produced in Japan were equally lightly armed. By comparison, France had a more powerful tank force—2,677 modern tanks, of which, however, only 172 were the Char B, armed with 75-millimetre guns. The largest force was the Soviet Union's, which, as a result of a massive production program, started in 1930–31, had about 20,000 tanks by 1939, considerably more than the rest of the world put together.

World War II. The most effective tank force proved to be the German, composed in 1939 of 3,195 vehicles, including 211 Pz. IVs (see Figure 24). What made the German tanks so formidable was that instead of being divided between various infantry and cavalry tank units they were all concentrated and used in massed formations in the panzer divisions. The successes of the panzer divisions during the first two years of World War II led the major armies to reorganize most of their tanks into similar formations; this resulted in a dramatic increase in production.

The campaigns of 1939–41, in which armoured forces played such an important role, also intensified the technical development of tanks and other armoured vehicles. The German Pz. IV and Soviet T-34 were rearmed in 1942 with longer-barreled, higher-velocity guns; soon afterward these began to be displaced by more powerfully armed tanks. In 1943 the Germans introduced the Panther medium tank with a long 75-millimetre gun having a muzzle velocity of 3,070 feet (936 metres) per second, compared with 1,260 feet per second for the original Pz. IV and 2,460 feet per second for its 1942 version. The 43-ton Panther weighed almost twice as much as its predecessor and was correspondingly better armoured. Germany also introduced the still more powerful Tiger tank, armed with an 88-millimetre gun. Its final version (Tiger II), at 68 tons, was to be the heaviest tank used during World War II. To oppose it, the Russians brought out the JS, or Stalin, heavy tank, which appeared in 1944 armed with a 122-millimetre gun. Its muzzle velocity was lower than that of the German 88-millimetre guns, however, and it weighed only 46 tons. At about the same time the T-34 was rearmed with an 85-millimetre gun.

German Panther and Tiger tanks

In contrast to the breakthrough role of the earlier heavy tanks, the Tiger and JS tanks functioned chiefly to support basic medium tanks by destroying enemy tanks at long range. German and Soviet armies also developed other heavy vehicles for this purpose, such as the 128-millimetre-gun Jagdtiger and the 122-millimetre-gun ISU, which, in effect, were turretless tanks. In addition, all armies developed lightly armoured self-propelled antitank guns. The U.S. Army developed a specialized category of tank destroyers that resembled self-propelled guns in being relatively lightly armoured but that, like tanks, had rotating turrets.

The turretless-tank type of vehicle originated with the Sturmgeschutz, or assault gun, introduced by the German army for infantry support but subsequently transformed into more versatile vehicles particularly suited for destroying enemy tanks. No such vehicles were produced in Britain or the United States. Throughout the war, however, the British army retained a specialized category of infantry tanks, such as the Churchill, and of cruiser tanks, such as the Crusader and Cromwell. The former were well-armoured, and the latter were fast, but none was well-armed compared to German and Soviet tanks. As a result, during 1943 and 1944, British armoured divisions were mostly equipped with U.S.-built M4 Sherman medium tanks. The M4 was preceded by the mechanically similar M3 medium, which was also armed with a medium-velocity 75-millimetre gun but mounted it in the hull instead of the turret, because this could be put into production more quickly when tanks were urgently required in 1940 and 1941. Production of the M4 began in 1942, and eventually 49,234 were built, making it the principal tank of U.S. and other allied armoured forces. Successful when first introduced, it was by 1944 no longer adequately armed and should have been replaced by a new medium tank. But the U.S. Army, like the British, adhered to the fallacious doctrine that armoured divisions should confine themselves to exploitation of infantry breakthroughs and did not, therefore, need powerfully armed tanks. Only

The U.S. M4 Sherman tank

By courtesy of (right) the U.S. Army; photograph, (left) Ullstein

Figure 24: *Tanks of World War II.*
(Left) German Pz. IV (foreground) and Pz. III (background) tanks, 1942. (Right) U.S. Army M26 Pershing tank with 90-millimetre gun (foreground) and M4 Sherman tank with 75-millimetre gun.

toward the end of the war did the U.S. Army introduce a few M26 Pershing heavy tanks with a 90-millimetre gun comparable to that of the original German Tiger (see Figure 24). Similarly, the British army introduced the prototypes of the Centurion tank with a 76-millimetre gun comparable to that of the German Panther. Otherwise, U.S. and British tanks were well behind the German and Soviet tanks in their gun power.

Postwar tank design. After World War II it was generally recognized that all tanks must be well-armed to fight enemy tanks. This finally ended the division of tanks into under-gunned categories of specialized infantry and cavalry tanks, which the British army retained longer than any other. Still not fully recognized, however, were the advantages of concentrating tanks in fully mechanized formations, and the British and U.S. armies continued to divide tanks between the armoured divisions and the less mobile infantry divisions. After World War II, tanks also suffered from one of the periodic waves of pessimism about their future. New antitank weapons, such as rocket launchers and recoilless rifles, and the mistaken belief that the value of tanks lay primarily in their armour protection, caused this attitude. The Soviet army, however, maintained large armoured forces, and the threat they posed to western Europe as the Cold War became more intense, together with the havoc created by Soviet-built T-34/85 tanks during the North Korean invasion of South Korea in 1950, provided a new impetus to development.

The development of tactical nuclear weapons in the mid-1950s provided further stimulus to the development of tanks and other armoured vehicles. Nuclear weapons encouraged the use of armoured forces because of the latter's mobility and high combat power in relation to their vulnerable manpower. Moreover, armoured vehicles proved capable of operating in relative proximity to nuclear explosions by virtue of their protection against blast and radioactivity.

As less emphasis was placed after a time on nuclear weapons and more on conventional forces, tanks retained their importance. This was based on their being recognized, particularly from the early 1970s, as the most effective counter to other armoured forces, which formed the principal threat posed on the ground by potential aggressors.

Armament. In keeping with the importance attached to the ability of tanks to defeat enemy tanks, great emphasis was placed after World War II on their armament. The result was progressive increases in the calibre of tank guns, the development of new types of ammunition with greater armour-piercing capabilities, and the introduction of more sophisticated fire-control systems to improve tank guns' ability to hit targets.

Growing calibre of tank guns

Increases in gun calibre are well illustrated by the British Centurion, which started in 1945 with a 76-millimetre gun but in 1948 was rearmed with an 83.8-millimetre gun and in 1959 with an even more powerful 105-millimetre gun. Moreover, during the 1950s the capabilities of British tank units were augmented by a small number of Conqueror heavy tanks armed with 120-millimetre guns, and in the early 1970s the Centurions were entirely replaced by Chieftains armed with a new and more effective 120-millimetre gun.

Similar increases took place in the calibre of Soviet tank guns. After World War II the basic T-34/85 tanks armed with 85-millimetre guns were replaced by T-54 and T-55 tanks armed with 100-millimetre guns. They were followed in the 1960s by the T-62, with a 115-millimetre gun, and in the 1970s and '80s by the T-64, T-72, and T-80, all with 125-millimetre smoothbore guns. The JS-3 and T-10 heavy tanks with their less powerful 122-millimetre guns had by then been withdrawn. This left the Soviet army in the same position as others of having a single type of battle tank as well armed as contemporary technology would allow (see Figure 25).

For a time the U.S. Army also subscribed to a policy of developing heavy as well as medium tanks. But the heavy M103 tank, armed with a 120-millimetre gun, was only built in small numbers in the early 1950s. As a result, virtually the only battle tanks the U.S. Army had were the M46, M47, and M48 medium tanks, all armed with 90-millimetre guns. After the mid-1950s the M47 tanks were passed on to the French, Italian, Belgian, West German, Greek, Spanish, and Turkish armies, and during the 1960s the M48 began to be replaced by the M60, which was armed with a U.S.-made version of the 105-millimetre gun developed for the British Centurion.

The same 105-millimetre gun was adopted for the Pz. 61 and Pz. 68 tanks produced in Switzerland, the West German Leopard 1, the Swedish S-tank, the Japanese Type 74, and the Mark 1 and 2 versions of the Israeli Merkava. It was also retained in the original version of the U.S. M1 tank developed in the 1970s, but the subsequent M1A1 version of the 1980s was rearmed with a 120-millimetre gun originally developed in West Germany for the Leopard 2 tank (see Figure 25). The British Challenger, introduced in the 1980s, was also armed with 120-millimetre guns, but these were still of the rifled type.

Armour-piercing ammunition

The last years of World War II saw the development of more effective antitank ammunition with armour-piercing, discarding-sabot (APDS) projectiles. These had a smaller-calibre, hard tungsten carbide core inside a light casing. The casing fell away on leaving the gun barrel, while the core flew on at an extremely high velocity. The APDS, which was adopted for the 83.8-millimetre gun of the Centurions, was fired with a velocity of 4,692 feet per second. By comparison, earlier full-calibre, armour-piercing projectiles had a maximum muzzle velocity of about 3,000 feet per second. With this shell the Centurion's 83.8-millimetre gun could penetrate armour twice as thick as could the 88-millimetre gun of the German Tiger II of World War II.

An alternative type of armour-piercing ammunition developed during the 1950s was the high-explosive antitank (HEAT) shell. This shell used a shaped charge with a conical cavity that concentrated its explosive energy into a very-high-velocity jet capable of piercing thick armour. The HEAT round was favoured by the U.S. Army for its 90-millimetre tank guns and also by the French army for the 105-millimetre gun of its AMX-30 tank, introduced in the mid-1960s. However, during the 1970s both APDS and HEAT began to be superseded by armour-piercing, fin-stabilized, discarding-sabot (APFSDS) ammunition. These projectiles had long-rod penetrator cores of tungsten alloy or depleted uranium; they could be fired with muzzle velocities of 5,400 feet per second or more, making them capable of perforating much thicker armour than all earlier types of ammunition.

During the 1960s, attempts were made to arm tanks with guided-missile launchers. These were to provide tanks with a combination of the armour-piercing capabilities of large shaped-charge warheads with the high accuracy at long range of guided missiles. The U.S. M60A2 and the U.S.-West German MBT-70 were armed with 152-millimetre gun/launchers firing Shillelagh guided antitank missiles, and the AMX-30 was armed experimentally with the 142-millimetre ACRA gun/launcher. But the high cost, unreliability, and slow rate of fire of the missiles, together with the appearance of APFSDS ammunition and greatly improved fire-control systems, led to abandonment of gun/launchers in the early 1970s.

The first major postwar advance in fire-control systems was the adoption of optical range finders, first on the M47 tank and then on the Leopard 1, the AMX-30, and other tanks. In the 1960s, optical range finders began to be replaced by laser range finders. In combination with electronic ballistic computers, these greatly increased the hit probability of tank guns. They became standard in all new tanks built from the early 1970s and were retrofitted in many of the earlier tanks.

Infrared night sights

Another major development was that of night sights, which enabled tanks to fight in the dark as well as in daylight. Originally of the active infrared type, they were first adopted on a large scale on Soviet tanks. Other tanks were fitted from the 1960s with image-intensifier sights and from the 1970s with thermal imaging sights. These latter were called passive because, unlike active infrared systems, they did not emit energy and were not detectable.

After World War II an increasing number of tanks were fitted with stabilized gun controls to enable them to fire

more accurately on the move (*i.e.,* to keep their gun barrels at a constant angle of elevation even while the tank was riding over bumps or depressions). At first some tanks, such as the T-54, had their guns stabilized only in elevation, but the Centurion already had stabilization in traverse as well as elevation, and this became standard beginning in the 1970s. Afterward, tanks were also provided with independently stabilized gunners', as well as commanders', sights, the better to engage targets on the move.

Armour. Until the 1960s, tank armour consisted of homogeneous steel plates or castings. The thickness of this armour varied from eight millimetres on early tanks to 250 millimetres at the front of the German Jagdtiger of 1945. After World War II, opinions differed about the value of armour protection. Tanks such as the Leopard 1 and AMX-30 had relatively thin armour for the sake of light weight and greater mobility, which was considered to provide a greater chance of battlefield survival. Other tanks, such as the Chieftain, had heavier armour, up to 120 millimetres thick at the front, and the Arab-Israeli Wars of 1967 and 1973 demonstrated the continued value of heavy armour.

At the same time, new types of armour were developed that were much more effective than homogeneous steel, particularly against shaped-charge warheads. The new types were multilayered and incorporated ceramics or other nonmetallic materials as well as steel. The first was successfully developed in Britain under the name of Chobham armour. Armour of its kind was first adopted in the early 1970s in the M1 and Leopard 2; it then came into general use in place of simple steel armour.

By courtesy of (bottom) General Dynamics Land Systems Division; photographs, (top) Tass from Sovfoto, (centre) Lothar Kucharz—Ullstein

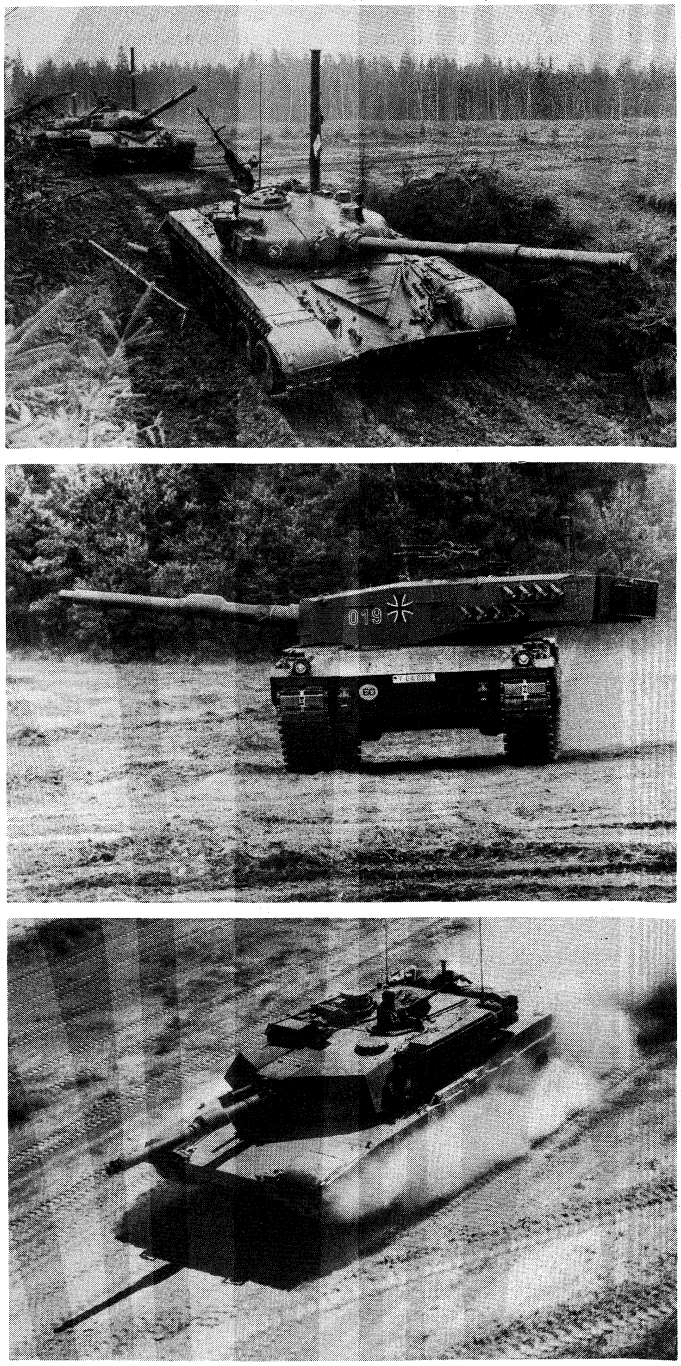


Figure 25: *Modern heavy battle tanks.*
(Top) Soviet T-72, with 125-millimetre gun. A snorkel is mounted for submerged fording. (Middle) West German Leopard 2, with 120-millimetre gun. Smoke dischargers are fitted onto the side of the turret for concealment. (Bottom) U.S. M1A1, with 120-millimetre gun adapted from the Leopard 2. The M1A1 is powered by a gas turbine engine.

Reactive armour

Fighting in Lebanon in 1982 saw the first use, on Israeli tanks, of explosive reactive armour, which consisted of a layer of explosive sandwiched between two relatively thin steel plates. Designed to explode outward and thus neutralize the explosive penetration of a shaped-charge warhead, reactive armour augmented any protection already provided by steel or composite armour.

The increased protection afforded to tanks inevitably increased their weight. Some tanks introduced during the 1950s and '60s, such as the T-54 and AMX-30, weighed only 36 tons, but the Chieftain already weighed 54 tons. Most tanks introduced during the 1980s, such as the M1 and the Leopard 2, also weighed more than 50 tons, and the Challenger weighed as much as 62 tons.

Mobility. In spite of the progressive increases in weight, tanks' speed and agility actually increased because they were provided with more powerful engines. After World War II, tank engines had an output of 500 to 800 horsepower, but, starting with the MBT-70, their output increased to 1,500 horsepower. Engines of this power were installed in the M1 and the Leopard 2, giving them power-to-weight ratios of more than 20 horsepower per ton.

Most tank engines of the immediate postwar years had 12 cylinders in a V-configuration and at first were of the spark-ignition gasoline type. But Soviet tanks already had diesel engines, and from the 1960s almost all tanks were diesel-powered. This increased their range of operation because of the greater thermal efficiency of the diesels, and it reduced the risk of catastrophic fires that could erupt if the armour was perforated by enemy weapons.

The development of gas turbines led in the 1960s to the use of one, in combination with a diesel engine, in the Swedish S-tank. After that, a 1,500-horsepower gas turbine was adopted to power by itself the M1 and M1A1. A gas turbine also powered the Soviet T-80, introduced in the 1980s. All other new tanks of the 1980s continued to be powered by diesels because of their greater fuel economy.

Since the speed of tanks over rough ground depended not only on the power of their engines but also on the effectiveness of their suspensions, the latter developed considerably in the postwar era. Almost all tanks adopted independently located road wheels, sprung in most cases by transversely located torsion bars. Exceptions to this were the Centurion and Chieftain and the Merkava, which used coil springs. To improve their ride over rough ground still further, most tanks built during the 1980s were fitted with hydropneumatic instead of metallic spring units.

Configuration. The great majority of postwar tanks continued the traditional configuration of driver's station at the front of the hull, engine compartment at the rear, and rotating turret at the centre. The turret mounted the main armament and was occupied by the tank's commander, gunner, and loader. This configuration, introduced by the Vickers-Armstrong A.10 tank designed in 1934, became almost universal after World War II, but after 1960 it was abandoned in some cases in favour of novel configurations. One widely adopted configuration retained the turret but replaced the human loader by an automatic loading mechanism. The first examples of this were on the T-64 and T-72 tanks, whose guns were automatically loaded from a carousel-type magazine below the turret.

Automatic loading

Another tank with an unconventional configuration was the Merkava, which had its engine compartment at the front and the ammunition at the rear of the hull, where it was least likely to be hit by enemy fire. The Merkava also had a turret with a low frontal area, which reduced the target it presented to enemy weapons.

AMPHIBIOUS VEHICLES

The need to cross rivers and other water obstacles led to the development of amphibious tanks, beginning with the British Medium D, designed in 1918. Amphibious tanks built during 1920s and '30s sacrificed armament and armour to achieve low weight in relation to their bulk and, therefore, sufficient buoyancy to float. This confined their usefulness to reconnaissance. A late example of this was the Soviet PT-76, introduced during the 1950s.

After the middle of World War II many other tanks were made amphibious by collapsible flotation screens, which, when erected, allowed even relatively heavy tanks to float. This method was first used with M4 medium tanks during the D-Day landings in Normandy in 1944. Later, flotation screens were permanently installed on the Swedish S-tank, the U.S. M551 Sheridan, and the Scorpion light reconnaissance tank introduced in 1969 by the British army.

An alternative method to crossing rivers was submerged fording, first tried in 1940 with the British A.9 cruiser and the German Pz. III and IV. After World War II, provision for submerged fording was built into several tanks, including the T-54, T-62, and T-72, the Leopard, and the AMX-30.

An entirely different problem was posed by amphibious landings from the open seas; to solve it the U.S. Marine Corps developed the Landing Vehicle Tracked, or LVT. Originally built in 1941 as an unarmoured cargo carrier, the LVT quickly acquired armour. Two types evolved: an armoured amphibious personnel and cargo carrier, and a turreted amphibious gun-vehicle for close fire support during landing operations. Altogether 18,620 LVTs were built during World War II; these played a prominent role in the Pacific campaigns from Guadalcanal onward. After World War II, LVTs were successfully used in Korea, notably for the 1950 Inch'ŏn landing. Two new models were built between 1951 and 1957: an LVTP-5 amphibious carrier, capable of carrying as many as 37 men, and an LVTH-6 armed with a turret-mounted 105-millimetre howitzer. They were followed in the 1970s by the 22.8-ton LVTP-7, which incorporated several improvements, the most important being a boatlike hull with a stern instead of bow loading ramp and two water-jet propulsion units that greatly improved its performance in comparison with that of the earlier LVTs (which were propelled in water as well as on land by means of their tracks). At the same time the LVTP-7 retained the seagoing qualities of the earlier LVTs, which could negotiate rough seas and Pacific surf in contrast to other amphibious vehicles intended primarily for crossing inland water obstacles. The use of water-jet propulsion units in the LVTP-7, however, had been preceded by their use in several amphibious reconnaissance vehicles, including the Soviet PT-76.

The Landing Vehicle Tracked

SELF-PROPELLED GUNS

Self-propelled guns are field artillery, antitank guns, or antiaircraft guns mounted on and fired from tracked or wheeled vehicles. The vehicles are armoured to varying degrees.

The development of self-propelled guns continued after World War II, except for self-propelled antitank guns. These became superfluous when it was recognized that all tanks needed to be sufficiently well-armed to fight enemy tanks. Turretless assault guns, much favoured during World War II by the German and Soviet armies, also virtually disappeared.

The general trend among the remaining self-propelled guns was either toward lightly but completely armoured models with guns mounted in turrets capable of all-around traverse, like most tanks, or toward partially armoured models that were essentially tracked chassis with guns mounted on top. Examples of the completely armoured self-propelled guns were the U.S.-built 155-millimetre M109 self-propelled howitzer (used not only by the U.S. Army but also by several others), the French 155-millimetre GCT, and the Soviet 122-millimetre SO-122 and 152-millimetre SO-152. Partially armoured models were represented by the U.S.-built M107 175-millimetre gun and M110 8-inch howitzer and by the Soviet 203-millimetre SO-203.

ARMOURED PERSONNEL CARRIERS

After World War II, armoured personnel carriers became the next most important armoured vehicle after battle tanks. Though a few experimental models were built in Britain at the end of World War I, development of armoured carriers did not really begin until they were adopted for the panzer division infantry at the beginning of World War II. Germany's example was quickly followed by the United States, which, by the end of the war, had produced 41,000 carriers. Both the German and U.S. carriers of World War II were of the half-track type and provided only light protection; nevertheless, they represented a major advance on the earlier method of transporting infantry into battle in unarmoured trucks. Moreover, the panzer grenadiers used them effectively as combat vehicles and fought from them on the move, thus greatly increasing the mobility of infantry on the battlefield.

In the postwar era the U.S. Army led in developing fully tracked carriers with all-around armour protection. The first postwar carrier was the large M44. This was followed in 1952 by the M75, which had a similar box body but carried 12 instead of 27 men. A few M75s were used successfully during the Korean War, and it became the first tracked armoured carrier to be used in large numbers.

In 1955 the M75 began to be replaced by the M59, which was similar in appearance but was less expensive and could swim across calm inland waters. In 1960 came the M113, which had a lower silhouette and was considerably lighter, owing partly to the use of aluminum armour (see Figure 26). The M113 was in fact the first aluminum-armoured vehicle to be put into large-scale production. After its appearance, several other armoured carriers, light tanks, and self-propelled guns were built with aluminum armour. Within 30 years more than 76,000 M113 carriers and their derivatives had been produced, making them the most numerous armoured vehicles outside the Soviet bloc. M113 carriers were used extensively in the Vietnam War, often as combat vehicles, although they were not designed for this and were at a disadvantage in spite of the addition of roof-mounted machine guns with shields.

The U.S. M113 armoured personnel carrier

Department of Defense photo

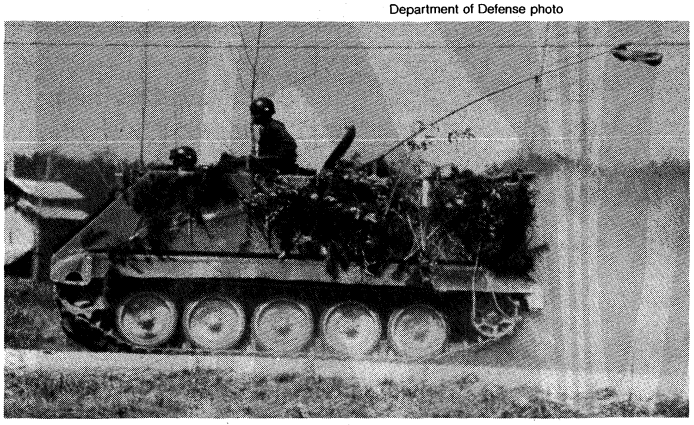


Figure 26: U.S. M113 armoured personnel carrier, capable of carrying 11 infantrymen besides its crew of two.

The first attempt to produce a tracked armoured carrier from which the infantry could fight to some extent was represented by the French AMX-VTT of 1958. A further step in this direction was taken by the West German army with the HS-30, which had a turret with a 20-millimetre cannon. The most significant advance was represented by the Marder, which was produced for the panzer grenadiers of the West German army during the 1970s, and the BMP, mass-produced for the Soviet armoured infantry from the mid-1960s. The Marder weighed 29.2 tons and

carried nine men in addition to mounting a turret with a 20-millimetre cannon. The BMP-1 weighed 13.5 tons and could carry up to 11 men. It was armed with a turret-mounted, low-performance 73-millimetre gun, while the BMP-2 was armed with a high-velocity 30-millimetre cannon; both versions carried externally mounted antitank guided missiles.

Another tracked armoured infantry vehicle was the U.S. M2 Bradley Infantry Fighting Vehicle, introduced in the 1980s. This 10-man vehicle weighed 22.6 tons and had a two-man turret with a 25-millimetre cannon and a TOW antitank missile launcher. Its British equivalent was the Warrior Mechanized Combat Vehicle, produced since 1986. This was also a 10-man vehicle of 24.5 tons with a 30-millimetre cannon in a two-man turret.

In addition to tracked armoured carriers or infantry fighting vehicles, which were intended to cooperate closely with tanks, most armies also developed wheeled armoured carriers for more general use. Examples included the VAB of the French army and the BTR-60, -70, and -80 of the Soviet army. (R.M.O.)

Naval ships and craft

Naval ships are the chief instruments by which a nation extends its military power onto the seas. Warships protect the movement over water of one's own military forces to coastal areas where they may be landed and used against enemy forces; warships protect one's own merchant shipping against enemy attack; they prevent the enemy from using the sea to transport his own military forces; and they attack the enemy's merchant shipping. Naval ships are also used in blockade—*i.e.,* in attempts to prevent an enemy from importing by sea the commodities necessary for his prosecution of the war. In order to accomplish these objectives, naval ships have been designed from earliest times to be faster and sturdier than merchant ships and to be capable of carrying offensive weapons.

In the modern era the word craft has come to denote small surface vessels that operate usually in coastal waters.

This section traces the development of the major surface warships and craft from their beginnings to the present day. Submarines, which operate below the surface, are discussed separately below. Detailed discussion of weapons used by warships can be found in other sections of this article. For early naval cannon, see *The gunpowder revolution;* for guided antiship and antiaircraft missiles, see *Tactical guided missiles;* for naval airplanes, jets, and helicopters, see *Military aircraft.*

THE AGE OF OAR AND RAM

The first craft fitted purposely to make war were conversions of the dugouts, inflatable bladders, papyrus rafts, or hide boats used in everyday transport. It is probable that the conversion at first consisted simply of a concentration of weapons in the hands of a raiding party. In time conversions added offensive and defensive powers to the craft itself. As vessels became more seaworthy and more numerous, warships designed as such developed both as marauders and as defenses against marauders. The first craft designed and built especially for combat may have sailed in the fleets of Crete and Egypt 5,000 years ago.

Egypt. The first recorded appearance of warships is on the Nile River, where Egypt's history has centred since antiquity. These boats were built of bundles of reeds lashed together to form a narrow, sharp-ended hull and coated with pitch, and they were hardly suited for tempestuous seas. By 3000 BC larger wooden seagoing versions of the reed craft sailed for distant cruising, trade, and conquest.

Egyptian wooden ships had both oars and sails, being fitted with a bipod (inverted V) mast and a single, large, square sail. The whole mast could be lowered when under oars. Large Egyptian ships had more than 20 oars to a side, with two or more steering oars. The war galley was built to the same pattern but was of stouter construction. Modifications that could be easily incorporated in a merchant ship's hull under construction included elevated decks fore and aft for archers and spearmen, planks fitted to the gunwales to protect the rowers, and a small fighting top high on the mast to accommodate several archers. Some galleys had a projecting ram, well above the waterline, which may have been designed to crash through the gunwale of a foe, ride up on deck, and swamp or capsize him.

Merchant galleys modified for war

Crete. By about 2000 BC Crete had evolved into a naval power exercising effective control of the sea in the eastern Mediterranean. Little record exists of Minoan seapower, yet these maritime people may have been the first to build a warship designed as such from the keel up, rather than as a modification of a merchant ship. Thus it was probably the Minoans who began to differentiate between war craft and merchantmen and between the rowing galley and sailing vessel.

Sometime in the 2nd millennium BC the commodious merchantman evolved as a beamy "round ship" powered by sails and emphasizing cargo capacity at the expense of speed. By contrast, the fast fighting "long ship" was narrower, faster, and more agile than the tubby cargo ship. Developing as both predator on and protector of maritime trade and coastal cities, it hoisted its sails for cruising but depended on oars when in action.

The Cretan warship had a single mast and a single bank of oars. The sharply pointed or "beaked" bows suggest an emphasis on the tactical use of the ram.

Phoenicia. Beginning about 1100 BC, the Phoenicians dominated the eastern Mediterranean for about three centuries. Information about Phoenician ships is fragmentary, but they appear to have been built primarily for trading, with a capacity to fight effectively if necessary.

Phoenician trading ships were apparently galleys, mounting a single pole mast with a square sail and with steering oars to port and starboard. Their war galleys show a Cretan influence: low in the bow, high in the stern, and with a heavy pointed ram at or below the waterline. Oars could be carried in a staggered, two-bank arrangement, allowing more oars to be mounted in a ship of a given length and increasing power and momentum. Because the ram was the principal weapon, the vessel's slender build and greater rowing power were important in providing more speed for the decisive shock of battle.

Greece. Unlike the Egyptians, for whom wood was scarce and costly, the Aegean peoples had an abundance of timber for shipbuilding. The earlier Greek warships were used more to carry attack personnel than as fighting vessels. No mention is made in the *Iliad,* for instance, of sea warfare. Even the pirates of the time were sea raiders seeking their booty ashore rather than in sea actions. The so-called long penteconter, mentioned by Herodotus, was employed in exploring, raiding, and communicating with outlying colonies. Light and fast, with 25 oars to a side, it played an important role in the early spread of Grecian influence throughout the Mediterranean. As the Greek maritime city-states sped the growth of commerce and thus the need for protection at sea, there evolved a galley built primarily for fighting. The first galleys, called uniremes (Latin: *remus,* "oar"), mounted their oars in a single bank and were undecked or only partially decked. They were fast and graceful with high, curving stem and stern. In Homeric times some carried an *embolon,* a beak or ram, which became standard in succeeding centuries.

Biremes and triremes. The bireme (a ship with two banks of oars), probably adopted from the Phoenicians, followed and became the leading warship of the 8th century BC. Greek biremes were probably about 80 feet (24 metres) long with a maximum beam around 10 feet. Within two or three generations the first triremes (ships with three vertically superimposed banks of oars) appeared. This type gradually took over as the primary warship, particularly after the Greeks' great sea victory at the Battle of Salamis (480 BC).

The first triremes

Like its predecessors, the trireme mounted a single mast with a broad, rectangular sail that could be furled. The mast was lowered and stowed when rowing into the wind or in battle. Built on an entirely different system from the Egyptians—with keel, frames, and planking—these were truly seagoing warships.

After Salamis, the trireme continued as the backbone of the Greek fleet, with the ram continuing as its primary

weapon. Its keel, like those of its predecessors, formed the principal-strength member, running the length of the ship and curving upward at each end. The ram, usually shod in bronze, formed a forward prolongation that gained effectiveness from the heavy keel back of it. Additional longitudinal strength came from a storming bridge, a gangway along the centreline from bow to stern along which the crew raced to board when a foe was rammed. Gradually, with ships becoming steadily heavier, boarding assumed greater importance and the ram lost some of its importance.

A trireme of the 5th century BC may have had a length of about 125 feet, a beam of 20 feet, and a draft of three feet. Manned by about 200 officers, seamen, and oarsmen (perhaps 85 on a side), with a small band of heavily armed *epibatai* (marines), under oars it could reach seven knots (seven nautical miles per hour; one knot equals 1.15 statute miles per hour or 1.85 kilometres per hour). Extremely light and highly maneuverable, the classical trireme represented the most concentrated application of human muscle power to military purposes ever devised. The oarsmen sat on three levels, which were slightly staggered laterally and fore and aft to achieve the maximum number of oarsmen for the size of the hull. In rowing, the oarsmen slid back and forth on leather cushions strapped to their buttocks; this enabled full use of the powerful muscles of the thighs and abdomen (see Figure 27).

With only scant room for provisions, such warships could not remain long at sea, and a voyage usually consisted of short hops from island to island or headland to headland. Even the largest triremes put into shore and beached for the night, resuming the passage in the morning, weather permitting. Light construction and little endurance made short distances between bases essential and frequent refits imperative.

Later developments. The trireme reached its peak development in Athens. By the middle of the 4th century BC, Athenians employed quadriremes (four-bank seating), with quinqueremes appearing soon thereafter. In the late 4th and early 3rd century BC an arms race developed in the eastern Mediterranean, producing even larger multibanked ships. Macedonia's rulers built 18-banked craft requiring crews of 1,800 men. Ptolemaic Egypt capped them with 20s and 30s. Ptolemy III even laid down a 40 (tesseraconter) with a design length of over 400 feet and calling for a crew of 4,000 rowers. The vessel was never actually used. (The multiplicity of "banks," once a puzzle to historians, signifies the number of rowers on each oar or row of oars rather than an almost unimaginable vertical piling-up of banks.)

Naval weapons of the galley era

This same arms race brought other changes of significance. Until the late 4th century BC, maneuver, marines, and the ram constituted a warship's offensive strength, and archers provided close-in fire. Demetrius I Poliorcetes of Macedonia is credited with introducing heavy missile weapons on ships at the end of the century, starting a trend that has continued to the present day. Demetrius' ships mounted crossbowlike catapults, for hurling heavy darts, and stone-throwing machines of the type the Romans later called ballistae. From this time on, large warships carried these weapons, enabling them to engage a foe at standoff ranges, though ramming and boarding also continued. Temporary wooden turrets—forecastles and sterncastles—were similarly fitted to provide elevated platforms for archers and slingers.

Following the fragmentation of the brief empire of Alexander the Great, sea power developed elsewhere. The city-state of Rhodes built a small but competent fleet to protect its vital shipping. Meanwhile to the west, Carthage, a state with ancient maritime origins, rose to prominence on the north coast of Africa and by about 300 BC had become the foremost Mediterranean naval power. Carthage's navy consisted probably of the same ram-galley types developed by its ancestral Phoenicians and by the Greeks.

Rome. Coincidentally, across the sea to the north the city-state of Rome expanded to include most of the southern Italian peninsula, with its extensive seacoast and maritime heritage. Rome's growth southward collided with Carthage's ambitions in Sicily, leading to the First Punic War, which began in 264 BC. Unlike their seafaring opponents, the Romans were not a naval power. When in the fourth year of the war Carthage sent a fleet against Sicily, Rome realized its fatal disadvantage and moved to remedy it. The Greek city-states it had conquered had long seagoing experience. Employing their shipbuilders and learning also from the foe, Rome built a fleet of triremes and quinqueremes, the latter patterned after a Carthaginian warship that ran aground in Roman territory early in the war.

Not content with copying the enemy's tactics, the Romans took land warfare to sea and forced the Carthaginians to fight on Roman terms. Each Roman galley had fitted in the bow a hinged gangplank with a grappling spike or hook (the *corvus*) in the forward end, thus providing a boarding ramp. They added to the crews many more marines than warships usually carried.

The gangplank and grappling hook

The Phoenicians and Greeks had emphasized ramming, with boarding as a secondary tactic. A Roman captain rammed and then dropped the gangplank. Ram and *corvus* locked the galleys together, and the Roman marines boarded, overwhelming the opponent. The Roman fleet had extraordinary success in the great naval Battle of Mylae off northeast Sicily, destroying or capturing 44 ships and 10,000 men. After other victories, and some defeats, by the end of the First Punic War, 241 BC, Rome had become the leading sea power.

As the Roman navy evolved, so did its warships. Though pictorial evidence is ambiguous, it seems clear that the gangplank and *corvus* disappeared as the Romans gained experience in sea warfare. Later Roman warships appear to have been conventional fully decked ram-galleys mounting one or two wooden turrets (probably dismountable) for archers. To the single mast with rectangular sail was added a bowsprit carrying a small sail, the artemon, Falces, or long spars tipped with blades, were used by Julius Caesar's fleet against the sailing vessels of the Veneti of northwestern Gaul to cut their rigging and immobilize them. Catapults and ballistae served as mechanical artillery, and it was under their fire that Caesar's legions landed in England.

Early Roman warships were all large; to escort merchantmen and combat pirates Rome found need for a lighter type, the liburnian. Probably developed by the pirates themselves, this was originally a light, fast unireme to

Redrawn from J. Morrison, *The Ship: Long Ships and Round Ships* (1980), The National Maritime Museum, London, after J. Morrison and R.T. Williams, *Greek Oared Ships* (1966), Cambridge University Press

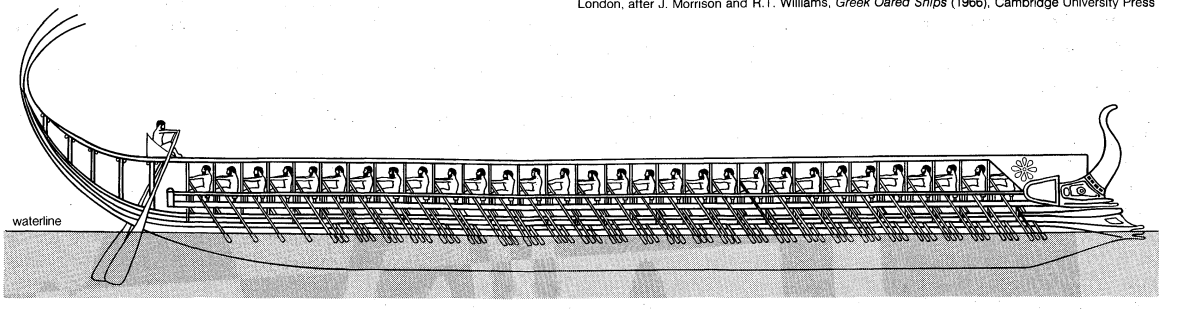


Figure 27: *Greek trireme of the 5th century BC.*
Powered by as many as 85 oarsmen on each side, the trireme was capable of seven knots and had a beaked prow for ramming enemy vessels.

which the Romans added a second bank of oars. In the Battle of Actium, 31 BC, Octavian's skilled fleet commander, Agrippa, used his liburnians to good effect. Although polyremes continued to be built after Actium, the liburnian became the predominant Roman man-of-war.

The Byzantine Empire. With the breakup of the western Roman Empire, naval organization and activity in the west decayed. In the eastern Roman Empire, however, the need for sea power was well appreciated. During the 11 centuries that the Roman Empire centred on Constantinople, the Byzantine rulers maintained a highly organized fleet. Their original type of warship was the liburnian, called in Greek the *dromōn;* it was built in several different sizes, the heavier designed to bear the weight of battle and the lighter single-bank dromons serving as cruisers and scouts.

Throughout the eastern Roman Empire's existence warships changed little except in rig and armament. An average large dromon measured up to 150 feet in length, with 100 oars and one or two fighting towers for marines. At some point early in the Christian era, the lateen sail, three-cornered and suspended from an angled yard, probably adopted from the Arabs, came into general use. Eastern warships had two or three masts. In a departure from classical customs, these were left in place in battle. Contemporary pictures show rams above the waterline.

Missile-launching weapons grew in size, some hurling projectiles as large as 1,000 pounds (450 kilograms) up to 750 yards (685 metres). Greek fire, a combustible material for setting fire to enemy ships, was invented in the 7th century or earlier. The various compounds passing under the name used a blend of some of the following: pitch, oil, charcoal, sulfur, phosphorus, and salt. As the composition of Greek fire was improved, tubes shaped into the mouths of savage monsters were placed in the bows of war galleys and the flaming substance, which water merely spread, was hurled on the enemy. Greek fire was an important factor in terrifying and repelling the Muslim fleet in sieges of Constantinople from the early 8th century on.

Greek fire

Viking vessels. By the beginning of the Viking period, about AD 800, the early and primitive Scandinavian craft had evolved into the well-known Viking ship, a sturdy, double-ended, clinker-built (*i.e.,* with overlapping planks) galley put together with iron nails and caulked with tarred rope. It had a mast and square sail, which was lowered in battle; high bow and stern, with removable dragon heads; and a single side rudder on the starboard (steerboard) quarter.

Viking vessels were essentially large open boats. Like the Homeric Greeks, the Vikings at first made no distinction between war and cargo ships, the same vessel serving either purpose as the occasion demanded. Later, however, they built larger ships specifically designed for war. By AD 1000 they sailed three categories of these: those with fewer than 20 thwarts (40 rowers); those with up to 30; and the "great ships" with more than 30, which might be considered the battleship of the time. Expensive and unwieldy, though formidable in battle, the great ships were never numerous. The middle group, maneuverable and fast, proved most valuable.

Viking "long ships" played an important role in exploration (reaching Greenland and America before Columbus), in the consolidation of kingdoms in Scandinavia, and in far-ranging raids and conquests. In them the Norsemen invaded the British Isles and established themselves in Normandy, whence their descendants under William I the Conqueror crossed the Channel in 1066.

THE AGE OF GUN AND SAIL

To about the end of the 13th century, the typical ship in northern European waters remained a clinker-built, single-masted, square-rigged descendant of the long ship. In that century, and even more in the 14th, changes began that would bring an end to the long dominance of the oar in battle. In about AD 1200 came one of the great steps in the history of sail: the introduction, probably in the Netherlands, of the stern rudder. This rudder, along with the deep-draft hull, the bowsprit and, in time, additional masts, transformed the long ship into the true sailing ship, which could beat into the wind as well as sail with it.

The rudder and the true sailing ship

Until the 15th century, northern ships probably continued to have single masts, though in the Mediterranean a two-mast rig carrying lateen (fore-and-aft) sails had existed for some time. Then change came rapidly in the north, spurred on by Henry V of England's construction of large and strongly built warships for his cross-Channel French campaigns. The remains of one of these, the *Grâce Dieu,* reflected the clinker-built construction of the Viking long ship, but they had a keel to beam ratio of about 2.5:1 and now carried a second mast.

Some historians believe that the *Grâce Dieu* carried a third mast. At any rate, in a few decades ships had three and, by the end of the century, large vessels mounted four masts carrying eight or more sails. A three-master carried a large sail on each mast and in addition a main topsail and the spritsail under the bowsprit—the rig, in fact, of Christopher Columbus' *Santa Maria* in 1492. Ships, no longer dependent on fair winds, could and did range the world.

The beamier round sailing ship used for commerce also became the warship when the need arose. In times of war, temporary wooden castles were added at the bow and stern to provide bulwarked platforms for archers and slingers. A complement of men-at-arms embarked, in addition to the ship's seamen. Tactics were usually simple and straightforward, opposing fleets closing and attempting to beat down each other's archers before grappling and boarding. At war's end, off came the castles, and the ship went back to trading.

The trading vessel that could be promptly adapted to war did not, however, fulfill the need of the European nations for navies. The coming of gunpowder and the period of world exploration brought changes that were to cause the sailing man-of-war to become more and more distinct from the merchantman.

Gun-armed warships. The employment of guns afloat, bringing a slow but progressive revolution in warship construction and naval tactics, had its first small beginnings by the 14th century. The first guns used at sea, undoubtedly hand weapons, were probably in Mediterranean galleys in the 13th or early 14th century. Such weapons played a minor role. In fact, in the numerous sea battles of the Greeks, Genoese, Moors, Turks, and Venetians during this period there is no mention of guns. But by the middle of the 14th century, the English, French, Spanish, and other navies mounted guns. Most were relatively small swivel pieces or breech-loading deck guns located in the castles fore and aft, but heavier guns were added later. The Mediterranean galleys of Venice, Turkey, and Spain at first simply mounted a heavy gun rigidly in a timber bed that was fixed to fire the gun forward over the bow. By the late 15th century these rigid mounts gave way to sliding mounts for the main centreline bow gun, as the pieces were called. Though some of these pieces were quite large, the light structure of a galley meant that there was only one large gun per vessel.

The first naval guns

European guns were originally built up of wrought-iron bars welded together to form a tube, then banded with a thick iron hoop. Initially, they were breechloaders with an open trough at the rear of the barrel through which the ball was loaded and a cylindrical chamber, filled with powder, inserted and wedged tight. They were replaced after 1500 by brass muzzle-loaders, cast in one piece. Some of these muzzle-loaders attained great size for their day; by the mid-16th century even some 60-pounders (firing 60-pound solid shot) were mounted in the largest ships. In this century also, increasing knowledge of iron metallurgy led to the production of cast-iron cannon that slowly replaced the brass guns in ships, though brass remained predominant for the lighter calibre well into the 19th century.

The Portuguese and Spanish, and then the French, seem to have been the first to cover transoceanic distances with cannon-armed warships. Vasco da Gama reached Calicut in India in 1498 with a squadron of cannon-armed carracks, and the Portuguese gained a number of signal victories over their Muslim opponents in the East in the early years of the 16th century using standoff artillery tactics that their foes could not match. The Spanish were pa-

trolling the waters of the Caribbean in ships well-provided with wrought-iron breech-loading cannon by the 1520s or '30s, if not before, and heavily armed French raiders were not far behind.

Gunports and heavy guns

Henry VII of England created the first true oceangoing battle fleet. The "king's ships" carried many guns, but most of these weapons were small breechloaders. Following him, Henry VIII initiated gunports in English warships, a development that was to have a far-reaching effect on man-of-war design. Neither stability nor structural strength favoured heavy guns in the high castles built upon the deck, so that Henry's introduction of gunports, at first low in the waist of the ship and afterward along the full broadside, made possible the true heavy-gun warship. The cutting of gunports in the hull must also have been a factor in causing the northern nations to shift from clinker-built ships to caravels with flush-fitting planks, a change that took place in the early 1500s.

The armament of an English man-of-war of the early 16th century consisted of four or five short-barreled cannon, or curtals, a similar number of demicannon, and culverins. The average cannon, a short-range gun, hurled an iron ball of about 50 pounds, and the demicannon one of 32 pounds. The culverin, a longer and stronger gun, fired a smaller shot over a longer range and was likely to be more accurate at other than point-blank range. Supplementing these standoff "ship killers," in descending size of ball fired (down to only several ounces), were the smaller demiculverin, saker (quarter culverin), falcon (half saker), falconet, and robinet.

A great warship of the 16th century mounted a total of large and small pieces approximating the numbers mounted in battleships of World War II. For its original complement in 1514, Henry VIII's best-known warship, the *Henry Grâce à Dieu,* had 186 guns. Most of these were small, but they also included a number of iron "great guns."

As the 17th century advanced, guns and gunpowder improved. Gun carriages were given heavy wooden sides called brackets, which had sockets for the gun trunnions and were joined by similar flat timbers called transoms. The carriage was supported on wooden trucks and hauled out after recoil by heavy tackle.

From oar to sail. *The galleass.* The coming of mighty men-of-war did not mean the immediate end of oared warships. In fact, some types of galleys and oared gunboats continued to serve well into the 19th century. Indeed, the Battle of Lepanto (1571), in which a combined European fleet defeated the Turkish fleet, differed little from traditional galley warfare with two exceptions. First, the scale of the action was very large, with more than 200 cannon-armed galleys on each side. Each of those galleys was propelled by 50 to 200 oarsmen and carried at least 50 additional men to fight and to man the guns and sails. Second, the European line of battle included six Venetian galleasses, a compromise type developed in the transition from oar to sail. These huge vessels, which depended on sail as well as oar, bristled with guns, including heavy ones in broadside. Although cumbersome to maneuver, their concentrated fire contributed importantly to victory.

Guns and oars at the Battle of Lepanto

Galleasses outside the Mediterranean differed somewhat from Venice's in that they were basically full-rigged sailing ships carrying broadsides of heavy guns and a bank of auxiliary oars for mobility. The hybrid existed only in small numbers and soon passed out of fashion to the north.

The galleon. The "great ships" of Henry VII and Henry VIII were carracks: starting basically with the lines of beamy, seaworthy merchant ships, designers had added stronger timbers, masts, sailpower, broadside guns, and high-built forecastles and aftercastles. In the galleon, the successor to the carrack, the general principles of design of sailing men-of-war were established, and they ruled, without fundamental change, for three centuries. The galleon retained certain characteristics of the galley, such as its slender shape, and in fact it had a greater length-to-beam ratio than the carrack. But the carrack's high-built forecastle, which tended to catch the wind and thus make the ship unmaneuverable, was eliminated from the galleon's design. The resulting ship was much more seaworthy. Like carracks, the larger galleons might carry a single mizzenmast or two relatively small masts, the second being called the bonaventure.

In the longer, leaner galleon, the number of heavy guns was increased until they ran the full length of the ship's broadside in one or two tiers (and later three).

The galleon came into favour in northern Europe during the middle of the 16th century. The far-ranging experience of mariners and improved construction techniques led to great fighting ships that were both lower in the water and more seaworthy than their predecessors. The sides now sloped inward from the lowest gun deck up to the weather deck. This "tumble home" helped concentrate the weight of the large broadside guns toward the centreline, improving the ship's stability.

By this time it had become normal for warships to mount powerful broadsides of 28 or more ship-smashing guns, a much heavier armament in proportion to their size than their predecessors. For their handy, maneuverable ships, the British had relatively large cannon carried in broadsides. Thus they were designed for off-fighting, permitting the English fleet to get the most out of its ships' superior maneuvering qualities. When the Spanish Armada arrived in 1588, the British sought and fought a sea battle with ship-killing guns, rather than the conventional fleet engagement of the past that concentrated on ramming, boarding, and killing men in hand-to-hand combat. With superior ability and long-range culverins, the English ships punished the invading fleet outside the effective range of the heavy but shorter-range cannon the Spanish favoured. This historic running battle of July 1588 closed one era and opened a greater one of big-gun sailing navies.

Guns and sails against the Spanish Armada

Ship of the line. The late Elizabethan galleon that began the true fighting ship of the line reached its culmination in England's *Prince Royal* of 1610 and the larger *Sovereign of the Seas* of 1637, along with similar great ships in other European navies. These two English ships mounted broadside guns on three decks; the *Sovereign of the Seas,* the most formidable ship afloat of its time, carried 100 guns. In this mobile fortress displacing approximately 1,500 tons, there was some reduction of height; the bonaventure mizzen disappeared, leaving the standard three masts that capital ships thereafter carried (see Figure 28).

Soon ships began to be standardized into different categories. James I organized his ships into four ranks, and, by the mid-17th century, six "rates" existed as a general concept, though not yet a system. The number of guns a ship carried determined its rate, with a first-rater mounting 100 guns and a sixth-rater 18. An important improvement came in the standardization of batteries in the higher rates

The National Maritime Museum, London

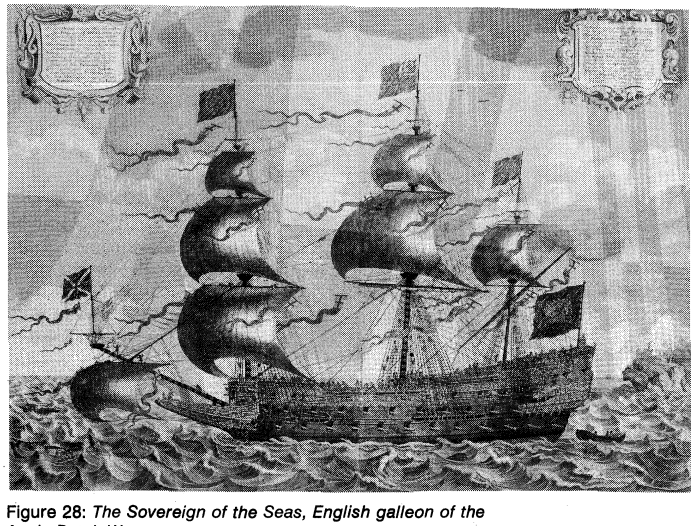


Figure 28: *The Sovereign of the Seas, English galleon of the Anglo-Dutch Wars.*
Launched in 1637, this was the largest warship of its time and the first to carry 100 guns. The prominent beak at its bow soon went out of fashion, but its three gun decks and low sterncastle and forecastle set the pattern for ships of the line for the rest of the sailing era. Contemporary engraving by J. Jayne.

so that guns on the same deck were of the same weight and calibre rather than mixed, as originally in the *Sovereign of the Seas.* Near the end of the century, guns began to be described by their weight and calibre, with the 32-pounder long gun favoured as the standard lower-deck weapon for British warships.

The frequent hard-fought sea battles of the 17th century, particularly in the Anglo-Dutch wars, led to the column formation of heavy warships called line ahead. In the line formation, each warship followed in the wake of the ship ahead so that every ship in the line had a clear field of fire for a broadside discharge of its guns. The adoption of line-ahead tactics made it necessary to standardize the battle line, which had consisted of ships of widely varying strength. Now only the more powerful warships were considered suitable "to lie in the line of battle." Hence the origin by the 1700s of the term line-of-battle ship, or the ship of the line, and, in the second half of the 19th century, the derived term battleship—ships that could hit the hardest and endure the most punishment.

The great two- and three-deckers

Some first-raters were built to carry as many as 136 guns, but, because the biggest ships were often cumbersome, relatively few were built. The handier 74-gun third-rater proved particularly successful, combining sufficient hitting power with better speed and maneuverability. Most of the ships of the line of the late 18th and early 19th centuries were 74s. One of these might be approximately 175 feet long with two full gun decks, the lower mounting the heaviest guns, by the Napoleonic Wars usually 32-pounders. The upper gun deck customarily carried 24-pounders, while the forecastle and quarterdeck mounted lighter guns. The bigger ships were similar but had three covered gun decks instead of two. Viscount Nelson's *Victory,* launched in 1765 and preserved in dry dock as it was at Trafalgar in 1805, is a classic example of this powerful type.

Warships gradually improved in design through the 17th and 18th centuries. New types of sails, providing more canvas and more versatile combinations for varying weather conditions, such as staysails and the jib sail, came into use in the 17th century. Soon thereafter the steering wheel replaced the old whip staff, or tiller.

Frigates and smaller vessels. Ships of the line, first to fourth rates, had strong, fast frigates as consorts. This ancestor of the modern cruiser evolved during the mid-18th century for scouting, patrol, and escort, as well as for attacking enemy merchantmen. The frigate carried its main battery on a single gun deck, with other guns on forecastle and quarterdeck. Like ships of the line, they varied in size and armament, ranging from about 24 guns in early small frigates to as many as 56 in some of the last. Two classic examples, still preserved, are the U.S. Navy's *Constitution,* with 44 guns, and *Constellation,* with 38.

Smaller vessels aided frigates in their blockade, escort, commerce raiding, and other duties. The single-masted cutter served as scout and coastal patrol craft. Brig and schooner-rigged types, generally called sloops of war, by the time of the American Revolution grew into the three-masted, square-rigged "ship sloop." Called a corvette on the Continent, the fast ship sloop complemented frigates on the fringes of the fleet. Smaller sloops, schooners, brigs, and luggers were widely used for special service. Fleets also needed ordnance and supply ships and other auxiliaries; these were usually merchantmen taken into service in war emergency. Converted merchantmen, such as John Paul Jones's *Bonhomme Richard,* often played combat roles. Fleets also had various special types, such as fire ships and bomb ketches. The latter, with two large mortars hurling bombs of about 200 pounds, were developed by France in the late 1600s and were used with devastating effect against Barbary pirate ports.

THE AGE OF STEAM AND IRON

As the Industrial Revolution unfolded in the 19th century, the age of wooden-hulled sailing ships gave way to that of steam-powered iron ships. Phenomenal changes took place in nearly every aspect of warship design, operation, and tactics. These changes ended the reign of the majestic ship of the line by the mid-1800s, but another half century elapsed before it was clear what form its replacement as the backbone of fleets would take.

Toward the ironclad. The change from wood to iron came slowly, in considerable part because the introduction of steam power required new techniques and experience in shipbuilding. The general use of iron for warships awaited the full realization of the value of the shell gun and the resulting need for armour, which were first demonstrated in the employment of armoured batteries in the Crimean War and in the battle between the *Monitor* and *Merrimack* in the American Civil War. The changes may be summarized under three headings: propulsion, armament, and armour.

Propulsion. Steam for propulsion of vessels was tried with varying success in several countries during the late 18th century. Engines and supporting machinery were at first not adequate for this fundamental advance in ship capability, but useful steam craft appeared in the early 1800s, suitable for operation on inland and coastal waterways. The earliest steam warship was the *Demologos* of the U.S. Navy (renamed *Fulton* after its designer, Robert Fulton). Built in the War of 1812, this well-gunned, double-hulled, low-powered ship cruised briefly in the New York Harbor area before the war ended and later was destroyed by an accidental fire.

First steam warship

The earliest steam warships in action were small paddle wheelers used by British and American navies against pirates and other weak foes. As engines gradually improved, navies experimented with them in standard warships, first as auxiliaries to sail, which was then essential for endurance. The paddle wheels were particularly vulnerable to enemy fire. In 1843, through the drive of Captain Robert Field Stockton of the U.S. Navy and the inventive skill of John Ericsson, a Swede whom Stockton brought to America, the United States launched the world's first screw-driven steam man-of-war, USS *Princeton,* a large 10-gun sloop.

The screw propeller was an old idea going back to Archimedes, but, with Stockton's assistance, Ericsson had made it effective for large warships, as Sir Francis Pettit Smith was doing at about the same time in England for large merchantmen. By the mid-1840s, boilers, engines, and machinery had improved to the point that thereafter practically all of the new warships had steam propulsion, though they also still carried sails.

Among the advances of this period were two other milestones. In 1834 Samuel Hall of England patented a type of steam condenser that made it possible to use fresh instead of corrosive salt water for boilers. In 1824 James Peter Allaire of the United States invented the compound-expansion steam engine, in which the steam was used in a second cylinder at a lower pressure after it had done its work in the first. Eventually it was made practical by progress in metallurgy and engineering; in 1854 John Elder, shipbuilder on the River Clyde, installed a successful two-stage engine in the merchant steamer *Brandon.* The higher efficiency was of great importance for ocean-keeping navies.

Armament. The basic changes in armament that were to take place in the 19th and 20th centuries had begun in the 18th century. In the British navy steps to make possible heavier long-range guns began with the introduction of strong springs to take up the first shock of the gun's recoil after firing, aided by inclined-plane wedges behind the trucks to coax the gun forward into firing position after recoil. Flintlocks pulled by a lanyard, instead of match, fired the guns. Sights also improved. In the early 1800s navies began to employ mercury fulminate in percussion caps to initiate firing. Efficient percussion locks came into use within a few years.

From smoothbore to rifle

Smoothbore guns were still inaccurate, and successful efforts were made to bring back the rifled barrels, as well as the breech loading, of early guns, thus increasing their speed and accuracy of fire. The bore of a rifled gun barrel had spiral grooves cut into it that caused a projectile fired from it to spin in flight; if this projectile was shaped in the form of a cylinder with a cone-shaped forward tip, spin enabled it to fly through the air with its pointed end forward at all times. This improved aerodynamics gave

the shell a more accurate course of flight and a longer range. Because a projectile could not be rammed down the muzzle of a rifled barrel, the use of rifling had to await the design of an efficient breech-loading mechanism. In the 1840s, Italian and Austrian inventors brought out sliding-wedge breechblocks. Later the French developed an interrupted screw system, originally an American invention. A British firm produced a rifled breech-loading gun that the Royal Navy used until 1864, when a number of accidents brought a temporary reversion to muzzle-loaders. But defects were eventually remedied, and breech loading brought phenomenal increases in rates of fire.

French 6.5-inch (165-millimetre) cast-iron rifled guns in the Crimean War demonstrated superiority in range, destructive power, and accuracy. They helped impress all of the navies with the need for rifling. Slower-burning powder was also badly needed. Black powder had gradually been improved during 600 years of use in firearms, but it still retained its primary defect, too-rapid burning (and hence the creation of gas pressures so high that they could burst a gun barrel upon firing). The use of rapid-burning powder required keeping the size of the charge down (and therefore the range) to prevent the bursting of even the best guns. Just before the American Civil War the U.S. Army developed large, perforated, dense grains of black powder that burned more slowly and thus were a start toward the controlled burning ultimately achieved with smokeless powder.

Exploding shell

A development equal in importance to the rifling of naval guns was the replacement of solid iron cannonballs with large shells that exploded upon impact. Shell guns in warships' main batteries were preceded by bombs fired from mortars, small shell guns, and solid hot shot heated to cherry red. A principal architect in bringing big shell guns to sea was Henri-Joseph Paixhans, a general of French artillery. The first large shell guns from Paixhans' design, chambered howitzers firing a 62.5-pound shell (thicker-walled than bombs to penetrate before exploding) was tested in 1824 against a moored frigate with remarkable accuracy and incendiary effect.

The new guns began to come into use afloat in the 1830s, a French squadron firing them in the bombardment of Vera Cruz, Mex. The U.S. Navy began installation of the new guns, including 16 eight-inch shell guns in the three-decker *Pennsylvania,* along with 104 32-pounder solid-shot guns. The British made similar installations. There was good reason for navies to proceed cautiously, as the production of shell guns at first encountered many manufacturing problems. (Indeed, in a gala demonstration of the 12-inch shell guns on the USS *Princeton* for President John Tyler, one of the guns blew up, killing the secretary of the navy and several others.) In the event, improvements in metallurgy, gun construction, and fire control—along with the maneuverability of steam warships—at last led to the important extension of range that the big gun had promised from the beginning.

In 1853 the dramatic destruction of a weaker Turkish squadron by a Russian fleet in the harbour of Sinop of Turkey's Black Sea coast attracted world attention and increased interest in shell guns. England, the United States, and others built big steam frigates (as they were misleadingly called) with big shell guns. Their great striking power and maneuverability under steam made them the capital ships of the day, superseding the ship of the line for a brief time before the ironclads took over.

Larger guns, increased powder charges, and greater tube pressures were made possible by the replacement of cast iron by built-up wrought-iron guns (later, cast steel and, eventually, forged steel were used). Hoops were shrunk on over the powder chamber and breech end of the tube to give the strength required for the greater internal pressures sustained by these guns upon firing.

Armour. The use of larger guns with more penetrating power and explosive shells made armour plating imperative. Among early experiments were floating armoured batteries built for the Crimean War. Heavy wrought-iron plates over a thick wooden backing gave these flat-bottomed vessels outstanding protection as they carried large-shell guns close inshore.

Iron ship armour

Other developments followed swiftly. The British soon built the first iron-hulled floating batteries. The French followed with the *Gloire,* the first seagoing armoured warship, protected throughout her entire length by a wrought-iron belt of 4.3- to 4.7-inch armour backed by 26 inches of wood. Displacing 5,617 tons, she mounted 36 large shell guns and could steam at 13.5 knots; a three-masted sailing rig supplemented the engines. *Gloire* was the first of a series of ironclads laid down by Napoleon III; 13 similar ships soon followed, then two-decker armoured rams. Great Britain countered with the *Warrior,* the first iron-hulled, seagoing, armoured man-of-war. Much larger than the *Gloire,* she displaced 9,210 tons, mounted 28 seven-inch shell guns, had slightly lighter armour, carried sails, and was one knot faster.

These first ironclads were commissioned on the eve of the American Civil War, in which ironclads were destined to take a decisive part. The war itself produced several spectacular developments, including pioneer submarines, the first aircraft carriers (to handle balloons for observation), and the torpedo boat, one of several means the Confederates explored in trying to break the blockade. These little craft had weak steam engines and mounted a torpedo lashed to a spar projecting from the bow. Called Davids, they were weak but definite forerunners of the torpedo boat and the versatile destroyer.

Ironclad warships were crucial, perhaps decisive, in the North's victory over the South. Partial ironclads appeared early on the western rivers and spearheaded Union general Ulysses Grant's victories in 1862. River and coastal ironclads (ultimately, mostly monitors) dominated the war against the South in attacks from the sea and in decisive support of land operations from the Mississippi system to the Chesapeake Bay and James River. Most memorable of the combats was the duel between the *Monitor* and *Virginia* (better known as the *Merrimack*). When the Federal forces lost Norfolk Naval Shipyard in Portsmouth, Va., in April 1861, they burned several warships, including the heavy steam frigate *Merrimack.* The Confederates raised the *Merrimack,* installed a ram and slanting casemates made from railroad track over thick wooden backing, as had been done in the *Gloire,* and renamed it *Virginia.* Mounting 10 guns, including four rifled ones, the *Virginia,* with yard workmen still on board finishing up, sailed on March 8, 1862, for its trial run. Defying concentrated fire of ship and shore batteries, it sank two ships of the Union's wooden blockading fleet before retiring with the ebbing tide. In this dramatic moment John Ericsson's *Monitor* arrived from New York during the middle of the night. Displacing fewer than 1,000 tons, less than one-third of the *Virginia,* the *Monitor* had a boxlike iron hull supporting an iron-plated wooden raft on which revolved the turret. The 172-foot-long vessel had little freeboard except for the thickly armoured rotating turret within which were mounted two 11-inch smoothbores.

The *Monitor* and *Merrimack*

The *Monitor* had many deficiencies. Not really a seagoing warship, it had nearly sunk on its voyage down from New York and did sink on its next sea voyage. Yet it proved the equal of its rival in their duel on March 9. The battle ended in a draw with neither ship seriously injured, but the repercussions of this first duel between completely ironclad warships swept the world.

On April 4, scarcely more time than required for a ship to cross the Atlantic, Great Britain ordered the 131-gun ship of the line, the *Royal Sovereign,* to be cut down, armoured, and fitted with turrets. Only three and a half weeks later Great Britain laid down the *Prince Albert,* the Royal Navy's first iron-hulled turret ship, mounting four turrets.

The Union Navy ordered 66 coastal and river monitors; these were low freeboard ships that were unsuitable for high-seas action and rarely suitable for long voyages. Many were not completed in time for war service. Besides the *Virginia,* the Confederates began a number of other ironclads. Several of these rendered valuable service and probably lengthened the war, but most had to be destroyed before completion. Out of a combination of characteristics of the *Monitor* and *Virginia* types evolved the battleship, which was next to rule the sea.

Toward the battleship. The later 19th century continued to be a time of great flux in warship design. European nations tried numerous arrangements of guns and armour, such as centreline turrets, a central armoured citadel with large guns on turntables at each corner (see Figure 29), lightly armoured big guns topside in barbettes (open-top breastworks), torpedoes in even the largest vessels, and substitution of high speed for armour.

For a time even the ancient ram was revived. When the Austrians won the Battle of Lissa from the Italians in 1866 by ramming, its value for the future seemed confirmed. Hence for years most large ships carried rams, which proved to be more dangerous to friend than foe when ships were sunk in peacetime collisions.

The first torpedoes

This period also saw a fundamental advance in underwater weaponry with the invention of the locomotive torpedo. After being presented with the idea by an Austrian naval captain in 1864, a British engineer named Robert Whitehead produced a projectile that was driven by compressed air and was designed to strike a ship's unprotected hull below the waterline. The Whitehead torpedo, as it was quickly adapted by the European navies, was about 16 inches in diameter and had a range of about 1,000 yards at approximately seven knots.

Engines for all the types of warships steadily improved as stronger metals made possible higher steam pressures and weight reduction. In the 1870s a third cylinder was added onto the two-stage compound steam engine to make the triple expansion engine, and in the 1890s a fourth cylinder was added. These improvements on the traditional reciprocating steam engine provided a marked increase in speed that was surpassed only by the radical innovation of the steam turbine at the end of the century.

The National Maritime Museum, London

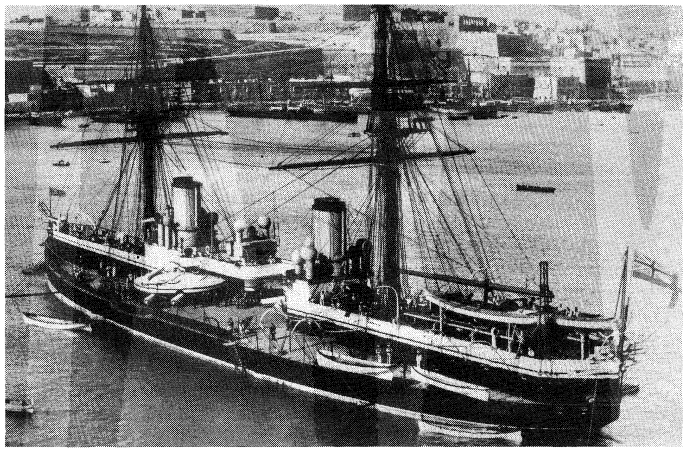


Figure 29: *HMS Inflexible, "central citadel" battleship of the Royal Navy.*
Launched in 1876, it mounted four 80-ton, 16-inch muzzle-loading guns in two hydraulically powered turrets. For greater stability, the engines and powder magazines were gathered toward the centre of the ship and protected by up to 24 inches of iron. The masts were removed in the 1880s.

Ships. A trend toward the centreline-turret, big-gun battleship finally became clear. In it were combined the seagoing hull, armour, and habitability of the *Virginia, Gloire,* and *Warrior* with the revolving turret and big guns of the *Monitor.*

HMS *Monarch,* 8,300 tons, mounting four 12-inch guns in two turrets, and commissioned in 1869, was perhaps the first true seagoing turret warship. HMS *Devastation,* 9,330 tons, four 12-inch guns in two turrets, and massively armoured, was completed four years later without sail and was a next step toward the ultimate 20th-century battleship, a ship with an armoured citadel around the propulsion plant, powder magazines, and handling rooms. Rising out of it, protecting big guns and crews, were barbettes and turrets. The main battery shrank to a few powerful guns, but these took the place of many in broadside because of their great size and ability to fire through a wide arc of bearings.

The change was vividly illustrated by the "new navy" the United States began building in the 1880s, consisting not of improved monitors but of powerful seagoing capital ships with mixed-calibre main batteries. Displacing 11,700 tons, these vessels had 18-inch belt armour and a speed of 15 knots and mounted four 13-inch guns in two turrets. They also mounted eight eight-inch guns in four turrets, smaller guns for defense against torpedo boats, and six torpedo tubes. The plan was, as in other navies, to employ the heavy guns against an enemy ship's armour-protected machinery and magazines while the faster-firing eight-inch guns attacked its relatively unprotected superstructure.

Armoured cruisers

The armoured cruiser was developed in this period as a large, fast vessel armed with intermediate-calibre guns and protected by armoured deck and medium-weight belt armour. Designed for commerce protection and raiding, as well as to cooperate with the battle line in fleet action, it was considered powerful enough and sufficiently protected to fight any ship capable of catching it and able to outrun battleships. Some even held it should become the principal warship.

Less heavily armoured was the protected cruiser, the engines and magazines of which were shielded by an armoured deck, but which lacked an armour belt. Unprotected cruisers had little or no armour, carried fairly light guns, and were designed primarily for scouting, patrolling, and raiding.

Carrying the new self-propelled torpedo, the torpedo boat had great potential, particularly under conditions of low visibility. Small, unseaworthy, and useful only in restricted waters with the then-short-range, slow torpedoes, the new boats did not immediately live up to expectations; nevertheless, as craft and torpedo improved, they were soon regarded as a major menace.

Armour. Early hull armour had been of wrought iron backed by wood. To increase resistance against ever more powerful rifled guns, compound armour of steel backed with iron was devised to combine steel's surface hardness with iron's resiliency. The firm Schneider & Cie in France invented an oil-tempering process to produce a homogeneous steel plate that had good resiliency and greater resistance than compound armour. The later addition of nickel further improved its resistance.

Steel-armour-piercing shells came into use in the late 1880s, again threatening the armoured ship. Accordingly, an American engineer, Hayward Augustus Harvey, perfected a face-hardening process, applying carbon to the face of the steel plate at very high temperatures for an extended period and tempering. Harvey nickel-steel armour superseded earlier types. Then, in 1894, the Krupp firm of Germany devised hot-gas tempering, based on Harvey's process, which in turn became standard with world navies. Later, the addition of chromium to nickel steel was found to be a further improvement.

Armament. The impact of developments in guns and powder exceeded even that of warship design in their effect upon navies. In the two decades after the American Civil War the main difficulties with breech mechanisms were resolved. Better guns, along with breech-loading, made possible both longer ranges and higher rates of fire.

New powders were equally important. About 1880 brown or cocoa powder appeared, employing incompletely charred wood. It burned slower than black powder and hence furnished a sustained burning that was effective ballistically but did not create excessive pressures within the gun barrel. To take advantage of this for longer-range firing, gun-barrel lengths jumped to 30–35 times bore diameter.

Armour-piercing projectiles

Several nations began to achieve success with smokeless powder of nitrated cellulose and usually some nitroglycerin. With greater striking power available, armour-piercing projectiles became more formidable. These were originally solid shot designed simply to punch through armour plate. In the 1890s, better steel and fuses made it possible to add an explosive charge. The resulting semi-armour-piercing shells became highly destructive, and in time all of the armour-piercing projectiles carried explosive charges.

In 1881 the British Admiralty advertised for an anti-torpedo-boat gun to fire cased ammunition at a rate of 12

shots per minute. Benjamin Berkeley Hotchkiss, an American ordnance engineer with a factory in Paris, produced a series of one-, three-, and six-pounder rapid-fire guns that vastly increased the rate of fire for small guns.

(E.McN.E./R.L.Sc./J.C.Re./J.F.G.)

THE AGE OF BIG GUN AND TORPEDO

From the late 19th century through World War I, the greatest driving force in warship development was the rivalry between the big gun and the torpedo. Improvements in these weapons had immense influence on the design and use of surface warships, from the huge dreadnought battleships to the small torpedo boat.

Armament. *Guns.* By 1900 a major change had occurred in the handling of the very heavy main guns, those of 11 to 13.5 inches calibre that fired shells weighing up to 1,300 pounds. In the 1890s such weapons often fired no faster than once every five minutes, compared to the five to 10 rounds per minute fired by a six-inch gun. As power control became easier and more precise, the big guns became more effective. By 1900 it was possible for a 12-inch gun to fire one or two aimed shots per minute.

Meanwhile, the standard of heavy-gun marksmanship began to improve. Although rifled guns had grown bigger and muzzle velocity had increased throughout the late 19th century, there had been no corresponding improvement in fire control. For this reason, effective battle ranges had not extended much beyond 3,000 to 4,000 yards. Then it was discovered that a ship's roll and pitch could be systematically compensated for, so that each shot could be fired at the same angle to the sea and reach almost exactly the same range. Greater accuracy could be achieved by firing groups of shells, or salvos, bunched around the estimated range. The pattern of splashes raised by a salvo would then make corrections possible. By the end of World War I, fire control had improved enough that guns firing 15,000 to 20,000 yards could attain a hit rate of 5 percent. This meant that a ship firing 10 heavy guns at the rate of once or twice per minute could expect a hit after two or three minutes.

The growing range of naval gunnery

Increased range was valuable for two reasons. First, a ship that could hit at ranges beyond the capabilities of its enemies could stand off and destroy them at leisure. Second, improved gun range increased protection against the new, longer-range torpedoes.

Torpedoes. Modifications and adaptations of the original Whitehead design quickly made the torpedo a formidable weapon. Directional control was greatly improved in the 1890s by the use of a gyroscope to control the steering rudders. Another significant improvement was the use of heat engines for propulsion. British firms, introducing both heat engines and contrarotating propellers, advanced to the high-performance, steam-driven Mark IV torpedo of 1917. Concurrently with this development, an American firm, E.W. Bliss Company, successfully used a turbine to drive a modified Whitehead design. (This Bliss-Leavitt torpedo remained in extensive use until World War II.) By 1914, torpedoes were usually 18 or 21 inches in diameter and could reach almost 4,000 yards at 45 knots or 10,000 yards at close to 30 knots.

Armour. The torpedo threat forced ship designers to provide battleships with underwater protection. Schemes to place coal bunkers near the outside of the ship proved impractical, but research during World War I showed that the basic idea of keeping the underwater explosion at a distance from the interior of the ship was correct. In the Royal Navy, existing ships were fitted with external bulges or "blisters" to keep the explosion farther outboard, and new ships were built with specially designed layers of compartments designed to absorb the shock of explosion.

During the war it also became apparent that the longer firing ranges meant that more shells would fall onto a ship's deck than on its side armour. Because these ranges were experienced at the Battle of Jutland, ships designed afterward with stronger deck armour were called post-Jutland.

Propulsion. *Steam turbines.* While weapons were the main driving force in warship development, changes in propulsion were also important. In 1890, propulsion was exclusively by reciprocating (*i.e.,* piston) steam engines, which were limited in power and tended to vibrate. To escape these limits, warship designers adopted steam turbines, which ran more smoothly and had no inherent limits. Turbines were applied to destroyers from about 1900 and to battleships from 1906.

The main drawback of turbine propulsion was that really efficient turbines ran too fast to drive efficient propellers. The solution was to reduce turbine speeds to acceptable propeller speeds through gearing. By 1918, single-reduction gearing was commonplace. Late in the interwar period, the U.S. Navy adopted double-reduction gearing, which permitted even higher turbine speeds without requiring propellers to run any faster.

From coal to oil

Fuel. Fuel also became a major issue. Coal was relatively inexpensive and easily available; however, it did not burn cleanly and was difficult to transfer from ship to ship at sea. Oil, on the other hand, burned cleanly, and it could be transferred easily at sea. Also, it had a higher thermal content than coal, so that the same weight or volume of oil could drive a ship much farther. The United States shifted to oil fuel in new ships in about 1910 and converted its remaining coal-burning warships after World War I. Beginning with the Queen Elizabeth class of battleships in 1915–16, Britain switched to oil. The other navies followed suit after the war.

Internal combustion. In contrast to the steam engine, a gasoline or diesel engine often needed no tending at all, could be very compact, and could start and stop quite easily. Such engines made it possible to build small, fast coastal minesweepers, subchasers, and motor torpedo boats. Internal combustion was thought to be especially suitable to subchasers, which would have to stop their engines while listening for a submarine and then start them up suddenly when something was heard.

Battleships. A battleship entering service in 1900 typically mounted a mixed battery of four heavy (11- to 13.5-inch) guns in two twin turrets, about a dozen secondary guns of six to nine inches, and small, fast-firing guns of three inches or less for beating off torpedo-boat attacks. These ships usually displaced 12,000 to 18,000 tons.

By 1904 studies reinforced by battle experience in the Spanish-American and Russo-Japanese wars indicated that fire from large guns at longer ranges was more effective than mixed-battery fire closer in. Only bigger shells could do serious damage to well-armoured ships. Moreover, the shells fired from guns of many different calibres produced a confusing pattern of splashes in the water that made the correcting of aim and range quite difficult. Effectively increasing range, then, depended upon abandoning the multiple-calibre pattern of previous battleship armament in favour of a single-calibre armament. Several navies reached this conclusion simultaneously, but the British were the first to produce such a ship, HMS *Dreadnought,* completed in 1906. Displacing about 18,000 tons, it carried 10 12-inch guns; its only other armament consisted of three-inch weapons intended to fight off destroyers.

Dreadnought and the all-big-gun ship

The *Dreadnought* gave its name to an entirely new class of battleships of the most advanced design. By 1914 the Royal Navy had 22 dreadnoughts (another 13 were completed during World War I), Germany built a total of 19 (five completed after 1914), and the United States completed 22 (14 of them after 1914). Japan and Italy built six, while Russia and France each built seven. Not all of these ships were strictly equivalent. Unlike its immediate German and American contemporaries, the *Dreadnought* had steam turbines in place of reciprocating engines. These enabled it to attain a speed of 21 knots, which was hitherto achieved only by cruisers. (Contemporary battleships were generally limited to about 18 knots.) Thus, in mobility as well as in size, the *Dreadnought* began a new era.

HMS *Dreadnought* also marked a beginning of rapid development in big-gun firepower. In 1909 the Royal Navy laid down HMS *Orion,* the first "super dreadnought," which displaced 22,500 tons and was armed with 13.5-inch guns (see Figure 30). The U.S. Navy followed with ships armed with 14-inch guns. Then, on the eve of World War I, the Royal Navy went a step further with HMS *Queen Elizabeth,* armed with 15-inch guns and capable,

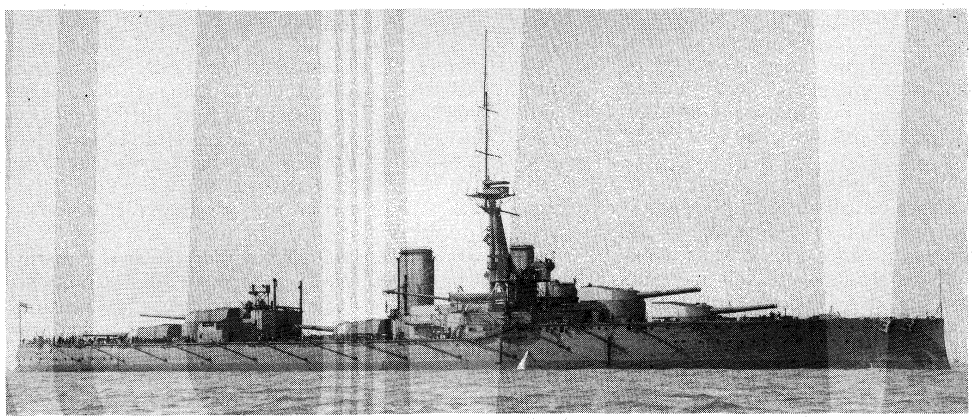


Figure 30: *HMS Orion, super dreadnought battleship of the Royal Navy.* Heavier than HMS *Dreadnought* but just as fast, this ship mounted 10 13.5-inch guns of greater armour-piercing power in five turrets along the centreline of the vessel. The *Orion* was present at the Battle of Jutland in 1916 and was scrapped under the Washington Treaty of 1922.

The National Maritime Museum, London

in theory, of 25 knots. World War I stopped the growth of British and German battleships, but the United States and Japan continued to build ships exceeding 30,000 tons displacement. In 1916 both countries adopted the 16-inch gun, which fired a shell of approximately 2,100 pounds. Such guns could be aimed to hit at ranges as great as 20,000 yards.

The battleship saw little combat in World War I, yet, despite submarines, aircraft, and destroyers, the outcome of the war still hinged upon control of the sea by the battleship. Had superiority in battleships passed to Germany, Britain would have been lost, and the Allies would have lost the war. The one moment when this might have happened was the only large-scale clash of battleships, the Battle of Jutland. Fought in May 1916 in mist, fog, and darkness, Jutland revealed the strengths and weaknesses of battleships and battle cruisers. Three British battle cruisers were lost. Several German battleships, thanks to watertight subdivision and efficient damage-control systems, survived despite more hits. But the British advantage in numbers was decisive, and Germany turned to the submarine to counter the Allied blockade.

Cruisers. HMS *Dreadnought* made earlier large cruisers obsolete, since it was nearly as fast as any of these ships. Consequently, the Royal Navy built a series of ships it called battle cruisers. These were as large as the newest battleships and were armed with battleship guns, but they were much faster (initially a top speed of 25 knots, compared with the 21 knots of battleships). The first was HMS *Invincible,* completed in 1907. Many of these ships were built: 10 for the Royal Navy before 1914, seven for Germany, and four for Japan.

The strength and weakness of battle cruisers

Battle cruisers gained their superior speed by sacrificing heavy armour; as a consequence, they could not stand up to battleships. This was proved at the Battle of Jutland, where the *Invincible* was blown in two by a single salvo and sunk along with two other battle cruisers. These losses led many to argue that the battle cruiser was a mistake, but during the war Britain laid down six more, three of which were eventually completed. The last of them, HMS *Hood,* launched in 1918, could be described as a new stage in warship development. It was so large, at 41,200 tons, that it could combine contemporary battleship armour and armament (equivalent to that of HMS *Queen Elizabeth*) with the very high speed of 31 knots. Although classed as a battle cruiser, it was actually the first of a new generation of very fast battleships.

At the other end of the cruiser spectrum were small, fast "scout" cruisers used for reconnaissance and escort duties. These ships displaced from 3,000 to 7,000 tons and, by 1915, attained speeds as high as 30 knots. They were armed with guns of smaller calibre, usually six or 7.5 inches. The British built many of this type of cruiser, as well as larger types that were nevertheless smaller than their battle cruisers.

Destroyers. The self-propelled torpedo had its greatest impact on the design of small surface ships. Beginning in the 1880s, many nations built hundreds of small steam torpedo boats on the theory that they could bar coastal waters to any enemy. Because their hulls could be crammed with machinery, torpedo boats were quite fast. By the early 1890s, speeds as high as 25 knots were being reported. As a defense against this new fast threat, Britain deployed oversized torpedo boats, calling them torpedo boat destroyers. These craft were successful in hunting down torpedo boats, and eventually they were renamed destroyers.

The first destroyers were essentially coastal craft, displacing only about 200 tons, but their larger successors could accompany battle fleets to sea. There it soon became apparent that a destroyer was in effect a superior sort of torpedo boat, capable of delivering its weapon against capital ships during or immediately after a fleet engagement. By 1914, 800- or even 1,000-ton ships were quite common.

During World War I British destroyer design changed radically, creating what became the postwar formula of the V and W destroyer classes: four four-inch guns superimposed fore and aft, a high forecastle forward for greater seakeeping ability, and two sets of twin (later triple) torpedo tubes amidships. These vessels, displacing about 1,200 tons and capable of 34 knots, made all earlier British destroyers obsolete.

Convoy protection

When Germany adopted unrestricted submarine warfare in February 1917, shipping losses soon forced the diversion of destroyers from fleet duty to convoy protection and antisubmarine warfare. Destroyers were not ideally suited to the escort role, as they had limited steaming range and their high-speed design made them less seaworthy than the merchant ships they were required to escort. The Royal Navy therefore built several types of specialized convoy escort, but the U.S. Navy found it easier to mass-produce its current destroyer design. These vessels, equipped with hydrophones and depth charges, as well as guns and torpedoes, overcame the submarine threat and had a large share in the safe convoy of two million American troops to Europe without loss of a single soldier.

THE AGE OF THE AIRCRAFT CARRIER

Although naval strategists continued to extol the battleship and battle cruiser after World War I, these capital ships soon were swept away by the new art of naval aviation. Conventional naval guns were limited to a range of perhaps 20 miles, but by World War II the aircraft carrier—a ship capable of launching, recovering, and storing aircraft that could themselves destroy ships—had extended the battle range of surface fleets by as much as 300 miles. In doing so, it had a profound effect on naval warfare.

The last capital ships. In 1922 the Five-Power Naval Limitation Treaty, signed in Washington, D.C., by emissaries of the victorious Allies of World War I plus Japan,

changed the character of navies by limiting battleship inventories. With a few exceptions, new battleship construction was prohibited until 1931, and most remaining predreadnought battleships were ordered scrapped. The new battleships allowed by the treaty could not mount guns of greater calibre than 16 inches, and they could not displace more than 35,000 tons.

Battleships were defined as warships armed primarily with guns over eight inches in calibre or displacing more than 10,000 tons. This definition of a battleship in effect defined a new kind of cruiser, which would displace about 10,000 tons and would be armed with eight-inch guns. In 1930 a new treaty, signed in London, extended the battleship-building "holiday" through 1936 and divided cruisers into two classes: ships armed with guns of up to 6.1 inches and ships armed with guns of 6.1 to eight inches. In U.S. parlance the former were light, and the latter heavy, cruisers.

One peculiarity of the Washington Treaty was that it defined warship size by devising new "standard" tonnages, which excluded the weight of fuel and reserve feed water. (Standard tonnage remains a means of measuring ship displacement in many cases, and it is used here when ship tonnages are listed.) The effect of the London Treaty's limit on cruiser tonnage was the saving of weight in warship design. Several navies used aluminum in structures not contributing directly to the strength of their ships, and there was considerable interest in welding (which was lighter than riveting) and in more efficient hull structures. Lighter machinery was also developed. The U.S. Navy, for example, built higher-pressure, higher-temperature boilers and more efficient turbines.

The new "treaty" warships

Most of the battleships that survived the scrappings were rebuilt during the 1920s and '30s with added deck armour and with new blisters to improve their resistance to underwater explosions. In many cases, lighter engines and boilers were fitted, so that weight and internal volume were freed for other purposes such as improved fire-control computers.

New battleships were also built. The Treaty of Versailles limited Germany to 10,000-ton capital ships, but in the 1930s that country built three large cruisers of about 12,000 tons, each armed with six 11-inch guns. These so-called pocket battleships, by combining heavy armour with great speed (provided by diesel engines), could defeat any contemporary cruiser. They also reignited the race in battleship construction. In 1935 France produced the *Dunkerque;* at 26,500 tons, armed with eight 13-inch guns, and reaching 30 knots, this was the first of the new generation of "fast battleships" presaged by HMS *Hood.* In 1937, after the Washington and London treaties had expired, Japan laid down the *Yamato* and *Musashi.* These two 72,800-ton ships, armed with 18.1-inch guns, were the largest battleships in history.

As World War II began, Britain was constructing five battleships of the King George V class. These displaced about 36,000 tons and carried 14-inch guns. The United States completed five 35,000-ton battleships before entering the war and one in 1942, and four 45,000-ton Iowa-class ships were built during wartime. The Iowa ships, carrying 16-inch guns, were the last battleships completed in the United States. Germany completed five ships (including the 42,000-ton *Bismarck* and *Tirpitz* and the 32,000-ton *Scharnhorst*), France completed four, Italy completed three, and Japan completed two. Most of these fast battleships could exceed 30 knots.

Before the war began, the new arts of dive-bombing and torpedo-bombing from carrier-based aircraft did not promise enough velocity and destructive power to penetrate battleship armour. But by the end of the war, even modern capital ships maneuvering at sea could be sunk by carrier aircraft. In October 1944 and April 1945, U.S. carrier-based airplanes sank the *Musashi* and *Yamato;* more than any other event, these marked the end of the long reign of the battleship.

Aircraft carriers. *World War I.* The airplane had just begun to go to sea on the eve of World War I. In November 1910 the American scout cruiser USS *Birmingham* launched the first airplane ever to take off from a ship, and two months later a plane was landed on an improvised flight deck built onto the armoured cruiser USS *Pennsylvania.* In 1913 a British cruiser, HMS *Hermes,* was converted to carry aircraft. In 1916, flying-off decks were built aboard several British ships, and by 1918 the Royal Navy had a converted passenger liner, HMS *Argus,* that could land and launch planes on a flight deck extending from bow to stern. The *Argus* was the world's first true through-deck aircraft carrier and was thus the prototype for all later carriers.

The first true aircraft carrier

Aircraft carriers were valuable in World War I primarily because their planes vastly extended a ship's ability to scout, or reconnoitre, large areas of ocean. The wartime Royal Navy developed a series of torpedo-carrying seaplanes and carrier-based light bombers, but both the aircraft and their weapons were too weak to pose a serious threat. For this reason, the aircraft carrier was considered an essential element of the fleet but not a replacement of the battleship.

Improvements between the wars. Throughout the interwar period, naval aircraft performance gradually improved, and dive bombers and torpedo bombers made aircraft carriers effective ship killers. In the opinion of many experts, this made other carriers so vulnerable that the only way to protect them was to find and destroy the enemy's carriers first. Another option was to protect the carrier with its own fighters. This option was not practical without some means of detecting an enemy air attack at a great distance, so that defending fighters could be sent up in time. The key to such a defense was radar. The phenomenon of radar was observed in the 1920s, and by the late 1930s prototype sets with huge antennas were operating. Radar was first installed aboard British and U.S. carriers in 1940–41.

As another defensive measure, in 1936 the Royal Navy decided to provide its new carriers with armoured hangars, the armour including part of the flight deck. The U.S. Navy, on the other hand, built its flight decks of wood, on the theory that damage from bombs to the decks could be repaired relatively easily. (Substantial armour lower in the ships was intended to preserve them from more serious bomb damage.)

Aircraft carrier operation required three elements: a means of launching from the ship, a means of recovering aircraft aboard ship, and a means of stowage. Landing aircraft were caught by arresting wires strung across the deck that engaged a hook fastened under the planes' tails. Originally, arresting wires were needed to keep the very light wood-and-cloth airplanes of the World War I era from being blown overboard by gusts of wind. After heavier steel-framed and steel-skinned airplanes were introduced, wires were no longer necessary. The Royal Navy abandoned arresting gear about 1926. The U.S. and Japanese navies continued to use it, but for a very different purpose: to keep landing airplanes from rolling into aircraft that were stowed at the forward end of the flight deck. In British practice this was unnecessary, because aircraft were stowed below immediately upon landing, so that each pilot faced a clear deck when he landed. Stowage was accomplished by elevator lifts, which were usually located in two or three places along the centreline of the flight deck.

Aircraft stowage: flight deck and hangar deck

World War II. The Washington Treaty of 1922 permitted each of the major powers to convert two capital ships to carriers, within a 33,000-ton limit. New carriers could not displace more than 27,000 tons, and no carrier could have guns of more than eight inches. The United States and Japan converted heavy battle cruisers just under construction into the USS *Lexington* and *Saratoga* and the Japanese *Akagi* and *Kaga.* These ships actually exceeded the 33,000-ton limit, the U.S. vessels carrying about 80 aircraft and the Japanese about 40. Two new U.S. carriers built in the 1930s to treaty specifications were the *Yorktown* and *Enterprise,* which displaced more than 20,000 tons and carried about 80 aircraft. Their Japanese equivalents were the *Hiryu* and *Soryu,* which operated about 50 aircraft. Britain, which had suspended new capital-ship construction during the war, converted two light battle cruisers completed in 1916, HMS *Courageous* and *Glorious.* For economic reasons Britain did not build a new

carrier to the treaty specifications until 1935, when HMS *Ark Royal* was laid down.

Under a new treaty of 1936, new carriers were limited to 23,000 tons, but the limit on the total number of carriers was removed. In response, the Royal Navy laid down the Illustrious class of 23,000-ton carriers. These vessels did not enter service until after the outbreak of World War II in 1939. With the commencement of war, the United States produced the 27,500-ton Essex class. Carrying more than 100 aircraft, these vessels became the principal fleet carriers of the Pacific Theatre (see Figure 31). Between 1940 and 1943, the United States also designed a series of 45,000-ton ships partly inspired by Britain's Illustrious carriers. Completed after the war ended in 1945, this Midway class was the first of the U.S. carriers to be built with armoured flight decks.

During the war Britain built second-line carriers, called light fleet carriers, which were designed for quick construction. These became the Colossus and Majestic classes, vessels of approximately 15,000 tons that carried about 40 aircraft each. The U.S. war program, meanwhile, included the conversion of a series of cruisers into light carriers of the 11,000-ton Independence class.

For protecting merchant convoys from submarine attack, escort carriers were built in large numbers, mainly in the United States. Many were converted merchant ships, and others were specially built on hulls originally designed for merchant service. The Royal Navy also added flight decks to some tankers and grain carriers, without eliminating their cargo role. These were called MAC ships, or merchant aircraft carriers.

Carriers played a dominant role in every aspect of operations at sea in World War II. The Pacific conflict began with the Japanese carrier strike against Pearl Harbor and ended with American and British carriers operating with impunity against the Japanese homeland. In between, the Battle of the Coral Sea, in May 1942, was the first naval battle in history in which opposing fleets fought without ever coming in sight of each other. A month later off Midway atoll, carriers again played the decisive role. The Battle of Midway reinforced a conviction already clear, especially from British operations in the Mediterranean with and without air support, that control of the sea also meant control of the air over the sea. In the autumn of 1942 the Solomon Islands campaign underlined the importance of both aircraft and submarines in fleet operations, emphasizing that modern sea power was a trident of air, surface, and undersea forces.

Naval aircraft at the Coral Sea and Midway

Destroyers and escort ships. Most destroyers built between the two world wars repeated Britain's V and W formula, sometimes with more powerful guns or with more torpedo tubes and generally displacing from 1,300 to 1,500 tons. The London Treaty of 1930 prohibited destroyers larger than 1,500 tons, but by the late 1930s several navies had exceeded the limits.

Besides delivering a bomb with enough velocity to damage a capital ship, the dive bomber forced the addition shipboard of large numbers of automatic guns, of 40 millimetres or less, to supplement the more powerful but slower-firing three- to five-inch antiaircraft guns. The Royal Navy converted some of its small World War I cruisers into antiaircraft ships, replacing their single six-inch guns with twin four-inch weapons controlled by special antiaircraft directors. The Japanese built large destroyers (the Akitsuki class) for much the same role; these were armed with a special 3.9-inch gun. The U.S. Navy provided virtually all of its destroyers with effective antiaircraft guns.

As in World War I, destroyers were used for convoy escort against submarines, if only because they were available in large numbers. However, they were not especially suited to that purpose; like their pre-1914 forebears, they were still primarily fast fleet escorts optimized to deal with surface torpedo attack. The likelihood of such attack declined as radar became widely available, but aircraft remained an important threat to major fleet units, so that the destroyer naturally evolved into an antiaircraft escort.

One important exception to the general abandonment of surface torpedo attack was the Imperial Japanese Navy. By 1941 Japanese doctrine envisaged concentrated night attacks by cruisers and destroyers carrying large numbers of unusually powerful, oxygen-fueled, wakeless torpedoes. These torpedoes were the Type 93 Long Lances, which proved extremely effective in the U.S.-Japanese naval battles around the Solomon Islands in 1942–43.

The new escort frigates

The submarine threat in World War II placed Britain, the United States, and Japan in desperate need of escorts for merchant convoys. Besides converting existing destroyers, each navy built huge numbers of specialized escorts adapted to mass-production techniques. Britain led in these measures, building relatively small escorts of limited endurance, which it called corvettes, and much larger escorts, which it called frigates. The U.S. Navy built a somewhat faster equivalent, which it called a destroyer escort. The Japanese built a series of escorts roughly equivalent to the British corvettes.

Torpedo boats. In the 1930s the German, Italian, British, and U.S. navies regained interest in motor torpedo boats, which had been largely discarded after World War I. All four navies built them in substantial numbers to fight in narrow seas during World War II. Against convoys in the English Channel and the North Sea, the Germans used their S-boats (*Schnellboote,* "fast boats"; often called E-boats by the British). The U.S. Navy's PT (Patrol Torpedo) boats harassed Japanese traffic in the South Pacific. Some of these wooden-hulled craft, which were powered by diesel or gasoline engines, could reach speeds of 40 knots. In addition to torpedoes, they could carry significant gun armament.

Amphibians. The internal combustion engine made

U.S. Navy photograph

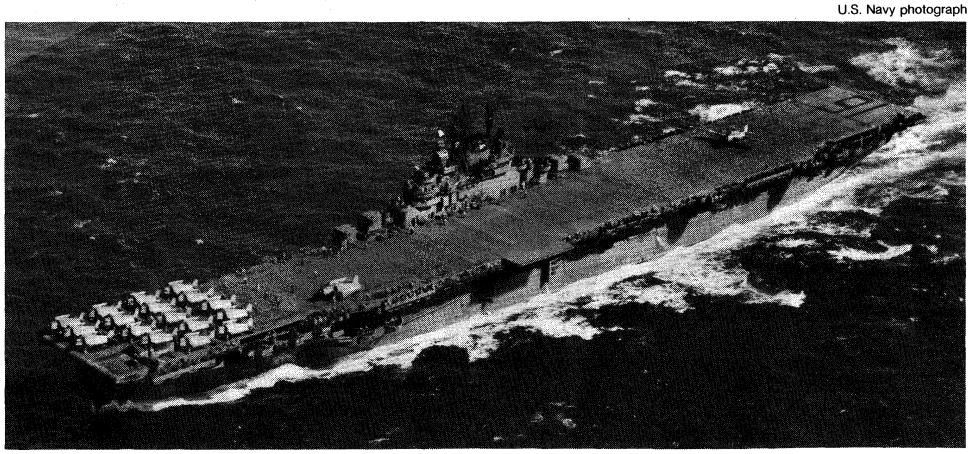


Figure 31: *USS Lexington, Essex-class aircraft carrier of the U.S. Navy.*
A TBF Avenger torpedo bomber lands over the stern; parked at the other end of the 875-foot flight deck are F6F Hellcats. Named for an earlier carrier sunk in the Pacific, in 1944 the *Lexington* took part in the battles of the Philippine Sea and Leyte Gulf. During the latter battle its planes helped sink the Japanese battleship *Musashi.*

possible the most spectacular naval innovation of World War II, the shallow-draft landing craft used to bring large forces quickly to enemy beaches during amphibious assaults. The most famous example of these was the LST (landing ship, tank), a large beaching craft that could embark and disembark troops and vehicles directly from shore to shore. The LST displaced about 4,000 tons full load and transported about 150 troops with equipment at 10 knots.

A beaching craft of intermediate size, which the U.S. Navy called the LCT (landing craft, tank), was carried over oceanic distances and launched at the time of assault. The LCT was too large to fit the davit of a conventional transport, so a new type of ship, the LSD (landing ship, dock), was created specifically to carry it. The LSD had a floodable well deck aft, like a miniature dry dock. It could carry tank-laden LCTs over oceanic distances then flood its well deck off a landing beach and launch the craft.

Not all of these vessels were powered by internal combustion engines; some LSTs and many LSDs used steam. Nevertheless, the vital smaller craft, such as the LCTs and a series of small infantry carriers called LCVPs (landing craft, vehicle, personnel), could not have been built without using diesel power plants. Only because their engines and fuel consumed so small a portion of their total displacement could these craft carry such massive loads on shallow drafts.

THE AGE OF THE GUIDED MISSILE

By the middle of World War II, carrier-borne aircraft had become so effective that the aircraft carrier was clearly replacing the battleship as the core of the modern navy. After the war, the development of jet aircraft and nuclear-powered ship propulsion magnified the range and speed of operations, but they did not alter the central role of the carrier.

The new threat to capital ships

At the same time, though, a new equalizer was being developed: the antiship guided missile. This weapon, which could be mounted onto the smallest surface vessels as well as aircraft and submarines, was especially dangerous to aircraft carriers because it could be launched outside antiaircraft range and, being unmanned, could not be distracted easily by defensive fire. The main defense was to provide the fleet with its own guided missiles capable of destroying either the missile or its launching platform.

Propulsion. *Nuclear power.* The ultimate development in steam propulsion was the use of the energy released by nuclear fission to heat the boilers of steam turbines. Nuclear power was proposed for ships, particularly submarines, in 1945, and by 1955 the United States had a nuclear submarine, USS *Nautilus,* in service. Other navies followed suit, so that within 20 years Britain, the Soviet Union, France, and China all operated nuclear submarines. In the 1950s the United States also developed nuclear power plants for surface ships, subsequently installing them aboard aircraft carriers and their escorts. The Soviet Union and France followed with more limited programs in the 1970s and '80s.

For a surface ship, the advantage of nuclear power was effectively infinite range at high speed. The disadvantage was the high cost, which limited such power to a few valuable ships.

Gas turbines. After 1945 the gas turbine, a turbine in which the combustion of fuel generates a stream of gases that turns the rotor, became available for ship propulsion. Gas turbines shared with internal combustion piston engines the great virtues of quick starting and stopping as well as relatively simple operation. They were also quite reliable. Their main defect was that they were efficient only over a relatively narrow speed range. For this reason, the first gas turbine warships employed combination power plants, such as combined steam and gas turbine (COSAG) or combined diesel and gas turbine (CODAG). Using such a plant, a relatively small ship, such as a frigate, could achieve much higher speed than with a conventional steam turbine. The next step was to combine two gas turbines, one sized for cruising and the other for high speed. Such an arrangement might be either combined gas and gas (COGAG), with both plants able to operate together, or combined gas or gas (COGOG), with only one plant being used at a time.

Systems employing the gas turbine proved useful in smaller escort ships such as destroyers and frigates, although they were also installed in cruiser-sized vessels. A related system, called combined diesel, electric, and gas turbine (CODLAG), was especially valuable in submarine warfare. In order to minimize engine noise, which might interfere with sonar sensors, diesel generators powered electric motors, which in turn drove the ship's propellers. For higher speeds, electricity was supplemented or replaced by gas turbines.

Armour. The role of armour greatly declined after 1945 because aircraft, the greatest threat to warships, now carried guided missiles and bombs capable of penetrating the thickest deck armour that any viable ship could accommodate. At the same time, warships' new missile weaponry occupied much more space than did the earlier guns, shells, and powder. Modern weapon systems also required room for computers and radars and for their operators. To cover such spaces with anything but the lightest plating would have added enormous weight and thus required very large and expensive hulls. The high cost of protection (in ship size as well as money) was a major reason for the abandonment of heavy, extensive armour in the guided-missile era.

The abandonment of heavy armour

Armour was not abandoned altogether, however. Thin armour, for example, could protect aircraft and missiles from the steel splinters of exploding warheads and thus could keep a ship hit elsewhere from being destroyed by a huge explosion of jet fuel or its own missiles. For this reason most modern warships adopted thin (about one- or two-inch) splinter protection around their missile magazines.

Aircraft carriers, at least in the U.S. Navy, retained armoured flight decks, though in their case the armour provided structural strength as well as limited protection.

Aircraft carriers. After World War II the heavy attack aircraft carrier developed three roles: to deliver air strikes (both conventional and nuclear) against sea and shore targets; to provide a long-range air-defense umbrella for other ships; and to support antisubmarine operations (leaving it to other ships actually to destroy the submarines). In order to carry out these roles, jet carriers became so huge that only a first-rate power could afford to build and operate them. Within 35 years of the end of World War II, only the United States and France operated full-scale carriers (although the 27,000-ton French Clemenceau class was closer in size to the World War II Essex carriers than to the 80,000-ton, 1,000-foot behemoths built by the United States in the 1970s and '80s). The Soviet Union began to build large carriers in 1983.

Navies that could not afford the large carrier divided its three roles among escort ships and new light aircraft carriers. To the light aircraft carriers was given the role of antisubmarine warfare, along with limited ground-attack and air-protection capabilities.

Large carriers. The main wartime technical development in aircraft carrier design was the hydraulic catapult, but this was barely powerful enough to launch the heavier jet aircraft coming into service after 1945. The problem was solved in 1951, when the British first tested an effective catapult driven by steam from a ship's boilers.

Jet aircraft landed at much higher speeds than had propeller-driven planes, making the installation of better arresting gear necessary. Also, landing control had to be improved, because the approaching pilot had to make crucial decisions much more quickly. As in the case of the steam catapult, the British supplied the solution, in the form of the angled deck and the mirror (later the Fresnel-lens) landing sight. By building an extension of the flight deck to one side and angling the landing strip onto that extension, the British system allowed a pilot to land away from aircraft parked at the end of the flight deck. If he missed the arresting wires, the pilot could fly off to try again. In this way mistakes became much less serious.

The angled flight deck

The mirror landing sight, in effect, allowed the pilot to see his own position relative to the required glide path and to make corrections instantly. Previously, an officer

on deck, observing the landing, had generally ordered the corrections.

By 1955 the modern jet aircraft carrier had emerged, with steam catapults, an angled deck, and a mirror landing system. The first full jet carrier was USS *Forrestal,* commissioned in 1955. The 60,000-ton Forrestal carriers were built with rectangular extensions to the after part of the flight deck; these considerably widened the deck and allowed the angled landing strip to be merely painted on rather than extended over the side. The elevators were shifted to the edge of the flight deck, so that they could operate while aircraft were landing and taking off.

The first nuclear-powered carrier, USS *Enterprise,* was commissioned in 1961. It was equipped with eight nuclear reactors and steamed for more than three years before refueling was necessary. The *Enterprise* displaced 75,700 tons, carried 100 jet aircraft, and could reach more than 30 knots. Beginning in 1975, the Nimitz class superseded the *Enterprise.* These 81,600-ton carriers were powered by only two nuclear reactors, yet they reached speeds comparable to the *Enterprise,* and their uranium cores needed replacement only once every 13 years. The smaller propulsion system created more room for the storage of aviation fuel, which greatly extended the operation of the 90 aircraft carried on these ships.

In 1983 the Soviet Union laid down a 65,000-ton nuclear-powered carrier, and in 1989 France laid down the 34,000-ton *Charles de Gaulle,* a nuclear-powered ship designed to carry 40 aircraft.

Light carriers. The expense of large carriers was due partly to the huge amounts of fuel, ammunition, and maintenance required to keep as many as 80 aircraft operational, but it was also due to the complexity and size of the catapults and arresting gear needed for jets. In the late 1960s Britain developed a jet fighter, the Harrier, that was capable of taking off vertically or (with a heavy payload) after a short roll. A carrier equipped with these V/STOL (vertical/short takeoff and landing) jets could be much smaller than a full jet carrier, because it would need neither catapults nor arresting gear. In the 1970s and '80s, Britain built three such ships, constituting the Invincible class. These 20,000-ton ships carried eight Sea Harriers and about a dozen antisubmarine helicopters. They also incorporated a further British contribution to aircraft carrier design: the upward-sloping "ski jump" at the end of the flight deck to assist the Sea Harriers in short takeoff.

V/STOL jets and the "ski-jump" flight deck

The Italian and Spanish navies also constructed light carriers for helicopters and V/STOL jets. Like the Invincibles, they were powered by gas turbines. The Soviet Kiev class, at more than 30,000 tons, carried a larger complement of rotary and V/STOL craft.

In the 1960s, '70s, and '80s the United States constructed the Iwo Jima, Tarawa, and Wasp classes of amphibious assault ships, descendants of the World War II escort carriers that could transport close to 2,000 marines as well as their weapons and vehicles. The Tarawa and Wasp classes, besides carrying helicopters and Harriers, were built with well decks for the launching of landing craft.

Occupying a position between cruisers and the through-deck light carriers were helicopter carriers, whose flight decks occupied only the after section of the ship. The 17,000-ton Moskva class of the Soviet Union, introduced in 1967, was a prominent example.

Fleet escort ships. In the surface ships supporting aircraft carriers, the most important trend after 1945 was an amalgamation of types. In 1945 cruisers were armoured big-gun ships that were capable of operating independently for protracted periods. Destroyers were part of the screen protecting a main fleet, and frigates were slower ships designed for merchant convoy protection against air and submarine threats—primarily the latter.

This series of distinctions began to collapse in the late 1950s. First, in order to hunt the new fast submarines, frigates had to match destroyer speeds. This made them more like small destroyers. At the same time, most cruisers were converted to carry long-range antiaircraft missiles. This conversion made it clear that cruisers were not solitary raiders or ship killers but fleet escorts—in effect, super destroyers. Eventually all three types became capable of antiaircraft, antisubmarine, and antiship warfare, although individual classes often specialized in one role.

The most prominent trend in armament was a shift from guns to guided missiles. Beginning in the mid-1950s, existing ships had at least some of their guns replaced by missiles, and thereafter new ships were built with missiles making up their main batteries. By 1990 the ranges of these weapons varied from about four nautical miles for a short-range antimissile missile to more than 300 nautical miles for a long-range antiship missile. Some of these traveled at more than twice the speed of sound.

From gun to missile

Main guns became fewer and smaller. By 1990 the most prominent guns were dual-purpose weapons (for antiaircraft as well as surface fire) measuring from 76 to 130 millimetres, or three to five inches. Close-in protection against missiles was provided by fully automatic or Gatling-type guns of 20 to 40 millimetres. All guns were now remotely controlled and directed by radar.

Cruisers. The era of big-gun cruisers ended with the completion of ships laid down during World War II. In 1961 the United States commissioned USS *Long Beach,* the first vessel designed from the keel up as a guided-missile cruiser and the first surface warship to steam under atomic energy. This 14,000-ton ship was followed by a series of nuclear-powered U.S. cruisers that ended, in the 1970s, with the 10,400-ton Virginia class. This class was supplemented in the 1980s by the 7,400-ton, gas-turbine-powered Ticonderoga cruisers. Both the Virginia and Ticonderoga ships were fitted with a broad array of weaponry, including surface-to-air and antiship missiles, tube-launched and rocket-launched antisubmarine torpedoes, and two five-inch and two 20-millimetre guns. In addition, they were supplied with Tomahawk cruise missiles, which could be fitted with conventional or nuclear warheads. The Ticonderoga vessels carried two submarine-hunting helicopters, and they were equipped with the extremely sophisticated Aegis radar system for tracking hostile targets and directing missile defense (see Figure 32).

Department of Defense photo

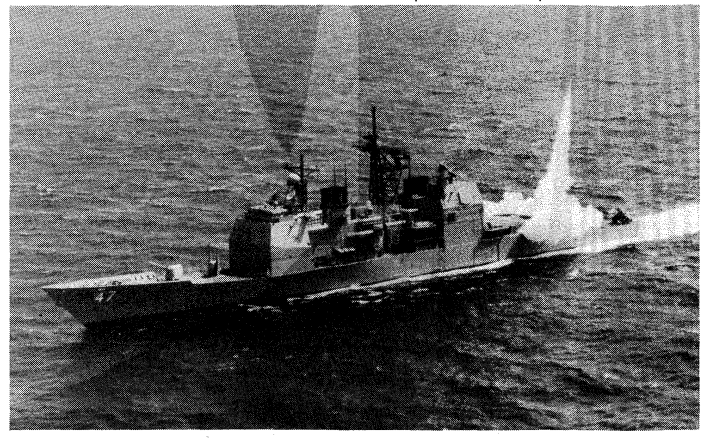


Figure 32: *USS Ticonderoga, guided-missile cruiser of the U.S. Navy.*
A Standard surface-to-air missile is fired from the aft launcher. Equipped with a phased-array radar system that can guide several missiles simultaneously to their targets, the Ticonderoga class has continued the evolution of cruisers into antiaircraft ships escorting carrier task forces.

As the guided-missile cruiser evolved into an escort for aircraft carriers, it ceased to be built by navies that had allowed their large carrier capacities to expire. Britain, for example, sold its County-class ships (which were officially classed as destroyers but were effectively cruisers) in the 1970s and '80s, relying thereafter on smaller escorts to protect its light carriers. The Soviet Union, on the other hand, laid down the first of its 22,000-ton, nuclear-powered Kirov cruisers in 1973. With armament, speed, and steaming range comparable to the Virginias, these cruisers were logical escorts for the new nuclear-powered aircraft carriers that were expected to give the Soviet Navy the ability to project its power around the world. Until then, Soviet guided-missile cruisers had emphasized a heavy complement of long-range antiship missiles, giving

some of them a ship-killing role similar to that of the big-gun cruisers.

The shift to anti-aircraft escort

Destroyers and frigates. Because of the high cost of cruisers, smaller escort ships became the backbone of lesser navies in the guided-missile age. The destroyer completed its transition, begun during World War II, from surface-ship killer to antiaircraft escort. To this duty was added antisubmarine warfare, the traditional role of the frigate. Often the latter was distinguished from the destroyer only by its lesser displacement, armament, and speed.

As submarines became faster, many classes of destroyer and frigate adopted the helicopter (often housed in a hangar in the after section) as a help in hunting them down. Like cruisers, they bristled with an array of sonar and radar sensors and were packed with electronic gear for the swift detection and identification of hostile targets and the computation of firing data. Such complex equipment, packed into ships that must also have high speed (30 knots and more), excellent seakeeping ability, and long endurance, meant that destroyers and frigates became larger than their World War II predecessors. Guided-missile destroyers ranged from 3,500 to 8,000 tons displacement, while frigates ranged between 1,500 and 4,000 tons.

Amphibians. The diesel-powered Newport LSTs, which entered service in the U.S. Navy in the 1960s, displaced more than 8,000 tons full load and transported amphibious craft, tanks, and other combat vehicles, along with 400 men, at speeds of up to 20 knots. Instead of merely beaching, like their World War II predecessors, they were fitted with an extendable ramp supported by huge projecting derrick extensions on each side of the bow. As the ship grounded, the ramp would shoot forward hydraulically 112 feet. Vehicles and troops would land over the ramp, while amphibious craft in the tank deck would disembark from stern gates.

The Alligator-class LST was a smaller Soviet equivalent of the Newport.

Beginning in the 1970s, gas-turbine power allowed the deployment of air-cushioned landing craft, which were naval hovercraft that could bring tanks and troops to shore at speeds of about 50 knots. (N.F.)

Submarines

A submarine is a naval vessel that is capable of propelling itself beneath the water as well as on the water's surface. This is a unique capability among warships, and submarines are quite different in design and appearance from surface ships. Submarines first became a major factor in naval warfare during World War I (1914–18), when Germany employed them to destroy surface merchant vessels. In such attacks submarines used their primary weapon, a self-propelled underwater missile known as a torpedo. Submarines played a similar role on a larger scale in World War II (1939–45), in both the Atlantic (by Germany) and the Pacific (by the United States). In the 1960s the nuclear-powered submarine, capable of remaining underwater for months at a time and of firing long-range nuclear missiles without surfacing, became an important strategic weapon platform. Armed with torpedoes as well as antiship and antisubmarine missiles, the nuclear attack submarine has also become a key element of naval warfare.

Following is a history of the development of submarines from the 17th century to the present. For a history of other warships, see above *Naval ships and craft.* For the weaponry of modern attack and strategic submarines, see above *Tactical guided missiles* and *Strategic missiles.*

EARLY HAND-POWERED SUBMERSIBLES

The first serious discussion of a "submarine"—a craft designed to be navigated underwater—appeared in 1578 from the pen of William Bourne, a British mathematician and writer on naval subjects. Bourne proposed a completely enclosed boat that could be submerged and rowed underwater. It consisted of a wooden frame covered with waterproof leather; it was to be submerged by reducing its volume by contracting the sides through the use of hand vises. Bourne did not actually construct his boat, and Cornelis Drebbel (or Cornelius van Drebel), a Dutch inventor, is usually credited with building the first submarine. Between 1620 and 1624 he successfully maneuvered his craft at depths of from 12 to 15 feet (four to five metres) beneath the surface during repeated trials in the Thames River, in England. King James I is said to have gone aboard the craft for a short ride. Drebbel's submarine resembled that proposed by Bourne in that its outer hull consisted of greased leather over a wooden frame; oars extended through the sides and, sealed with tight-fitting leather flaps, provided a means of propulsion both on the surface and underwater. Drebbel's first craft was followed by two larger ones built on the same principle.

A number of submarine boats were conceived in the early years of the 18th century. By 1727 no fewer than 14 types had been patented in England alone. In 1747 an unidentified inventor proposed an ingenious method of submerging and returning to the surface: his submarine design had goatskin bags attached to the hull with each skin connected to an aperture in the bottom of the craft. He planned to submerge the vessel by filling the skins with water and to surface by forcing the water out of the skins with a "twisting rod." This arrangement was a forerunner of the modern submarine ballast tank.

The submarine was first used as an offensive weapon in naval warfare during the American Revolution (1775–83). The *Turtle,* a one-man craft invented by David Bushnell, a student at Yale, was built of wood in the shape of a walnut standing on end (see Figure 33). Submerged, the craft was powered by propellers cranked by the operator. The plan was to have the *Turtle* make an underwater approach to a British warship, attach a charge of gunpowder to the ship's hull by a screw device operated from within the craft, and leave before the charge was exploded by a time fuse. In the actual attack, however, the *Turtle* was unable to force the screw through the copper sheathing on the warship's hull.

Bushnell's *Turtle*

Robert Fulton, famed U.S. inventor and artist, experimented with submarines several years before his steamboat *Clermont* steamed up the Hudson River. In 1800, while in France, Fulton built the submarine *Nautilus* under a grant from Napoleon Bonaparte. Completed in May 1801, this craft was made of copper sheets over iron ribs. A collapsing mast and sail were provided for surface propulsion, and a hand-turned propeller drove the boat when submerged. A precursor of a conning tower fitted with a glass-covered porthole permitted observation from within the craft. The *Nautilus* submerged by taking water into ballast tanks, and a horizontal "rudder"—a forerunner of the diving

By courtesy of the U.S. Navy

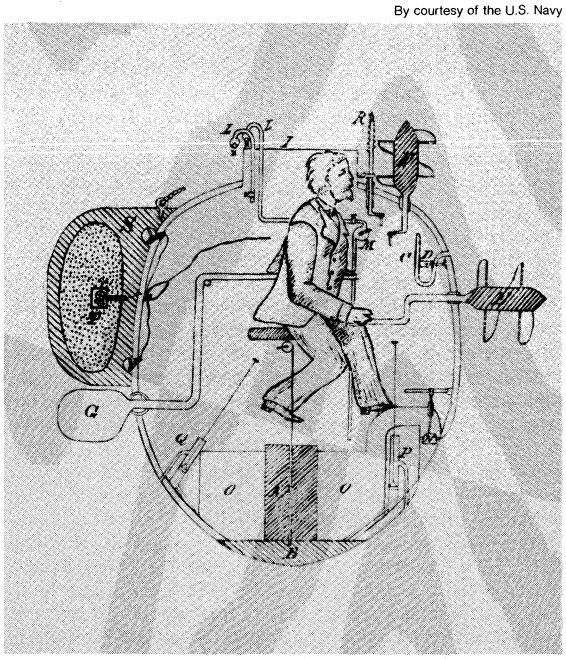


Figure 33: Bushnell's submarine torpedo boat, 1776. Drawing of a cutaway view made by Lieutenant Commander F.M. Barber in 1885 from a description left by Bushnell.

plane—helped keep the craft at the desired depth. The submarine contained enough air to keep four men alive and two candles burning for three hours underwater; later a tank of compressed air was added.

The *Nautilus* was intended to attach an explosive charge to the hull of an enemy ship in much the same manner as the *Turtle*. Fulton experimentally sank an old schooner moored at Brest but, setting out to destroy British warships, was unable to overtake those he sighted. France's interest in Fulton's submarine waned, and he left for England, offering his invention to his former enemy. In 1805 the *Nautilus* sank the brig *Dorothy* in a test, but the Royal Navy would not back his efforts. Fulton then came to the United States and succeeded in obtaining congressional backing for a more ambitious undersea craft. This new submarine was to carry 100 men and be powered by a steam engine. Fulton died before the craft was actually finished, however, and the submarine, named *Mute,* was left to rot, eventually sinking at its moorings.

During the War of 1812 between the United States and England, David Bushnell built another submarine, which attacked HMS *Ramillies* at anchor off New London, Conn. This time the craft's operator succeeded in boring a hole in the ship's copper sheathing, but the screw broke loose as the explosive was being attached to the ship's hull.

The next U.S. attempt at submarine warfare came during the Civil War (1861–65) when the Confederate States resorted to "unconventional" methods to overcome the Union Navy's superior strength, exerted in a blockade of Southern ports. In 1862 Horace L. Hunley of Mobile, Ala., financed the building of a Confederate submarine named *Pioneer,* a craft that was 34 feet long and was driven by a hand-cranked propeller operated by three men. It probably was scuttled to prevent its capture when Union forces occupied New Orleans (although some records say the *Pioneer* was lost with all hands during a dive while en route to attack Union ships).

The second submarine developed by the same builders was a remarkably advanced concept: a 25-foot iron boat intended to be propelled by a battery and electric motors. Not surprisingly, no suitable motors could be found, so a propeller cranked by four men was again adopted. The submarine sank without loss of life in heavy seas off Mobile Bay while seeking to attack the enemy.

Confederate submarines

The third submarine of the Confederacy was the *H.L. Hunley,* a modified iron boiler lengthened to between 36 and 40 feet. Ballast tanks and a system of weights submerged the craft; it could travel at a speed of four miles an hour, powered by eight men cranking its propeller. Its armament consisted of a "torpedo," filled with 90 pounds (40 kilograms) of gunpowder, towed behind the submarine at the end of a 200-foot line. The *Hunley* was to dive under an enemy warship and drag the torpedo against its hull. After a successful test against a barge, the *Hunley* was moved by railroad to Charleston, S.C. There the vessel suffered several disasters, sinking three times and drowning a number of crewmen including Hunley himself. Manned for a fourth time, the *Hunley* was fitted with a "torpedo" on the end of a long spar, and the craft made several successful dives. On the night of Feb. 17, 1864, the submarine attacked the Union warship *Housatonic* in Charleston harbour. The torpedo's detonation exploded the warship's magazines: the *Housatonic* sank in shallow water with the loss of five men, but the *Hunley* was also destroyed by the explosion, and its crew was killed.

One of the more intrepid submarine inventors of the same period was Wilhelm Bauer, a noncommissioned officer of Bavarian artillery who built two boats, *Le Plongeur-Marin* (1851) and *Le Diable-Marin* (1855). The first boat sank in Kiel harbour on Feb. 1, 1851, but Bauer and his two assistants escaped from a depth of 60 feet after the craft had been on the bottom for five hours. His second craft, built for the Russian government, was successful and reportedly made 134 dives before being lost at sea. In September 1856, during the coronation of Tsar Alexander II, Bauer submerged his submarine in Kronshtadt harbour with several musicians on board. An underwater rendition of the Russian national anthem was clearly heard by persons inside ships in the harbour.

TOWARD DIESEL-ELECTRIC POWER

A major limitation of the early submarines was their lack of a suitable means of propulsion. In 1880 an English clergyman, George W. Garrett, successfully operated a submarine with steam from a coal-fired boiler that featured a retractable smokestack. The fire had to be extinguished before the craft would submerge (or it would exhaust the air in the submarine), but enough steam remained in the boilers for traveling several miles underwater.

Similarly, the Swedish gun designer Torsten Nordenfelt constructed a steam-powered submarine driven by twin propellers. His craft could be submerged by vertical propellers to a depth of 50 feet and was fitted with one of the first practical torpedo tubes. Several nations built submarines to Nordenfelt's design.

In an effort to overcome the problems of propulsion, two French naval officers built the 146-foot submarine *Le Plongeur* in 1864, powered by an 80-horsepower compressed-air engine, but the craft quickly exhausted its air tanks whenever it got under way. Development of the electric motor finally made electric propulsion practicable. The submarine *Nautilus,* built in 1886 by two Englishmen, was an all-electric craft. This *Nautilus,* propelled by two 50-horsepower electric motors operated from a 100-cell storage battery, achieved a surface speed of six knots (nautical miles per hour; one knot equals 1.15 statute miles per hour or 1.85 kilometres per hour). But the battery had to be recharged and overhauled at short intervals, and the craft was never able to travel more than 80 miles without a battery recharge. In France, Gustave Zédé launched the *Gymnote* in 1888; it, too, was propelled by an electric motor and was extremely maneuverable but tended to go out of control when it dived.

The end of the 19th century was a period of intensive submarine development, and Zédé collaborated in a number of designs sponsored by the French navy. A most successful French undersea craft of the period was the *Narval,* designed by Maxime Laubeuf, a marine engineer in the navy. Launched in 1899, the *Narval* was a double-hulled craft, 111.5 feet long, propelled on the surface by a steam engine and by electric motors when submerged. The ballast tanks were located between the double hulls, a concept still in use today. The *Narval* made a large number of successful dives. Further French progress in submarines was marked by the four Sirène-class steam-driven undersea craft completed in 1900–01 and the *Aigrette,* completed in 1905, the first diesel-driven submarine of any navy.

The *Narval:* steam–electric propulsion

Similarly, there were submarine successes in the United States by rival inventors John P. Holland (an Irish immigrant) and Simon Lake. Holland launched his first undersea craft in 1875. This one and its successors were significant in combining water ballast with horizontal rudders for diving. In 1895, in competition with Nordenfelt, Holland received an order from the U.S. Navy for a submarine. This was to be the *Plunger,* propelled by steam on the surface and by electricity when submerged. The craft underwent many design changes and finally was abandoned before completion. Holland returned the funds advanced by the navy and built his next submarine (his sixth) at his own expense. This was the *Holland,* a 53.25-foot craft launched in 1897 and accepted by the navy in 1900. For underwater propulsion the *Holland* had an electric motor, and it was propelled on the surface by a gasoline engine. The submarine's armament consisted of a bow torpedo tube, for which three torpedoes were carried, and two dynamite guns. With its nine-man crew the *Holland* was a successful boat; it was modified many times to test different arrangements of propellers, diving planes, rudders, and other equipment.

Holland's chief competitor, Simon Lake, built his first submarine, the *Argonaut I,* in 1894; it was powered by a gasoline engine and electric motor. This and Lake's other early boats were intended as undersea research craft. In 1898 the *Argonaut I* sailed from Norfolk, Va., to New York City under its own power, predating the cruises of the French *Narval* and marking the first time an undersea craft operated extensively in the open sea. Lake's second submarine was the *Protector,* launched in 1901.

Of the major naval powers at the turn of the century,

only Britain remained indifferent toward submarines. Finally, in 1901, the Royal Navy ordered five of the Holland-design undersea craft. Germany completed its first submarine, the *U-1* (for *Unterseeboot 1*), in 1905. This craft was 139 feet long, powered on the surface by a heavy oil engine and by an electric motor when submerged, and was armed with one torpedo tube. Thus, the stage was set for the 20th-century submarine, a craft propelled on the surface by diesel engines and underwater by battery-powered electric motors, submerging by diving planes and taking on water ballast, and armed with torpedoes for sinking enemy ships. The quarters inside these early craft were cramped, generally wet, and stank from diesel oil.

The first U-boat

World War I. By the eve of World War I all of the major navies included submarines in their fleets, but these craft were relatively small, were considered of questionable military value, and generally were intended for coastal operations. The most significant exception to the concept of coastal activity was the German Deutschland class of merchant U-boats, each 315 feet long with two large cargo compartments. These submarines could carry 700 tons of cargo at 12- to 13-knot speeds on the surface and at seven knots submerged. The *Deutschland* itself became the *U-155* when fitted with torpedo tubes and deck guns, and, with seven similar submarines, it served in a combat role during the latter stages of the war. In comparison, the "standard" submarine of World War I measured slightly over 200 feet in length and displaced less than 1,000 tons on the surface.

The prewar submarines generally had been armed with self-propelled torpedoes for attacking enemy ships. During the war submarines also were fitted with deck guns. This permitted them to approach enemy merchant ships on the surface and signal them to stop for searching (an early war policy) and later to sink small or unarmed ships that did not warrant expenditure of torpedoes. Most war-built submarines had one and sometimes two guns of about three- or four-inch calibre; however, several later German submarines carried 150-millimetre guns (including the Deutschland class in military configuration).

An important armament variation was the submarine modified to lay mines during covert missions off an enemy's harbours. The Germans constructed several specialized submarines with vertical mine tubes through their hulls; some U-boats carried 48 mines in addition to their torpedoes.

Also noteworthy was the development, during the war, of the concept of an antisubmarine submarine. British submarines sank 17 German U-boats during the conflict; the early submarine-versus-submarine successes led to British development of the R-class submarine intended specifically for this role. These were relatively small craft, 163 feet long and displacing 410 tons on the surface, with only one propeller (most contemporary submarines had two). Diesel engines could drive them at nine knots on the surface, but once submerged, large batteries permitted their electric motors to drive them underwater at the high speed of 15 knots for two hours. (Ten knots was a common speed for submerged submarines until after World War II.) Thus, they were both maneuverable and fast. Advanced underwater listening equipment (asdic, or sonar) was installed, and six forward torpedo tubes made them potent weapons. Although these submarines appeared too late to have any actual effect on the war, they pioneered a new concept in the development of the submarine.

Anti-submarine submarines

All World War I-era submarines were propelled by diesels on the surface and by electric motors submerged, except for the British Swordfish and K class. These submarines, intended to operate as scouts for surface warships, required the high speeds then available only from steam turbines. The K-boats steamed at 23.5 knots on the surface, while electric motors gave them a 10-knot submerged speed.

World War II. Interest in submarines continued high within the world's navies during the period between World Wars I and II. Britain, France, and Japan built improved types, and during this period the U.S. Navy built its first large long-range submarine, the *Argonaut.* Completed in 1928, it was 381 feet long, displaced 2,710 tons on the surface, was armed with two six-inch guns and four forward torpedo tubes, and could carry 60 mines. The *Argonaut,* the largest nonnuclear submarine ever built by the U.S. Navy, led to the highly successful Gato and Balao classes of U.S. submarines used in World War II.

During the 1930s the rejuvenated Soviet shipyards began producing large numbers of submarines, primarily coastal craft, in an attempt to make the Soviet Union a sea power without major expenditures for surface warships. But though the Soviet program achieved quantity, their ships were unsuitable for operations against the German Navy, their crews were poorly trained, and Soviet bases were blocked by ice much of the time.

From J.P. Mallmann Showell, *U-Boats under the Swastika* (1987)

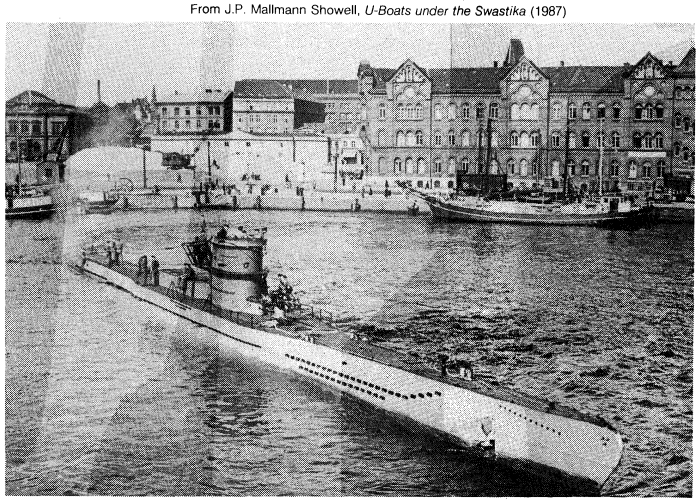


Figure 34: *Launching of U-218 at Kiel, Ger., in 1941.* With the exception of mine-laying shafts just behind the conning tower, this Type VIID submarine was essentially identical to the Type VIIC, which hunted Allied ships in the Atlantic during World War II.

World War II saw extensive submarine campaigns on all of the world's oceans. In the Atlantic the principal German U-boat was the VII type, a relatively small but effective craft when properly employed. The Type VIIC variant was 220.25 feet long, displaced 769 tons on the surface, and was powered by diesel-electric machinery at a speed of 17 knots on the surface and 7.5 knots submerged. Armament consisted of one 90-millimetre deck gun, various antiaircraft guns, and five torpedo tubes, four forward and one aft. Either 14 torpedoes or 14 tube-launched mines were carried. Manned by a crew of 44, these submarines had a surface endurance of 6,500 miles at 12 knots, but, when they were submerged, their batteries would remain active a little less than a day at four knots.

Type VII U-boats

The ultimate diesel-electric submarine evolved in the war was the German Type XXI, a 250-foot, 1,600-ton craft that could attain 17½ knots submerged for more than an hour, could travel at six knots underwater for two days, or could "creep" at slower speeds for four days. These submarines were fitted with snorkel devices (see below), which made it unnecessary for them to surface fully to recharge their batteries after operating submerged. The Type XXI had an operating depth of 850 feet, more than twice what was then normal, and was armed with four 33-millimetre guns and six forward torpedo tubes (23 torpedoes carried). These properties made all earlier submarines obsolete. Existing Allied antisubmarine forces would have had serious trouble coping with these craft had the war continued past the spring of 1945.

A final German war design of particular interest was the Walter turbine propulsion plant. The need for oxygen for combustion had previously prevented the use of steam turbines or diesels while the submarine was submerged and air was at a premium. Hellmuth Walter, a German scientist, developed a turbine propulsion system using oxygen generated by hydrogen peroxide to operate the turbine while submerged. A simplified submarine, the *V-80,* built in 1940 and propelled by a Walter turbine system, could attain speeds of more than 26 knots submerged for a short period of time. After many delays, the first Walter-

propelled Type XVII combat submarines were completed and could reach 25 knots underwater for brief periods, and a submerged run of 20 knots for 5½ hours was achieved on trials. But these submarines, like the Type XXI, were not ready for full-scale operations when the war ended.

The snorkel

A notable German submarine development of World War II was the *schnorchel* device (anglicized by the U.S. Navy to "snorkel"). Its invention is credited to a Dutch officer, Lieutenant Jan J. Wichers, who in 1933 advanced the idea of a breathing tube to supply fresh air to a submarine's diesel engines while it was running submerged. The Netherlands Navy began using snorkels in 1936, and some fell into German hands in 1940. With the advent of radar to detect surfaced submarines, the Germans fitted hundreds of U-boats with snorkels to permit the operation of diesels at periscope depth (to recharge batteries for underwater propulsion) with less of a possibility of detection by Allied radar-equipped ships and aircraft.

In the Pacific war the Japanese employed a large number of submarines of various sizes and types, including aircraft-carrying submarines, midget submarines, and "human torpedoes" carried by larger submarines. The Japanese I-201 class was a high-speed submarine, of 259 feet and 1,291 tons displacement, that had diesel propulsion for 15 knots on the surface; while underwater, large batteries and electric motors could drive the vessel at a speed of 19 knots for almost one hour. Each boat had two 25-millimetre guns and four forward torpedo tubes and carried ten torpedoes.

U.S. submarines

The highly successful U.S. submarine campaign in the Pacific war was waged mainly with the Gato- and Balao-class submarines. These were approximately 311.5 feet long, displaced 1,525 tons, and had diesel-electric machinery for 20-knot surface and nine-knot underwater speeds. The principal difference between the two designs was the 300-foot operating depth for the Gato class and 400-foot depth for the Balao boats. Manned by 65 to 70, these submarines had one or two five-inch deck guns plus smaller antiaircraft weapons and 10 torpedo tubes (six forward, four aft) and carried 24 torpedoes.

Postwar developments. After the war the Allies were quick to adopt advanced German submarine technology. The British built two peroxide turbine-propelled experimental submarines, but this concept lost favour because of the unstable properties of hydrogen peroxide and because of American success with nuclear propulsion. The Soviet Union began building modifications of the Type XXI submarine. Some 265 of these submarines, labeled Whiskey and Zulu class by NATO observers, were completed between 1950 and 1958, more submarines than built by all of the world's other navies combined between 1945 and 1970. (In that period Soviet shipyards produced a total of 560 new submarines.)

The U.S. Navy studied German technology and converted 52 war-built submarines to the Guppy configuration (an acronym for greater underwater propulsive power with the "y" added for phonetics). These submarines had their deck guns removed and streamlined conning towers fitted; larger batteries and a snorkel were installed; four torpedoes and, in some craft, one of the four diesel engines were removed. The result was an underwater speed of 15 knots and increased underwater endurance. (N.C.P./Ed.)

Although the major powers switched to nuclear power after World War II, the great bulk of the world's navies continued to buy—or in a few cases, build—submarines descended directly from the fast diesel-electric U-boats of the war. (Indeed, many of them were designed and built in West Germany.) The main advances were in weapons and sensors. Deck guns were abandoned, in some cases for antiship missiles. Torpedoes, which soon could exceed 50 knots, either homed onto their targets acoustically with self-contained sonar or were guided by electronic commands passed to them through a threadlike wire paid out behind the speeding projectile. Submarine sonars, for detecting both surface ships and other submarines, were enormously improved.

Improved underwater endurance

On the other hand, maximum submerged speed increased only somewhat (to more than 20 knots) over the German Type XXI, and endurance at top speed was no greater than at the end of World War II. Improvements in battery design concentrated more on increasing endurance at low speed. Some modern submarines, for example, could remain submerged (at about three knots) for as long as a week to 10 days. That was an important improvement, since during so long a period sea conditions could easily arise that would allow a submarine to escape or force submarine hunters on the surface to disperse.

Postwar diesel-electric submarines continued to be equipped with snorkels, but hunters adopted improved radars that could detect even the small head of the snorkel, just as aircraft with more primitive radars could detect surfaced U-boats during World War II.

For these reasons, diesel-electric submarines remained furtive platforms, conserving their energy for the postattack escape. Because their electric motors were quieter than nuclear units (and could even be shut off for a time), they were sometimes proposed as antisubmarine ambushers that would silently await their prey in areas through which enemy submarines were known to pass. Britain's Upholder class, 230 feet long and displacing about 2,000 tons at the surface, was built in the 1980s to operate in this capacity alongside nuclear vessels.

NUCLEAR PROPULSION

In 1954, with the commissioning of USS *Nautilus,* nuclear power became available. Since the nuclear reactor needed no oxygen at all, a single power plant could now suffice for both surface and submerged operation. Moreover, since a very small quantity of nuclear fuel (enriched uranium) provided power over a very long period, a nuclear submarine could operate completely submerged at high speed indefinitely.

This change was revolutionary. In the typical prenuclear submarine attack, the submarine approached the target on the surface to avoid draining the battery and submerged only just before coming within sight of the target. The submerged approach had to be made at very low speed, perhaps no more than two or three knots, again to avoid wasting battery power. The submarine commander had to husband his battery charge until after the attack, when he would have to use full underwater power (and a speed of perhaps seven to 10 knots) to evade the counterattack. Even then, a full battery charge would last only about one or two hours at top speed. This necessity of conserving battery power, which forced diesel-electric submarines to approach their targets as quietly and slowly as possible, meant that they could not engage most fast surface warships, such as aircraft carriers and battleships.

Operating freely underwater

Nuclear submarines were in an altogether different class. Not only could they evade freely (that is, at top speed for indefinite periods) after attacking, they could also operate freely before attacking and keep up with fast surface ships. This principle was illustrated by the only instance of a nuclear submarine's firing of a weapon in anger. During the Falkland Islands conflict in 1982, a British nuclear submarine, HMS *Conqueror,* followed the fast Argentine cruiser *General Belgrano* for more than 48 hours before closing in to sink it. That performance would have been entirely beyond the capability of any prenuclear submarine. For the first time, a submarine commander could maneuver freely underwater, without worrying that he was exhausting his vessel's batteries, and fast surface warships were vulnerable to submarine attack.

Initially, the major powers continued to build diesel-electric submarines alongside nuclear vessels, but some later gave in to the expense of maintaining two categories of submarine in parallel. After 1959 the U.S. Navy effectively ceased construction of nonnuclear submarines. The Royal Navy, which completed its first nuclear submarine, HMS *Dreadnought,* in 1963, followed a similar policy until the 1980s, when it began the Upholder class. France completed its first nuclear submarine, *Le Redoutable,* in 1971 and effectively abandoned diesel-electric construction in 1976. Although the Soviets continued to build diesel submarines, the bulk of their new construction shifted to nuclear power after their first nuclear submarines, of the November class, entered service in 1958. Beginning in 1968, the Chinese built a few nuclear submarines while continuing to build large numbers of nonnuclear submarines.

Nuclear power plants. *Reactors.* A nuclear reactor provides the heat that powers a steam turbine, which in turn drives a propeller. There are three main types of marine nuclear reactor: pressurized-water, natural-circulation, and liquid-metal.

Generally, uranium in a reactor produces heat by nuclear fission. In the reactor, the uranium is surrounded by a moderator, which is required to slow the reaction neutrons so that they will interact more efficiently with the uranium. In most reactors the moderator is water, which is also used to carry away the heat of reaction. This heated water is called the primary loop water. Pressurized to prevent it from boiling, it runs through a heat exchanger, in which the heat is passed to another, secondary, water circuit. The heat exchanger is essentially a boiler, and the secondary circuit, or loop, provides the steam that actually turns the turbine. So long as a sufficient seal is maintained, the water of the primary loop cannot contaminate the rest of the power plant.

Pressurized and natural-circulation reactors

In most cases the water in the primary loop is circulated by pump. Reactors can also be arranged so that differences in temperature—for example, between that portion of the reactor containing the reacting fuel and the rest of the reactor—force the water to circulate naturally. Typically, in these natural-circulation reactors cooled water from the heat exchanger is fed into the bottom of the reactor, and it rises through the fuel elements as they heat it.

The liquid-metal-cooled reactor operates on the principle that molten metal can carry much more heat than water, so that a more compact turbine can be used. Against that advantage, molten metal can be made highly radioactive, so that leaks, which are dangerous enough in a pressurized-water plant, become much more so. Second, pumps in these reactors must be much more powerful, and the simplicity of using the same substance as moderator and heat sink is lost. Finally, there is always the possibility that enough heat will be lost for the plant to seize up, the metal solidifying in the pipes, with catastrophic results.

Adoption by navies. Under the direction of Captain (later Admiral) Hyman Rickover, the U.S. Navy developed both pressurized-water and liquid-metal prototypes. It completed its first two nuclear submarines, the *Nautilus* and *Seawolf,* to test the two types, but problems (including leakage) in the *Seawolf* reactor led to the abandonment of the liquid-metal scheme. Later the navy also developed natural-circulation reactors. U.S. attack submarines (except for USS *Narwhal,* the natural-circulation prototype) were built with pressurized-water reactors, but the Ohio-class strategic submarines of the 1980s were powered by natural-circulation reactors. The latter were inherently quieter than pressurized-water units because they required no pumps, at least at low and moderate power.

Soon after the *Nautilus* was completed, the Royal Navy adopted American-type pressurized-water reactors, but by the early 1980s it was developing natural-circulation reactors of its own design.

The first generation of French nuclear power plants, built for strategic submarines, were pressurized-water types. French attack submarines, on the other hand, were powered by natural-circulation units, at least at low and moderate speeds.

The Soviet Union revealed little about its nuclear submarine program, but most of its plants were apparently of the pressurized-water type. However, its most compact and powerful plant, in the very fast Alfa-class attack submarines of the 1970s and '80s, was thought to use liquid metal.

The nuclear navies. The advent of the new nuclear submarines had two great consequences. One was the rise of an altogether new kind of submarine, the strategic submarine. The other was a revolution in antisubmarine warfare, with attack submarines becoming the primary antisubmarine weapons. Attack submarines were armed with torpedoes and, in some cases, with antiship missiles. Strategic submarines carried similar weapons, but their primary weapons were submarine-launched ballistic missiles (SLBMs), such as the U.S. Polaris, Poseidon, and Trident and the Soviet SS-N-6 Sawfly, SS-N-18 Stingray, and SS-N-20 Sturgeon.

Strategic submarines. Strategic submarines were valuable because they were so difficult to find and kill, and they became even more important as long-range SLBMs became more accurate. Accurate missiles could destroy missiles in fixed land sites; were all strategic missiles so based, the side firing first could hope to disarm its enemy. However, with many missiles now based at sea, such a first strike became virtually impossible—barring some breakthrough in submarine detection. To the extent that preemptive attack was impractical, therefore, a force of strategic submarines became an effective deterrent against enemy attack. For this reason, the United States, the Soviet Union, Great Britain, France, and China all built submarines armed with SLBMs.

Strategic submarines actually predated the nuclear-propulsion era, in that during the 1950s both the U.S. and Soviet navies developed missile-carrying diesel-electric submarines. The U.S. submarines were armed with Regulus cruise missiles, and the Soviet ships carried SS-N-3 Shaddock cruise missiles and SS-N-4 Sark short-range SLBMs. However, these missiles had to be launched from the surface, and the submarines themselves could not remain submerged indefinitely. Strategic submarines did not become truly effective until nuclear power plants and dive-launched missiles enabled them to operate continuously without exposing themselves on the surface in any way.

The growing size and missile range of strategic submarines

The first modern strategic submarines were of the U.S. George Washington class, which became operational in 1959. These 5,900-ton, 382-foot vessels carried 16 Polaris missiles, which had a range of 1,200 nautical miles. In 1967 the first of the Soviet Union's 8,000-ton Yankee-class submarines were delivered, which carried 16 SS-N-6 missiles of 1,300-nautical-mile range. These were followed a decade later by Delta-class vessels fitted with 16 SS-N-18 missiles. Each SS-N-18 had a range of 3,500 nautical miles. In 1982 the Soviet Union began to deploy its Typhoon class, of 18,500 tons displacement at the surface and about 560 feet in length. These submarines carried 20 SS-N-20 SLBMs, each of which could carry its warheads a distance of 4,500 nautical miles.

Beginning in 1970, the United States fitted its Lafayette-class submarines with 16 Poseidon SLBMs, which could launch its warheads a distance of 2,500 nautical miles. To carry 24 Trident missiles, improved versions of which could travel about 6,500 nautical miles, the U.S. Navy commissioned the first Ohio-class submarine in 1981 (see Figure 35). These vessels displaced 16,600 tons at the surface and were about as long as the Soviet Typhoons.

Britain's first strategic submarines, of the Resolution class, entered service in 1967 with 16 Polaris missiles. The

U.S. Navy photo by PH1 Dale L. Anderson

Figure 35: *USS Ohio, strategic nuclear submarine of the U.S. Navy.* Commissioned in 1981, it carries 24 Trident ballistic missiles in a double row of vertical launch tubes (shown with hatches open). The average patrol time at sea of Ohio-class submarines is 70 days, and their nuclear reactor cores need replacement only once every nine years.

first Vanguard-class vessel was laid down in 1986 to carry 16 Trident missiles.

To supplement the Redoutable class of the 1970s, France built *L'Inflexible.* This 8,000-ton submarine, which entered service in 1985, carried 16 M-4 SLBMs, each with a range of 2,800 nautical miles. In 1988 the first of the Triomphant class was laid down; as replacements of the Redoutable class, these were designed to carry SLBMs of 6,000-nautical-mile range.

In 1981 China launched the fist Xia-class strategic submarine. It was armed with 12 CSS-N-3 missiles with a range of 1,500 nautical miles.

Attack submarines. After the rise of nuclear-powered strategic submarines, it seemed that only other nuclear submarines could maneuver in three dimensions and remain in contact long enough to destroy them. Surface ships were clearly handicapped because their sonars could not operate as freely as those of a submarine. That situation changed somewhat when surface warships began to tow passive sonar arrays at submarine-like depths and when ship- or helicopter-launched homing torpedoes acquired a fair chance of holding and killing their targets. Both submarines and surface ships, therefore, became effective antisubmarine weapons, but only submarines could operate near an enemy's bases, where hostile submarines would be easier to find, and only they could lie in ambush with little chance of being detected.

Most modern nuclear attack submarines adopted a dual function: to attack enemy surface ships and to destroy enemy submarines. For example, the U.S. Sturgeon submarines of the 1960s and '70s and their successors, the Los Angeles class, carried torpedoes and rocket-launched nuclear depth bombs for antisubmarine warfare as well as underwater-launched Harpoon missiles for attacking surface ships from as far away as 70 nautical miles. From 1984 both classes were fitted with Tomahawk cruise missiles, which could be programmed to strike ships 250 nautical miles away or, in a strategic variant, to hit land targets with a nuclear warhead at ranges up to 1,300 nautical miles.

Soviet cruise-missile submarines

The Soviets tended to divide their attack submarines between antisubmarine and cruise-missile duties. Beginning in 1971, the SS-N-7 Star Bright cruise missile, which could be launched underwater and could strike ships 35 nautical miles away, was deployed in Charlie-class submarines. The SS-N-7 began a series of dive-launched antiship cruise missiles of increasing range, culminating in the SS-N-19 Shipwreck, a supersonic missile that could carry a nuclear warhead 340 nautical miles. Twenty-four of these weapons were carried aboard the 13,000-ton, 500-foot Oscar submarines, which entered service in 1980.

The most prominent submarine-hunting submarines of the Soviet Union were of the three Victor classes. The Victor I vessels, which entered service beginning in 1968, introduced the "tear-drop" hull configuration to the underwater Soviet Navy. These and the 6,000-ton Victor II and III classes of the following decades were fitted with rocket-launched torpedoes or nuclear depth bombs, giving them a battle range extending to 50 nautical miles.

Adding a further role to Soviet attack submarines after 1987 was the SS-N-21 Sampson cruise missile, a weapon with a nuclear capability and range similar to those of the U.S. Tomahawk.

The British Valiant, Swiftsure, and Trafalgar classes of the 1960s, '70s, and '80s displaced between 4,000 and 4,500 tons at the surface and were about 285 feet long. They were armed with torpedoes and dive-launched Harpoon missiles. In France, the first Rubis-class submarine was laid down in 1976 with torpedo and sonar systems inherited from the diesel-electric Agosta class. Beginning in 1984, new and existing vessels of this class were given improved sonar and silencing and were fitted with dive-launched Exocet antiship missiles. They displaced about 2,400 tons at the surface and were about 235 feet long.

Design principles. After 1955 the three main trends in nuclear attack submarine design in the United States, Britain, and the Soviet Union were increased speed, increased diving depth, and silencing.

Speed. Increased speed required increased power. Since the resistance a submarine encounters is a function of its surface area, the ideal was to achieve greater power without increasing the volume or weight of the power plant and, therefore, the size of the submarine. A more powerful (and therefore noisier) engine could be silenced, but only by increasing the size of the submarine, which in turn would lower its speed. These complex trade-offs were illustrated by the Sturgeon and Los Angeles submarines. Reactor power approximately doubled between these two generations, but overall size increased enormously, from about 3,600 to 6,000 tons surfaced. The Soviets, meanwhile, achieved very high speed (about 40 knots, compared to slightly over 30 knots for fast Western submarines) in their Alfa class, but probably at the cost of a great deal of noise at high speed.

Speed was prized for several quite different reasons. At first, the U.S. and Soviet navies developed fast submarines primarily as antiship weapons. In the 1950s the Guppy-style hull design of USS *Nautilus* gave it a submerged speed of over 20 knots, which was fast enough to evade surface ships but not to counterattack them. To make up this deficit, U.S. submarines then under design were altered by adapting nuclear power to the tapered "tear-drop" hull of the experimental submarine *Albacore.* The resulting Skipjack class, which entered service in 1959, came up with a top speed in excess of 30 knots.

The "tear drop" hull

In a spectacular demonstration of the Soviets' fast attack capabilities, a Soviet nuclear submarine intercepted the nuclear aircraft carrier USS *Enterprise* in February 1968. The submarine was not quite as fast as the *Enterprise,* but it was fast enough to keep the carrier within weapon range while the carrier accelerated to top speed.

With the commencement of the Soviet fast nuclear program, the U.S. Navy shifted its emphasis to dual-purpose vessels capable of attacking submarines as well as surface ships. High speed, as achieved in the 1970s and '80s by the Los Angeles class, was then required to keep up with the fast surface targets that the Soviet submarines were expected to attack.

High sustained speed also made it possible for submarines to deploy more efficiently to distant patrol stations. Although nuclear submarines' fuel supplies were effectively unlimited, they were limited in their capacity for stores and could not expect to remain at sea for more than about 60 to 90 days. The more rapidly they could reach their patrol area, therefore, the more productive time they could spend there.

As in the case of nonnuclear submarines, higher speed was also valued for evasion after an attack. However, when that higher speed was bought at the cost of louder operation, submarines became easier to detect. Also, from the mid-1950s the main antisubmarine weapons were homing torpedoes, which became significantly faster than the submarines they sought, and nuclear depth bombs, which might be dropped effectively anywhere in the vicinity of a submarine. In all of these cases, sheer speed was no longer a guarantee of evasion, although it did make attack more difficult.

Depth. Deeper diving was valued for several reasons. As in the past, it could be combined with higher speed for better evasion. In addition, a deep-diving submarine could make better use of its own sonar, partly because it could operate in several quite different layers of the sea. This advantage was reflected in a change in U.S. submarine sonars that began about 1960. Previous submarine units had been cylindrical, producing broad, fan-shaped beams that could determine target range and bearing but not target depth. The new sonars were spherical, producing narrow, pencil-shaped beams that could distinguish between targets at different depths. They could also make better use of sonar reflection off the sea bottom and surface to achieve greater range.

Finally, greater maximum operating depth became particularly important at high speed, when there was always a possibility that a submarine would accidentally tip down and descend below a safe operating depth before the downward motion could be corrected. It is no surprise, then, that the greatest reported diving depth (about 2,800 feet) was associated with the highest reported maximum speed

Diving depth of attack submarines

(about 43 knots), in the Soviet Alfa class. (Diving depth of most other modern attack submarines was reportedly between 1,000 and 1,500 feet.)

Greater depth required a stronger (and heavier) hull, and increased power required a stronger power plant. Attempts to combine the two required a larger hull (to provide enough buoyancy); that in turn added underwater resistance, which cut the speed advantage gained from the more powerful engine. This tension between different requirements explains the characteristics of many modern submarines. For example, the Los Angeles class was said to have sacrificed some diving depth in order to achieve higher speed. In the Alfa class, weight was saved by adopting an expensive titanium-alloy hull and a very compact power plant.

Silencing. Until the late 1950s, submarines were usually detected by active sonar; that is, by sound waves bounced off their hulls. Because these sound waves could also be detected by the hunted submarine, they gave it warning that it was in danger of attack. Also, because water can support only so much sonar energy, active sonar was limited in range. Beginning in the early 1950s, then, the U.S. and British navies began to investigate passive sonar, in which sensors detected noises emanating from the submarine itself. Early nuclear submarines were quite susceptible to such detection because their machinery was very noisy. In particular, the pumps required to circulate the coolant, which could not be turned off without melting the reactor core, could be heard at a considerable distance.

Beginning at that time, silencing became a major thrust in submarine design. The pumps of pressurized-water reactors were redesigned to be quieter, and in many submarines the machinery was carried clear of the hull on sound-absorbing mounts. All of this added to the size and weight of the machinery and to the expense of construction; it also added to the attraction of natural-circulation plants.

As a further step in silencing, hulls were coated with sound-absorbing material. Even relatively simple coatings could drastically reduce the effectiveness of homing torpedoes. (N.F.)

Military aircraft

Aircraft have been a fundamental part of military power since the mid-20th century. Generally speaking, all military aircraft fall into one of the following categories: fighters, which secure control of essential airspaces by driving off or destroying enemy aircraft; bombers, which are larger, heavier, and less maneuverable craft designed to attack surface targets with bombs or missiles; ground-support, or attack, aircraft, which operate at lower altitudes than bombers and air-superiority fighters and attack tanks, troop formations, and other ground targets; transport and cargo planes, big-bodied craft with large amounts of interior space for carrying weapons, equipment, supplies, and troops over moderate or long distances; and helicopters, which are rotary-winged aircraft used for ground support, to transport assault troops, and for short-distance transport and surveillance.

EARLY HISTORY

When the first practical aircraft were produced, in the form of hot-air and hydrogen balloons in 1783, they were adopted quickly for military duties. In 1793 the French Convention authorized formation of a military tethered-balloon organization, and a company of "Aérostiers" was formed on April 2, 1794. Two months later the first military reconnaissance from such a balloon was made before the city of Maubeuge. Until the Aérostiers were disbanded in 1799, their reports contributed to the success of French armies in many battles and sieges. Similar reconnaissance balloons were used later by other armies, notably by both armies during the American Civil War and by the British in Africa from 1884 to 1901.

The beginning of military aviation

True military aviation began with the perfection of the navigable airship in the late 19th century and the airplane in the first decade of the 20th century. The brothers Wilbur and Orville Wright, who made the first powered, sustained, and controlled flights in an airplane on Dec. 17, 1903, believed such an aircraft would be useful mainly for military reconnaissance. When they received the first contract for a military airplane from the U.S. government in February 1908, it called for an aircraft capable of carrying two persons at a speed of at least 40 miles (64 kilometres) per hour for a distance of 125 miles. The aircraft they delivered in June 1909 was listed as "Airplane No. 1, Heavier-than-air Division, United States aerial fleet."

The most formidable aircraft of the years before World War I were airships rather than airplanes. Airships were large, self-propelled craft usually consisting of a rigid, fabric-covered metal frame within which were gas bags containing a lighter-than-air gas such as hydrogen. The most ambitious examples of this type of craft were the huge airships designed and built in Germany by Ferdinand, Count von Zeppelin. A typical zeppelin could carry five 110-pound (50-kilogram) high-explosive bombs and 20 6.5-pound incendiary bombs at a time when most military airplanes were without any form of weapons, being intended only for reconnaissance.

Experiments with arming airplanes were made spasmodically after 1910, when August Euler took out a German patent on a machine-gun installation. Bombing techniques evolved simultaneously. Dummy bombs were dropped on a target in the form of a ship by the American designer Glenn Curtiss on June 30, 1910. This test was followed by the dropping of a real bomb and the devising of the first bombsight. In England the Royal Flying Corps (RFC) fitted some of its aircraft with bomb carriers, which consisted of a kind of pipe rack beside the observer's cockpit in which small bombs were retained by a pin. The pin was pulled out over the target by tugging on a string. It was primitive but it worked. The Naval Wing of the RFC subsequently attempted to drop torpedoes from Short and Sopwith seaplanes, with some success, and efforts were soon under way to develop means to launch and recover such craft on shipboard. In 1910–11 a Curtiss biplane had been flown from and onto wooden platforms erected over the decks of anchored U.S. Navy cruisers, and in May 1912 a pilot of the Naval Wing, RFC, flew a Short S.27 biplane from HMS *Hibernia* while the ship was steaming at 10.5 knots. The following year the old cruiser *Hermes* was fitted with a short deck from which seaplanes took off on wheeled trolleys that were fitted under their floats and dropped away as the machines became airborne.

Thus, by 1914, reconnaissance, bomber, and carrier-based aircraft all were evolving, and some had been used in combat. The first use of an airplane in war was on Oct. 23, 1911, during the Italo-Turkish War, when an Italian pilot made a one-hour reconnaissance flight over enemy positions near Tripoli, Libya, in a Blériot XI monoplane. The first bombing raid came nine days later, when a pilot dropped four grenades on Turkish positions. The first reconnaissance photographs of enemy positions were taken on Feb. 24–25, 1912, in the same conflict.

The first bombing raid

WORLD WAR I

Airships. At the start of the war the German armed forces had 10 zeppelins and three smaller airships, but this impressive offensive capability was largely offset by the highly explosive nature of the hydrogen gas that gave the zeppelins their lifting power. After losing three zeppelins in daylight raids over heavily defended areas in the first month of the war, the army abandoned airship operations, but the navy, with its battle fleet blockaded in port by the Royal Navy, mounted a night bombing offensive—the first aerial strategic bombardment campaign in history.

The finest of the zeppelins was the LZ-70; this craft was 740 feet (220 metres) long, was able to fly above 16,000 feet, and had a range of 7,500 miles. The LZ-70 was shot down late in the war, however, and large rigid (metal-framed) airships were never again employed as combat aircraft. Smaller, nonrigid airships were used throughout World War I by the British for antisubmarine patrol, convoy escort, and coastal reconnaissance, achieving a remarkable record of protecting coastal convoys from German submarines. They were revived by the U.S. Navy during World War II for the same use.

Unpowered, captive balloons also were used extensively for observation and artillery spotting in World War I, but by World War II they had become so vulnerable that they were used only as unmanned antiaircraft barrage balloons. Anchored to the ground or ships by cables, they compelled attacking enemy aircraft to fly high to avoid the cables; they also brought down many German pilotless V-1 "buzz bombs" over England in 1944–45.

Reconnaissance aircraft. At the outbreak of World War I, heavier-than-air craft were used only for visual reconnaissance, since their feeble engines could carry little more than a pilot and, in some cases, an observer aloft. They soon proved their worth in this mission, however, and RFC aviators provided reconnaissance that enabled the British and French armies to counterattack in the decisive Battle of the Marne on Sept. 6–12, 1914, turning back the invading Germans just short of Paris.

More powerful engines and better aircraft designs soon made possible specialized reconnaissance aircraft that could fly at high altitudes to avoid interception. The Germans, for example, had Rumpler two-seaters in service by 1917 that could operate as high as 24,000 feet. Radios were carried aloft to permit aerial observers to spot and adjust artillery fire, at first with transmitters only and then, as radios became lighter, with receivers for two-way communication.

Fighters. The importance of aerial reconnaissance and artillery spotting (particularly the latter) made it clear that the belligerent able to deny the enemy use of airspaces above the battlefield would enjoy enormous advantages. This realization led to the emergence of fighters as a distinct category of aircraft. In the early days of the war, pilots and observers blazed away at enemy aircraft with pistols, rifles, and even shotguns, but to little effect. Machine guns were the obvious solution. In 1913 the Vickers company in Britain had exhibited a two-seat biplane of pusher configuration (*i.e.,* with the propeller behind the engine) that was armed with a machine gun fired by an observer who sat ahead of the pilot in a tublike crew compartment. A development of this machine, the Vickers F.B.5 Gunbus, entered service in early 1915 as the first production aircraft designed from the outset with air-to-air armament. The French armed similarly configured Voisin pushers with machine guns (one had shot down a German aircraft as early as Oct. 5, 1914), but, burdened with the extra weight of observer and gun, such aircraft were slow and unmaneuverable, and their successes were mostly the result of accidental encounters. Light, single-seat aircraft of tractor configuration (*i.e.,* with the propeller at the nose) had much better performance, but efforts to arm them with machine guns firing at an angle to avoid hitting the propeller produced little success.

The first true fighter aircraft

The solution to the problem emerged in the spring of 1915 in the form of an interrupter gear, or gun-synchronizing device, designed by the French engineer Raymond Saulnier. This regulated a machine gun's fire so as to enable the bullets to pass between the blades of the spinning propeller. The interrupter itself was not new: a German patent had been taken out on such a device by the Swiss engineer Franz Schneider before the war. The real breakthrough was made by Roland Garros, a famous sporting pilot before the war and a friend of Saulnier, who perceived that a machine gun fitted with such a device and mounted rigidly atop the fuselage could be aimed accurately simply by pointing the airplane in the desired direction. Though the French machine gun had a tendency to "hang fire," so that steel deflector plates had to be fitted onto the rear of the propeller blades to prevent their being shot off, Saulnier quickly perfected his device and fitted it to Garros's Morane L monoplane. With this machine, Garros shot down three German aircraft on April 1, 13, and 18. Then, on April 19, Garros himself force-landed with a ruptured fuel line and was taken prisoner. His efforts to burn his aircraft failed, and the secrets of Saulnier's interrupter gear were laid bare. The Germans reacted quickly, putting the designer Anthony Fokker to work on a similar device. With Saulnier's gear as his inspiration (and perhaps drawing on earlier German work), Fokker swiftly came up with an efficient interrupter gear, which he fitted onto a monoplane of his own design—ironically, a copy of a French Morane. The result was the Fokker Eindecker ("monoplane"; see Figure 36), which entered service in July 1915 and reigned supreme in the air over the Western Front until the following October—a period known among Allied aviators as the "Fokker Scourge."

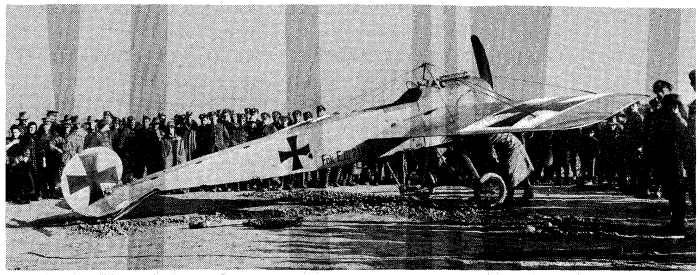


Figure 36: Fokker Eindecker, German fighter plane of World War I.
By courtesy of John W.R. Taylor

The Fokker Eindecker

The Eindecker's mastery was ended by new versions of the French Nieuport with a machine gun mounted above the top wing, allowing it to fire clear of the propeller arc, and by British D.H.2 and F.E.2b pushers with nose-mounted guns. Though a superb flying machine, the Nieuport was limited by its light armament, while the two British machines had brought the aerodynamically inefficient pusher configuration to its limit and were soon outclassed. Thereafter, the pace of fighter development began to be set by improvements in engine design—a phenomenon that was to persist well into the jet age.

Most Allied fighters at that time were powered by rotary radial engines (*i.e.,* with the cylinders, arranged radially about the crankcase like the spokes of a wheel, rotating around a stationary crankshaft). These engines were relatively powerful in relation to their weight, but their large frontal areas produced a great deal of drag, and the gyroscopic forces induced by their whirling mass posed serious aircraft control problems. In mid-1916 Germany took the lead in fighter design on the basis of its superb Daimler and Benz water-cooled, in-line engines, such as those which powered the streamlined Albatros D.I, D.II, and D.III series of fighters. These were faster than their Allied opponents and, most important, could carry two machine guns without sacrificing performance. The Albatros D.I pioneered a fighter configuration that was to prevail into the 1930s: a compact, single-seat, externally braced tractor biplane armed with two synchronized machine guns mounted ahead of the pilot on the upper fuselage decking and aimed with a simple ring-and-bead sight. Albatros fighters gave British airmen a terrible drubbing above the Arras battlefield during the "Bloody April" of 1917, but a new generation of French and British fighters with more powerful engines soon tilted the balance toward the Allies. Prominent among these were the French Spad fighters and the British S.E.5, both powered by the Spanish-designed and French-built Hispano-Suiza water-cooled V-8, as well as the British Sopwith Camel and new versions of the French Nieuport, powered by improved rotary radial engines.

Though Germany fell decisively behind France and Britain in aircraft production in 1917, and thus lost the war in the air, perhaps the definitive single-seat fighter of World War I was the Fokker D.VII of 1918. Typically powered by a 160-horsepower Mercedes engine, the D.VII was a fabric-covered biplane that differed from others in having a sturdy fuselage structure of welded steel tubing. Armed with two machine guns, it had a top speed of 117 miles per hour. Even more powerful engines made two-seat fighters possible; the best of these was the British Bristol F.2b, powered by the 220-horsepower, water-cooled Rolls-Royce Falcon, a V-12 engine that gave the Bristol a top speed of almost 120 miles per hour. The F.2b was armed with a synchronized machine gun for the pilot and two flexible machine guns for the observer.

The Fokker D.VII

Ground attack. The Allies fielded specialized aircraft for ground attack only at the very end of the war. Notable among these was the Sopwith Salamander, a development

of the Sopwith Camel with an armoured cockpit and two machine guns firing downward through the floor at a fixed angle to rake enemy trenches while flying low over them. The Germans produced a number of specialized two-seat aircraft for this purpose—notably the Halberstadt CL.III of 1917, which was armed with a forward-firing synchronized machine gun as well as a flexible gun and racks of grenades for the observer. At the Battle of Cambrai in November and December 1917, the Germans sent large formations of such aircraft over the British trenches and into the rear areas with devastating effect. By the end of the war, they were using numbers of armoured, all-metal Junkers J.1 ground-attack aircraft, one of the most advanced machines to see combat during the war.

Bombers. Since they had to carry heavy disposable loads over long distances in order to be effective, specialized bombers were slower to develop. The first bombing raids to achieve significant success (and the first to cross national boundaries) were mounted against the Zeppelin works at Friedrichshafen from Belgian bases by airmen of the Royal Naval Air Service (RNAS) on Oct. 8 and Nov. 21, 1914. However, their spectacular success owed more to the highly flammable nature of the zeppelins themselves than to the destructive power of the 20-pound bombs used. These raids prompted the Admiralty to commission the development of the first specialized heavy night bomber, the Handley Page H.P. O/100, which flew for the first time in December 1915. Meanwhile, other air forces began building and putting into service strategic day bombers. Among the first were French Voisins. The type L was used in early 1915 to carry about 130 pounds of small bombs that simply lay in the bottom of the cockpit until the time came for the observer to drop them overboard. Later models had more powerful engines and were equipped alternatively as attack aircraft, carrying up to 660 pounds of bombs or having a 37-millimetre gun mounted in the nose. None flew faster than 84 miles per hour, so the Voisins operated mainly under cover of darkness in the last year of the war.

Italy, too, was quick to appreciate the value of bombing attacks on enemy targets. Its big three-engined, twin-tail-boom Capronis were among the finest bombers of World War I (see Figure 37). Even larger were the Russian Ilya Muromets bombers of the tsar's Squadron of Flying Ships. Designed by Igor Sikorsky, now remembered mainly as a helicopter pioneer, these biplanes spanned about 100 feet and were descended from his "Russky Vityaz" of May 1913, the world's first successful four-engined airplane. About 80 were built, and they made 400 raids on German targets with the loss of only one plane. The best-known German strategic bombers of World War I were twin-engined Gotha "pusher" biplanes, which made several daylight raids on London in formation in the summer of 1917 before reverting to night operations. The German air force also operated a family of giant four-engined metal bombers known as Riesenflugzeug, or R-planes. Typical of these was the Staaken R.VI number R.25, which was powered by four 260-horsepower Mercedes engines. This had a takeoff weight of 25,269 pounds, which included a crew of seven and a bomb load of up to 4,000 pounds.

The Gotha bombers

By courtesy of John W.R. Taylor

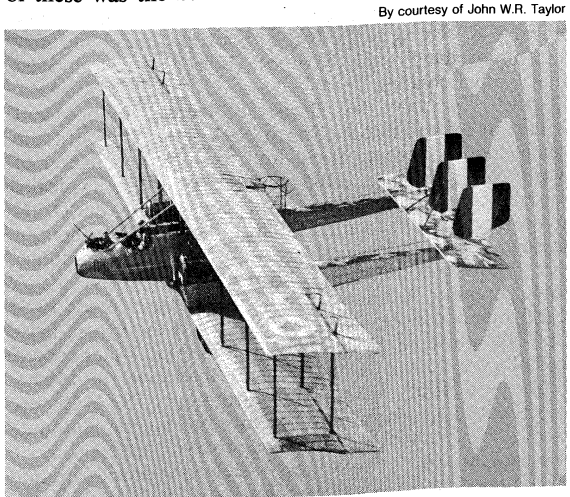


Figure 37: Italian Caproni bomber of World War I.

Naval aviation. Equally significant progress was made in naval flying in World War I. Three distinct categories of combat aircraft emerged: long-range over-water reconnaissance and antisubmarine aircraft operating from shore bases, shorter-range floatplane reconnaissance and fighter aircraft, and ship-borne aircraft. Long-range flying boats (so called because their fuselages were shaped like the hull of a boat) were used extensively by the British. These pioneered the technique of searching for submarines with methodical, mathematically developed search patterns. The German navy made extensive use of reconnaissance and fighter floatplanes from Belgian coastal bases to counter Allied air patrols and coastal naval operations. Some of these, notably Hansa-Brandenburg machines designed by Ernst Heinkel, rivaled their land-based equivalents in performance.

The most efficient of the long-range coastal-based airplanes were large, twin-engined flying boats designed by Glenn Curtiss and others. Despite their bulk, these aircraft were sufficiently fast and maneuverable to engage enemy zeppelins and aircraft in combat. Curtiss' flying boats were the only aircraft of U.S. design to see frontline combat service in World War I.

Carrier-based air power also advanced rapidly. In early 1916 the first landplanes (British Sopwith Pups) were flown off the 200-foot decks of primitive carriers that had been converted from merchant ships, and on Aug. 2, 1917, a pilot landed a Pup on the takeoff deck of HMS *Furious* while the ship was under way. The concept of the true aircraft carrier had been born.

Britain went on to develop more formidable naval aircraft, and in October 1918 a squadron of Sopwith Cuckoos, each able to carry an 18-inch torpedo, was embarked on HMS *Argus*. The war ended before the squadron could go into action; but the RNAS had already used torpedoes dropped from Short seaplanes to sink enemy ships in the Mediterranean, and the Cuckoo, with its modest top speed of 103 miles per hour and endurance of four hours, heralded the eventual demise of the battleship in the face of air-power dominance at sea.

Air transport and training. Military air transport showed little development in 1914–18. Aircraft were used on occasion to drop supplies to cut-off or besieged forces, but the methods were primitive in the extreme: bags of food, medical supplies, or munitions were dropped from bomb racks or simply heaved over the side.

Conversely, training made enormous strides during the war. At the RFC School of Special Flying at Gosport, Eng., Major Robert Smith-Barry introduced a curriculum based on a balanced combination of academic classroom training and dual flight instruction. Philosophically, Smith-Barry's system was based not on avoiding potentially dangerous maneuvers (as had been the case theretofore) but on exposing the student to them in a controlled manner so that he could learn to recover from them, thereby gaining confidence and skill. Technologically, it was based on the Avro 504J, a specialized training aircraft with dual controls, good handling characteristics, adequate power, and in-flight communication between instructor and student by means of an acoustic system of soft rubber tubing—the so-called Gosport tube. For the first time, military pilots flew into action as masters of their airplanes. The Gosport system of training was eventually adopted at training schools throughout the world, remaining the dominant method of civil and military flight instruction into the jet age.

The Gosport system

INTERWAR DEVELOPMENTS

In the two decades between the end of World War I and the start of World War II, military aviation underwent a complete transformation. The typical combat aircraft of 1918 was a fabric-covered, externally braced biplane with fixed landing gear and open cockpits. Few aero engines developed as much as 250 horsepower, and top speeds of 120 miles per hour were exceptional. By 1939, the first-line combat aircraft of the major powers were all-

metal monoplanes with retractable landing gear. Powered by engines that developed 1,000 horsepower or more and that were supercharged to permit flight at altitudes above 30,000 feet, fighters were capable of exceeding 350 miles per hour, and some bombers flew faster than 250 miles per hour. Gyroscopically driven flight instruments and electrical cockpit lighting permitted flying at night and in adverse weather. Crews were seated in enclosed cockpits, were provided with oxygen for breathing at high altitudes, and could converse with other aircraft and ground stations by voice radio. Parachutes, worn by a few German fighter pilots in the last days of World War I, were standard equipment.

Most of these changes occurred after 1930. The end of World War I left the victorious Allies with huge inventories of military aircraft, and this combined with economic strictures and a lack of threat to retard the development of military aviation in the 1920s. Provisions of the Treaty of Versailles prohibiting developments in military aviation had the same effect in Germany. Nevertheless, advances in key technologies, notably high-performance aero engines, continued. The U.S. government, for instance, sponsored a systematic program of aerodynamic research under the aegis of the National Advisory Committee for Aeronautics (NACA), which was to yield enormous dividends in aircraft performance through drag-reduction, engine-cooling, and airfoil technologies. Still, the most significant technical advance in the 1920s was the abandonment of wooden structures in favour of metal frames (still fabric-covered) to provide the strength needed to cope with increasingly powerful engines and to resist harsh climates around the world.

Civilian design improvements. When more drastic changes came, they emerged not from military requirements but from civilian air racing, particularly the international seaplane contests for the coveted Schneider Trophy. Until the appearance of variable-pitch propellers in the 1930s, the speed of landplanes was limited by the lengths of existing runways, since the flat pitch of high-speed propellers produced poor takeoff acceleration. Seaplanes, with an unlimited takeoff run, were not so constrained, and the Schneider races, contested by national teams with government backing, were particularly influential in pushing speeds upward. During the 1920s the Curtiss company built a remarkable series of high-speed racing biplanes for the U.S. Army Air Corps and Navy. These were powered by the innovative D-12, a 12-cylinder, liquid-cooled engine, also of Curtiss design, that set international standards for speed and streamlining. One of the Curtiss planes, an R3C-2 piloted by Lieutenant James Doolittle, won the 1925 Schneider race with a speed of 232.5 miles per hour—in sharp contrast to the winning speed of 145.62 miles per hour in 1922, before the Curtiss machines took part in the event. The influence of the Curtiss engine extended to Europe when British manufacturer C.R. Fairey, impressed with the streamlining made possible by the D-12, acquired license rights to build the engine and designed a two-seat light bomber around it. The Fairey Fox, which entered service in 1926, advanced the speed of Royal Air Force (RAF) bombers by 50 miles per hour and was faster than contemporary fighters. Nor were British engine manufacturers idle; when the U.S. Army and Navy standardized on air-cooled radial engines in the 1920s, Curtiss ceased developing liquid-cooled engines, but British engine designers, partly inspired by the D-12, embarked on a path that was to produce the superlative Rolls-Royce Merlin.

The role of the Schneider races

The year that Doolittle won the Schneider Trophy, an even more revolutionary design appeared—the S.4 seaplane designed by R.J. Mitchell of the British Supermarine Company. A wooden monoplane with unbraced wings, the S.4 set new standards for streamlining, but it crashed from wing flutter before it could demonstrate its potential. Nevertheless, it was the progenitor of a series of monoplanes that won the trophy three times, giving Britain permanent possession in 1931. The last of these, the S.6B, powered by a liquid-cooled Rolls-Royce racing engine with in-line cylinders, later raised the world speed record to more than 400 miles per hour. The S.6B's tapered fuselage and broad, thin, elliptical wings were clearly evident in Mitchell's later and most famous design, the Spitfire (see Figure 38).

In the United States the Thompson Trophy, awarded to the winner of unlimited-power closed-circuit competitions at the National Air Races, was won in 1929 for the first time by a monoplane, the Travel Air "R" designed by J. Walter Beech. Powered by the Wright Cyclone, a 400-horsepower radial engine with a streamlined NACA cowling that contributed 40 miles to its maximum speed of 235 miles per hour, the "R" handily defeated the far more powerful Curtiss biplanes flown by the army and navy. Embarrassed, the military withdrew from racing—and the army soon ordered its first monoplane fighter, the Boeing P-26. In 1935 the industrialist Howard Hughes set a world landplane speed record of 352 miles per hour in a racer designed to his own specifications and powered by a 1,000-horsepower, twin-row radial engine built by Pratt & Whitney. The Hughes H-1 was a low-wing monoplane built with unbraced wings with a "stressed-skin" metal covering that bore stress loads and thereby permitted a reduction in weight of the internal structure. These features, along with a flush-riveted, butt-joined aluminum fuselage, an enclosed cockpit, and power-driven, retractable landing gear folding flush into the wing, anticipated the configuration, appearance, and performance of the fighters of World War II.

Fighters. By the 1930s the advantages of monoplanes with unbraced wings and retractable landing gear were evident, and fighters of this description began to appear. The first of these to see operational service was the Soviet I-16, designed by Nikolay Polikarpov. The I-16 first flew in 1933 and enjoyed considerable success against German and Italian biplanes in the Spanish Civil War of 1936–39. Powered by a radial engine derived from the Wright Cyclone, it had manually retracted landing gear and an open cockpit; its armament of four 7.62-millimetre machine guns, two in the wings and two in the engine cowling, was heavy for the time.

As the I-16 entered combat in Spain, two important British fighters were under development: the Supermarine Spitfire, a cleanly elegant fighter of stressed-skin aluminum construction, and the Hawker Hurricane, a more traditional design with a structural frame of welded steel tubes and a fabric covering over the rear fuselage. Both were powered by a Rolls-Royce Merlin engine of some 1,200 horsepower, and both carried an unprecedented armament of no fewer than eight .303-inch Browning machine guns, mounted in the wings outboard of the propeller arc so that no interrupter gear was needed. Meanwhile, in Germany the nascent Luftwaffe (Air Force) was taking delivery of the first versions of the Bf 109, designed by Willy Messerschmitt for the Bayerische Flugzeugwerke ("Bavarian Aircraft Factory"). Like the Spitfire, the Bf 109 was a low-

The Supermarine Spitfire

Quadrant/Flight

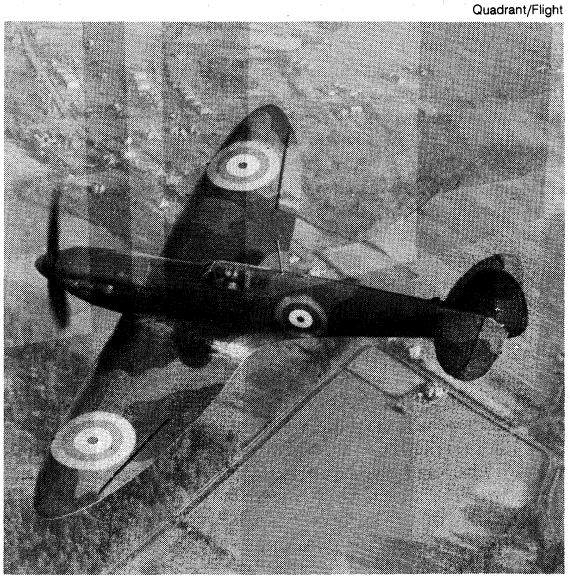


Figure 38: Supermarine Spitfire, Britain's premier fighter plane from 1938 through World War II.

wing monoplane of all-metal, stressed-skin construction. Early versions, fitted with fixed-pitch propellers, fought on a par with the I-16 in Spain, but later versions, powered by a Daimler-Benz engine that was equivalent to the Merlin and fitted with variable-pitch propellers for optimal performance at low and high altitudes, totally outclassed the Russian fighter.

Bombers. Bombers evolved in parallel with fighters, changing to high-strength metal construction in the late 1920s and to monoplane design, which brought higher speeds, in the early 1930s. In 1931 the Boeing Aircraft Company produced the B-9 bomber. Anticipating all-metal fighters, the B-9 was the first operational combat aircraft with all-metal cantilever monoplane design, semiretractable undercarriage, and variable-pitch propellers. Two 600-horsepower engines gave it a speed of 188 miles per hour, representing a 50-percent improvement over the biplane bombers then in service, without any reduction in bombload. Within months of its first flight, the B-9 was overshadowed completely by the Martin B-10 of 1932, which brought the biggest single advance in bomber design since the Handley Page night bomber of World War I. To the innovations of the B-9 it added enclosed cockpits and an internal bay for its 2,260-pound bombload. Maximum speed went up to 213 miles per hour, making the B-10 faster than the fighters of its day. Following this success, Boeing built in 1935 a four-engined craft known as the Model 299, which became the prototype of the B-17 Flying Fortress. This famous plane was based on the concept that a bomber could penetrate to any target in daylight as long as it had sufficient defensive armament to battle past fighter opposition. Gun turrets for defensive machine guns had already been pioneered by Machines Motrices in France, and a license-built version of their turret had appeared on the British Boulton Paul Overstrand bomber in 1934. Meanwhile, the U.S. Army Air Corps claimed that its highly secret Norden bombsight provided such accuracy that "a bomb could be placed in a pickle barrel from 20,000 feet."

The Flying Fortress

An important type of bomber to emerge in the interwar period was the dive-bomber, designed to release its bombs at a low point of a steep dive. Accuracy was maintained by the use of airbrakes, which were flaps that could be extended outward to slow the dive by increasing the aircraft's drag. The dive-bomber as a distinct type of aircraft was a product of tests undertaken during the 1920s by the U.S. Navy. These demonstrated the advantages of bombing the lightly armoured upper decks of warships and resulted in the appearance of the first real dive-bomber, the Curtiss F8C Helldiver, in 1929. Impressed by a Helldiver demonstration, the Luftwaffe, whose doctrine stressed the direct support of ground forces, requested a more advanced aircraft with similar capabilities. The result was the Ju 87 "Stuka" (for *Sturzkampfflieger,* or "dive-bomber"; see Figure 39), which gained a fearsome reputation for destructiveness during the Spanish Civil War.

Carrier aircraft. By the 1930s, ship-based aircraft were fitted under the tail with arrester hooks that engaged cables strung across the landing deck in order to bring them to a halt after landing. Folding wings then enabled them to be taken by elevator to below-deck hangars. Japanese and U.S. aircraft carriers had mixed complements of single-seat fighters, dive-bombers, and torpedo planes; the Royal Navy pursued a less successful course, developing two-seat reconnaissance fighters, such as the Fairey Fulmar, which were outperformed by their land-based equivalents.

Figure 39: German Junkers Ju 87 Stuka dive-bomber.
UPI

WORLD WAR II

Fighters. *Day fighters.* Air superiority was crucial to the outcome of most of the decisive campaigns of World War II, and here the performance of single-seat fighters was generally the critical factor. First-class fighters required extremely powerful aero engines suitable for compact, low-drag installation, and in this respect Britain, Germany, and the United States were in a class by themselves. The only significant exception was the Japanese Mitsubishi A6M carrier fighter, known as the Zero. Designed by Horikoshi Jiro, the Zero was so remarkably strong and light that it achieved first-class performance with a second-class engine—though at the cost of being vulnerable to battle damage.

The outstanding fighters of the early war years (1939–41) were the Spitfire, Bf 109 (known to the wartime Allies as the Me 109), Zero, Hurricane, and Grumman F4F Wildcat (this last a U.S. Navy fighter powered by a turbosupercharged, twin-row radial engine by Pratt & Whitney). The Lockheed P-38 Lightning, a novel twin-boom interceptor designed before the war by Clarence ("Kelly") Johnson, had exceptional performance, but until 1943 it was available only in small numbers. The main U.S. Army Air Force fighters of the early war, the Curtiss P-40 Warhawk and the Bell P-39 Airacobra, were badly outclassed by the Bf 109 and Zero as a result of production decisions that deprived their high-performance Allison engines of scarce turbosuperchargers, assigning them instead to bombers. The best Soviet fighters were similarly outclassed: the MiG-3, designed by a bureau headed by Artem Mikoyan and Mikhail Gurevich, was fast, but it had marginal handling characteristics at low altitudes, and the performance of Semyon Lavochkin's LaGG-3 was ruined by a disastrously heavy airframe.

The Spitfire and Hurricane were determined opponents of the Bf 109 during the Battle of Britain, the first battle fought entirely in the air. The German fighter was armed with two machine guns in the cowling and two wing-mounted cannon firing 20-millimetre exploding shells. Aerial cannon, perfected by the Germans during the war, was intended to ensure the greatest possible destruction against metal-skinned aircraft in the short periods during which a target could be kept in the gunsights at rapidly increasing speeds. It was superior in fighter-to-fighter combat, while the massed batteries of .303-inch machine guns in the British fighters were highly effective in destroying bombers. Aiming was accomplished by gyroscopic lead-computing gunsights that projected the aim point onto a transparent screen in front of the pilot.

The Battle of Britain

More powerful and heavily armed versions of the Spitfire and Bf 109 were tactically viable through the end of the war, but they were hampered by a short radius of action (the farthest distance to which they could fly, engage in combat, and return to base). In 1942–43, fighters began to enter service fitted with newer and more powerful engines and designed on the basis of the most recent aerodynamic data. Notable among these were the German Focke-Wulf Fw 190, designed by Kurt Tank, and the U.S. Republic P-47 Thunderbolt, Grumman F6F Hellcat, and North American P-51 Mustang. All were heavily armed, the Fw 190 with as many as two 7.6-millimetre machine guns and four 20-millimetre cannon, the P-47 eight .50-inch machine guns, and the F6F and P-51 six .50-inch machine guns. The Fw 190, P-47, and F6F had distinctively bulky fuselages widened to accommodate their twin-row radial

engines, while the slimmer P-51, designed in 1940 by J.H. ("Dutch") Kindleberger under a British contract, was fitted with in-line engines and incorporated the latest drag-reduction and airfoil data provided by NACA. Powered by the Rolls-Royce Merlin, the P-51 became the outstanding high-altitude escort fighter of the war. It was at least competitive with contemporary versions of the Spitfire, Bf 109, and Fw 190 in speed, rate of climb, and maneuverability, but it had a more spacious fuselage, a more efficient wing, and, fitted with droppable fuel tanks, a far greater radius of action of more than 1,000 miles. During 1943 the Soviet Red Air Force also gained technical parity with the Luftwaffe with its radial-engined Lavochkin La-5 and La-7 and the in-line-powered Yakovlev Yak-3 and Yak-9.

By war's end, piston-engined fighter technology reached its peak in later versions of the Fw 190, powered by in-line Jumo engines by Junkers, and in the Hawker Tempest, powered by the massive 2,200-horsepower, 24-cylinder, in-line Napier Sabre. Armed with four 20-millimetre cannon and able to attain speeds in excess of 435 miles per hour, the Tempest was the fastest piston-engined fighter ever to see service.

The fastest piston-engined fighter

Night fighters. During the Battle of Britain, the RAF converted twin-engined bombers such as the Bristol Blenheim into night fighters by installing offensive ordnance and radar, but these had little success, since they were no faster than their prey. On the other hand, Messerschmitt's Me 110, a disastrous failure as a twin-engined, two-seat day fighter, became highly successful at night fighting, as did similarly modified Ju 88 bombers. The RAF later used radar-equipped versions of the de Havilland Mosquito to protect its bombers during the battle for the night skies over Germany in 1943–45.

Ground attack. The most effective attack aircraft of the war was the Soviet Ilyushin Il-2 Stormovik. Heavily armoured for protection against ground fire and defended by a gunner in the rear of the two-seat cabin, the Il-2 could fly at up to 280 miles per hour at treetop level and was able to attack ground targets with cannons, bombs, and rockets. It was the first close-support type to employ rockets in vast quantities and had a great influence on the adoption of such weapons by other Allied forces. Another important ground-attack aircraft was Britain's Hawker Typhoon, originally intended to be a high-altitude fighter but limited to low altitudes by its thick wing. Armed with rockets and 20-millimetre cannon, it specialized in attacking trains, tanks, and other moving ground targets.

Bombers. The Stuka dive-bomber was used to great effect during the invasions of Poland, France, and the Low Countries in 1939–40, but its slow speed rendered it vulnerable to fighter attack. The Germans' principal bombers of the Battle of Britain were the twin-engined Heinkel He 111, Dornier Do 17, and Ju 88. The Ju 88 was fast, with a top speed of 280 miles per hour, but it carried a modest bombload; the other German bombers had mediocre performance and were lightly armed by British or American standards. The later Do 217 had a range of 1,500 miles and could carry a bombload of 8,800 pounds, but it was built only in small numbers. The Germans never built a successful four-engined bomber.

Combat experience showed that the heavily armed British and U.S. bombers were more vulnerable to fighter attack than expected. This was dramatically revealed on Dec. 18, 1939, when a formation of Vickers Wellingtons—the most battle-worthy bombers of the day with four-gun Boulton Paul tail turrets—was decimated over the Heligoland Bight by cannon-armed German fighters. In time this led to the adoption of self-sealing fuel tanks, armour protection for crews, and even heavier defensive armament, but the British responded immediately by abandoning daylight bombing except under special circumstances. Bombing at night reduced vulnerability to fighters, but finding and hitting targets proved difficult: nothing smaller than a city could be effectively attacked, and, as operational analysis revealed in 1941–42, ordinary crews had trouble doing even that. The problem was solved partly by using specially trained "Pathfinder" crews to mark targets with flares and partly by electronic navigation aids. During the Battle of Britain, the Germans used electronic beams to guide bombers to their targets at night, and the British later developed such on-board radars as the H_2S blind bombing system, which could produce maplike pictures of terrain beneath the aircraft through clouds or in darkness. From 1943, powerful four-engined bombers such as the Handley Page Halifax and Avro Lancaster, carrying H_2S radar and heavy armament, kept RAF bomber losses within barely acceptable limits.

Night bombing

An independent British development was the de Havilland Mosquito. Constructed entirely of wood, powered by two Rolls-Royce Merlin engines, and carrying a crew of two and no defensive armament, this extraordinarily fast aircraft remained effectively immune to interception until the appearance of jet fighters, and it could reach Berlin with a 4,000-pound bomb. It was perhaps the most successful multimission aircraft ever made, serving with distinction as a low-level day bomber, radar-equipped night fighter, and long-range photoreconnaissance aircraft.

The U.S. Army Air Force armed later versions of its B-17 Flying Fortresses and Consolidated-Vultee B-24 Liberators with 12 or more .50-inch machine guns, eight of them in twin-gun, power-driven turrets in nose, tail, ventral, and belly positions. Still, losses were high, reaching unacceptable numbers in raids against the Schweinfurt ball-bearing works on Aug. 17 and Oct. 14, 1943. Daylight bombing had to be curtailed until the arrival of P-38, P-47, and P-51 escort fighters equipped with drop tanks to provide the necessary range. For high-altitude attacks from 25,000 feet, the B-17 could carry 4,000 pounds of bombs at 215 miles per hour with a radius of action of some 800 miles. The B-24 carried more bombs and was slightly faster, but it could not fly as high and was more vulnerable to enemy fire. British heavy bombers carried larger bombloads—the Lancaster could carry 7,000 pounds with a radius of action of nearly 1,000 miles or a bombload of 14,000 pounds over a radius of 500 miles—but only at medium altitudes of less than 20,000 feet. The heaviest bomber of World War II was the Boeing B-29 Superfortress (see Figure 40), which entered service in 1944 with a fully pressurized crew compartment (previously used only on experimental aircraft) and as many as 12 .50-inch machine guns mounted in pairs in remotely-controlled turrets. Although these features were intended to optimize the B-29 for very high-altitude missions at 35,000 feet, it was most effectively used when, stripped of almost all its heavy defensive armament, it carried bombloads as heavy as 12,000 pounds in low altitude firebombing attacks against Tokyo and other Japanese cities from bases 2,000 miles away in the Mariana Islands. Specially modified B-29s dropped atomic bombs on Hiroshima and Nagasaki.

Boeing photo

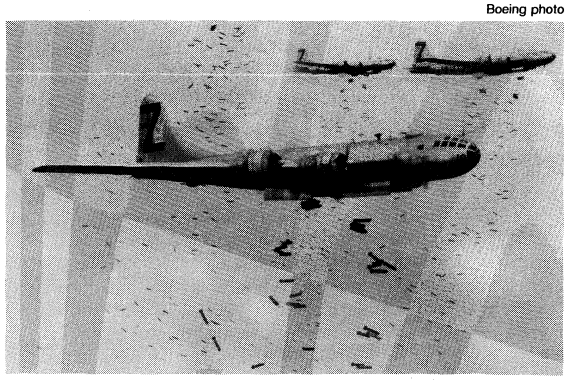


Figure 40: Boeing B-29 Superfortresses, U.S. long-range bombers built for the high-altitude bombing of Japan.

Naval aviation. During World War II, carrier-based attack aircraft replaced the big guns of capital ships as the dominant offensive weapon of naval warfare. This was first demonstrated by the destruction of Italian battleships at Taranto by Fairey Swordfish torpedo biplanes on the night of Nov. 11–12, 1940; by the Japanese attack on Pearl Harbor on Dec. 7, 1941; and by the decisive Battle of Midway (June 3–6, 1942), in which surface vessels never exchanged gunfire while U.S. aircraft destroyed four Japanese aircraft carriers for the loss of only one of their

Air power at sea

own. In addition to such fighters as the F6F, Zero, and modified Spitfires and Hurricanes, notable carrier aircraft of the war included dive-bombers such as the U.S. Douglas SBD Dauntless and Japanese Aichi 99 as well as torpedo planes such as the Grumman TBF Avenger and Nakajima B5N.

Land-based torpedo planes were also effective, as shown in attacks on the British battleships *Repulse* and *Prince of Wales* by twin-engine Japanese Mitsubishi G3M bombers off Malaya on Dec. 10, 1941.

Kamikaze attacks, a Japanese suicide tactic first used in the Battle of Leyte Gulf on Oct. 25, 1944, were very destructive as long as the supply of skilled volunteer pilots held out. First conducted with bomb-armed Zero fighters, they later expanded to encompass bombers and such special craft as a piloted, rocket-propelled winged bomb called the Ohka ("Cherry Blossom"). By the end of the war, however, there were no more skilled kamikaze volunteers, and the tactic became no more effective than traditional dive bombing.

Reconnaissance. For military staffs contemplating offensive operations, aerial photography became the most important source of detailed information on enemy dispositions. British reconnaissance aircraft were especially capable. Modified versions of the Spitfire and Mosquito, stripped of armament and fitted with extra fuel tanks, proved essentially immune to interception at high altitudes. Stripped-down versions of the P-38 and P-51, called the F-4 and F-5, were also effective photoreconnaissance platforms, the latter excelling at high-resolution coverage from low altitudes.

Training. Japan and Germany entered World War II with exceptionally well-trained aviators, but their provisions for training replacements were inadequate. The British Commonwealth and the United States gained a vital advantage over the Axis by establishing large, well-organized air-crew training programs. Outstanding training aircraft included the British de Havilland Tiger Moth, the U.S. Stearman PT-19, and the German Bücker Bü 133 Jungmeister—all biplanes. Only the United States built specialized single-engined trainers with such features characteristic of operational craft as retractable landing gear and variable-pitch propellers. Notable among these was the North American AT-6.

Air transport. Major advances in air transport were made during the war. Mass drops of parachute troops had been pioneered by the Soviet Union in the 1930s, but the Luftwaffe first used the technique operationally, notably during the invasion of Crete, in which 15,000 airborne and parachute troops were landed onto that island by 700 transport aircraft and 80 gliders. The troop-carrying glider was one of the developments of World War II that had no continuing place in postwar air forces, but the transport airplane was only at the beginning of its useful life. The Germans built transports such as the Ju 52 only in small quantities, but the twin-engined Douglas C-47, which had revolutionized American commercial aviation in the mid-1930s as the DC-3, was produced in huge numbers and was the backbone of tactical air transport in every Allied theatre of the war. One of the few transports with a large side door suitable for dropping paratroopers, the C-47 was also the mainstay of British and American airborne operations. Douglas also manufactured the four-engined C-54, which entered service in 1943–44 as the first land-based transport with intercontinental flight capabilities. The C-54 was particularly important in the vast distances of the Pacific-Asian theatre of operations.

The Douglas C-47

Helicopters. In the years before World War II, both the U.S. Army and the RAF had experimented with autogiros; these were craft that employed a propeller for forward motion and a freely rotating, unmotorized rotor for lift. In the event, autogiros proved too expensive and mechanically complex and were supplanted by conventional light aircraft. Meanwhile, during the late 1930s Igor Sikorsky in the United States and Anton Flettner and Heinrich Focke in Germany had perfected helicopter designs with serious military potential. The Sikorsky R-4, powered by a single lifting rotor and an antitorque tail rotor, was used for local rescue duties at U.S. air bases in the Pacific and was also used in several combat rescues in Burma. The German navy used a handful of Flettner Fl 282s, powered by two noncoaxial, contrarotating lifting rotors, for ship-based artillery spotting and visual reconnaissance.

THE JET AGE

Beginning in the 1920s, steady advances in aircraft performance had been produced by improved structures and drag-reduction technologies and by more powerful, supercharged engines, but by the early 1930s it had become apparent to a handful of farsighted engineers that speeds would soon be possible which would exceed the capabilities of reciprocating engines and propellers. The reasons for this were not at first widely appreciated. At velocities approaching Mach 1, or the speed of sound (about 745 miles per hour at sea level and about 660 miles per hour at 36,000 feet), aerodynamic drag increases sharply. Moreover, in the transonic range (between about Mach 0.8 and Mach 1.2), air flowing over aerodynamic surfaces stops behaving like an incompressible fluid and forms shock waves; these in turn create sharp local discontinuities in airflow and pressure, creating problems not only of drag but of control as well. Because propeller blades, describing a spiraling path, move through the air at higher local velocities than the rest of the aircraft, they enter this turbulent transonic regime first. For this reason, there is an inflexible upper limit on the speeds that can be attained by propeller-driven aircraft. Such complex interactions in the transonic regime—and not the predictable shock-wave effects of supersonic flight, which ballisticians had understood since the late 19th century—presented special problems that were not solved until the 1950s. In the meantime, a few pioneers attacked the problem directly by conceiving a novel power plant, the turbojet.

The first turbojets

While still a cadet at the Royal Air Force College, Cranwell, in 1928, Frank Whittle advanced the idea of replacing the piston engine and propeller with a gas turbine, and in the following year he conceived the turbojet, which linked a compressor, combustion chamber, and turbine in the same duct. In ignorance of Whittle's work, three German engineers independently arrived at the same concept: Hans von Ohain in 1933; Herbert Wagner, chief structural engineer for Junkers, in 1934; and government aerodynamicist Helmut Schelp in 1937. Whittle had a running bench model by the spring of 1937, but backing from industrialist Ernst Heinkel gave von Ohain the lead. The He 178, the first jet-powered aircraft, flew on Aug. 27, 1939, nearly two years before its British equivalent, the Gloster E.28/39, on May 15, 1941. Through an involved chain of events in which Schelp's intervention was pivotal, Wagner's efforts led to the Junkers Jumo 004 engine. This became the most widely produced jet engine of World War II and the first operational axial-flow turbojet, one in which the air flows straight through the engine. By contrast, the Whittle and Heinkel jets used centrifugal flow, in which the air is thrown radially outward during compression. Centrifugal flow offers advantages of lightness, compactness, and efficiency—but at the cost of greater frontal area, which increases drag, and lower compression ratios, which limit maximum power. Many early jet fighters were powered by centrifugal-flow turbojets, but, as speeds increased, axial flow became dominant.

Fighters. *Subsonic flight.* Though Whittle was first off the mark, the Germans advanced their programs with persistence and ingenuity. The Messerschmitt Me 262, powered by two Jumo engines and with wings swept back 18.5°, was capable of 525 miles per hour. Armed with four 30-millimetre cannon and unguided rockets, it was an effective bomber destroyer, but it entered service too late to have a major effect on the war. The Gloster Meteor entered service on July 27, 1944, about two months before the Me 262; though it was less capable than the German fighter, it was effective in intercepting V-1 "buzz bombs." Desperate to combat Allied bombers, the Germans also turned to rocket propulsion, fielding the tailless Me 163 Komet in the final months of the war. Powered by a hydrogen peroxide rocket designed by Helmuth Walter, the Komet had spectacular performance, but its short range and ineffective cannon armament made it an operational

failure. In addition, the propellants were unstable and often exploded on landing.

Meanwhile, the U.S. aviation industry entered the jet race with the receipt by General Electric of a Whittle engine in 1941. The first U.S. jet, the Bell P-59A Airacomet, made its first flight the following year. It was slower than contemporary piston-engined fighters, but in 1943–44 a small team under Lockheed designer Kelly Johnson developed the P-80 Shooting Star. The P-80 and its British contemporary, the de Havilland Vampire, were the first successful fighters powered by a single turbojet.

The jets of World War II inaugurated the first generation of jet fighters, in which turbojet propulsion was applied to existing airframe technology and aerodynamics. (Indeed, some early postwar jets—notably, the Soviets' Yakovlev Yak-15 and Yak-23 and the Swedish Saab 21R—were simply re-engined propeller-driven fighters.) These aircraft generally outperformed their piston-engined contemporaries by virtue of the greater thrust that their jets provided at high speeds, but they suffered from serious deficiencies in range and handling characteristics owing to the high fuel consumption and slow acceleration of early turbojets. More fundamentally, they were limited to subsonic speeds because the relatively thick airfoils of the day were prone to the compressibility problems of transonic flight—especially at high altitudes, where the higher speeds required to produce lift in thin atmosphere brought aircraft more quickly to transonic speed. For this reason, first-generation jets performed best at low altitudes.

Limitations of early jet fighters

Other first-generation fighters included the U.S. McDonnell FH Phantom and the British Hawker Sea Hawk (the first jet carrier fighters), the McDonnell F2H Banshee, and the French Dassault Ouragan. These single-seat day fighters were in service by 1950, while first-generation all-weather fighters, burdened with radar and a second crew member, entered service through the late 1950s.

Transonic flight. As the first generation of jet fighters entered service, many aerodynamicists and engineers believed supersonic flight a practical impossibility, owing to transonic drag rise or compressibility, which threatened to tear an aircraft apart. Nevertheless, on Oct. 14, 1947, U.S. Air Force Captain Charles Yeager, flying a rocket-powered Bell X-1 launched from the bomb bay of a B-29, became the first human to exceed the speed of sound. Designed exclusively for research, the X-1 had thin, unswept wings and a fuselage modeled after a .50-inch bullet. Yeager's flight marked the dawn of the supersonic era, but it was only part of a broad wave of testing and experimentation that had begun during World War II. Germany had experimented then with swept-back and delta-shaped wings, which delayed transonic drag rise, and, after extensive testing, these configurations were widely adopted in the postwar years. At the same time, the development of slats, slotted flaps, and other sophisticated high-lift devices for landing and takeoff enabled designers to use smaller wings, which in turn allowed them to achieve higher speeds. Turbojets became more powerful, and, in the late 1950s, afterburning, or reheat, was introduced; this permitted large temporary thrust increases by the spraying of fuel into hot exhaust gases in the tailpipe—in effect turning the turbojet into a ramjet.

As these developments took hold, a second generation of fighters appeared that were capable of operating in the transonic regime. These aircraft had thinner lifting and control surfaces than first-generation jets, and most had swept-back wings. Aerodynamic refinements and more powerful, quicker-accelerating engines gave them better flight characteristics, particularly at high altitudes, and some could exceed the Mach in a shallow dive. In addition, airborne radars became more compact and reliable, and radar-ranging gunsights began to replace the optically ranging sights used in World War II. Air-to-air missiles, using radar guidance and infrared homing, became smaller and more capable (see above *Rockets and missile systems: Tactical guided missiles*). Outstanding fighters of this generation were the U.S. North American F-86 Sabre and its opponent in the Korean War (1950–53), the Soviet MiG-15. The F-86 introduced the all-flying tail (later a standard feature on high-performance jets), in which the entire horizontal stabilizer deflects as a unit to control pitch, yielding greater control and avoiding the compressibility problems associated with hinged surfaces. This and a radar-ranging gunsight helped the F-86 achieve a favourable kill ratio over the MiG-15, despite the Soviet fighter's greater speed, higher service ceiling, and heavier armament. Other jets of this generation were Britain's superlative Hawker Hunter, the MiG-17, and the diminutive, British-designed Folland Gnat. The latter two, introduced in the mid-1950s, later became successful low-altitude dogfighters—the Gnat against Pakistani F-86s in the Indo-Pakistani conflict of 1965 and the MiG-17 against U.S. aircraft in the Vietnam War (1965–73).

The F-86 and MiG-15

Supersonic flight. A third generation of fighters, designed around more powerful, afterburning engines and capable of level supersonic fight, began to enter service in the mid-1950s. This generation included the first fighters intended from the outset to carry guided air-to-air missiles and the first supersonic all-weather fighters. Some were only marginally supersonic, notably the U.S. Convair F-102 Delta Dagger, an all-weather interceptor that was the first operational "pure" delta fighter without a separate horizontal stabilizer. Other aircraft included the Grumman F11F Tigercat, the first supersonic carrier-based fighter; the North American F-100 Super Sabre; the Dassault Mystère B-2; the Saab 35, with a unique double-delta configuration; and the MiG-19.

To this point, jet fighters had been designed primarily for air-to-air combat, while older aircraft and designs falling short of expectations were adapted to ground attack and reconnaissance. Since land-based surface attack was to be carried out by bombers, the first operational jets of fighter size and weight designed to attack surface targets were based on aircraft carriers. These paralleled the third generation of fighters, but they were not supersonic. One example was the British Blackburn Buccaneer, capable of exceptional range at low altitudes and high subsonic speeds. The Douglas A-4 Skyhawk, entering service in 1956, sacrificed speed for ordnance-delivery capability. One of the most structurally efficient aircraft ever built, it carried the burden of U.S. Navy attacks on ground targets in North Vietnam and was often used by Israeli pilots in the Middle Eastern conflicts. The Grumman A-6 Intruder, which entered service in the 1960s, was another subsonic carrier-based aircraft. The first genuine night/all-weather, low-altitude attack aircraft, it was highly successful over North Vietnam.

Ground-attack jets

Mach 2. A fourth generation of fighters began to appear in the 1960s, capable of maximum speeds ranging from about Mach 1.5 to 2.3. Top speeds varied with the intended mission, and increasing engine power, aerodynamic sophistication, and more compact and capable radars and avionics began to blur the differences between two-seat, all-weather fighters and single-seat air-superiority fighters and interceptors. By this time, military designers had become persuaded that air-to-air missiles had made dogfighting obsolete, so that many interceptors were built without guns. This generation included the first land-based jet fighters designed with surface attack as a secondary or primary mission—a development driven by the appearance of surface-to-air missiles such as the Soviet SA-2, which denied bombers medium- and high-altitude penetration.

Precursor to this generation was the Lockheed F-104 Starfighter, designed by a team under Kelly Johnson and first flown in 1954. Capable of speeds well above Mach 2, this interceptor was built with short and extremely thin wings to reduce the generation of shock waves. However, light armament, limited avionics, and poor maneuverability made it an ineffective air-to-air fighter, and only with the installation of up-to-date bombing and navigation systems in the 1960s did it become a useful low-level attacker. The truly outstanding fighters were the U.S. McDonnell F-4 Phantom II and the MiG-21. A large, twin-engined two-seater, the F-4 was originally a carrier-based interceptor armed only with air-to-air missiles, but it was so successful that the U.S. Air Force adopted it as its primary fighter. When combat in Vietnam showed that gun armament was still valuable for close-range dogfight-

The F-4 and MiG-21

ing, later versions of the F-4 were fitted with an internally mounted, 20-millimetre rotary cannon. The MiG-21 was a small, delta-wing, single-seat aircraft designed as a specialized daylight interceptor, but it soon proved amenable to modification for a broad range of missions and became the most widely produced jet fighter ever. It was a formidable threat to U.S. airmen over North Vietnam and to Israeli pilots over the Sinai Peninsula and Golan Heights in 1973. Also outstanding was the Republic F-105 Thunderchief, one of the largest single-engined fighters ever built. Designed to carry a nuclear bomb internally as a low-altitude penetrator and therefore exceptionally fast at low altitudes, the F-105, with heavy loads of conventional bombs under the wings, carried out the brunt of U.S. Air Force attacks against North Vietnam. Also noteworthy in this generation were the British Electric Lightning, one of the first Mach-2 interceptors to enter service and one of the fastest at high altitudes; the Soviets' twin-engined, all-weather Yak-28 Firebar; the Convair F-106 Delta Dart, a single-seat air-defense interceptor with superior speed and maneuverability; the Dassault Mirage III, the first successful pure delta in the air-to-air role and an enormous export success; the Soviet Sukhoy Su-21 Flagon, a tailed-delta, single-seat, all-weather interceptor; and the Vought F-8 Crusader, an outstanding carrier-based dogfighter over Vietnam.

Multimission. By the 1970s, steady improvements in engine performance, aerodynamics, avionics, and aircraft structures resulted in a trend toward multimission fighters. Also, as engine acceleration characteristics improved dramatically and radars, fire-control systems, and air-to-air missiles became more compact and capable, the performance of aircraft themselves became less important than the capabilities of their missiles and sensors. It was now clear that, even with supersonic aircraft, almost all aerial combat occurred at transonic and subsonic speeds. Thenceforth, speed and operating ceiling were traded off against sustained maneuvering energy, sensor capabilities, mixed ordnance of guns and missiles, range, takeoff and landing qualities, multimission capability, political goals, and—above all—cost. A dramatic manifestation of the complexity of this new design equation was the Hawker Harrier, the first vertical/short takeoff and landing (V/STOL) fighter. Transonic and short-ranged but able to dispense with runways, the Harrier became operational with the RAF in 1967 and over the following decades was fitted with avionics of growing capabilities. The Royal Navy's Sea Harrier version (see Figure 41) distinguished itself in the 1982 Falkland Islands War both against Argentine ground positions and in dogfights with A-4s and Mirage IIIs.

The Harrier

The new generation of fighters was characterized by Mach 2+ performance where necessary, multimission capability, and sophisticated all-weather avionics. Many aircraft of this generation employed variable-geometry wings, permitting the amount of sweep to be changed in flight to obtain optimal performance for a given speed. Important aircraft in this generation included, roughly in order of operational appearance, the following: the MiG-25 Foxbat, a large single-seat interceptor and reconnaissance aircraft with a service ceiling of 80,000 feet and a top speed on the order of Mach 2.8 but with limited maneuverability and low-altitude performance; the MiG-23 Flogger, a variable-wing interceptor able to acquire and engage with missiles below it in altitude; the MiG-27 Flogger, a ground-attack derivative of the MiG-23; the Saab 37 Viggen, designed for short takeoff with a main delta wing aft and small delta wings with flaps forward; the fixed-wing Sepecat Jaguar, developed by a French-British consortium in ground-attack and interceptor versions; the Grumman F-14 Tomcat, a highly maneuverable, twin-engined, two-seat, variable-geometry interceptor armed with long-range missiles for the defense of U.S. aircraft-carrier fleets; the Dassault-Breguet Mirage F1, designed for multimission capability and export potential; the McDonnell Douglas F-15 Eagle, a single-seat, twin-engined, fixed-geometry air-force fighter designed for maximum sustained maneuvering energy and the first possessor of a genuine "look-down/shoot-down" capability, which was the product of pulse-Doppler radars that could detect fast-moving targets against cluttered radar reflections from the ground; the Panavia Tornado, a compact, variable-geometry aircraft developed jointly by West Germany, Italy, and Great Britain in no fewer than four versions, ranging from two-seat, all-weather, low-altitude attack to single-seat air-superiority; the U.S. General Dynamics F-16 Fighting Falcon (see Figure 41), a high-performance, single-seat multirole aircraft with impressive air-to-ground capability; the MiG-29 Fulcrum, a single-seat, twin-engined, fixed-geometry interceptor with a look-down/shoot-down capability; the MiG-31 Foxhound interceptor, apparently derived from the MiG-25 but with less speed and greater air-to-air capability; and the McDonnell Douglas F/A-18 Hornet, a single-seat, carrier-based aircraft designed for ground attack but also possessing excellent air-to-air capability.

Bombers. *High-altitude craft.* The Luftwaffe fielded the first operational jet bomber, the Arado Ar 234, in the waning months of World War II, but it had minimal impact. The jet bombers of the immediate postwar years enjoyed only indifferent success, mostly serving to test engineering and operational concepts and being produced in small numbers. By the mid-1950s, however, first the Americans and then the British and Soviets began to field highly capable jet bombers. The first of these to be produced in large numbers was the swept-wing, six-engined Boeing B-47 Stratojet, used by the U.S. Strategic Air Command as a long-range nuclear weapons carrier. Deployed in 1950, it was followed in 1955 by the eight-engined Boeing B-52

The B-52

By courtesy of (left) General Dynamics Corp., (right) British Aerospace

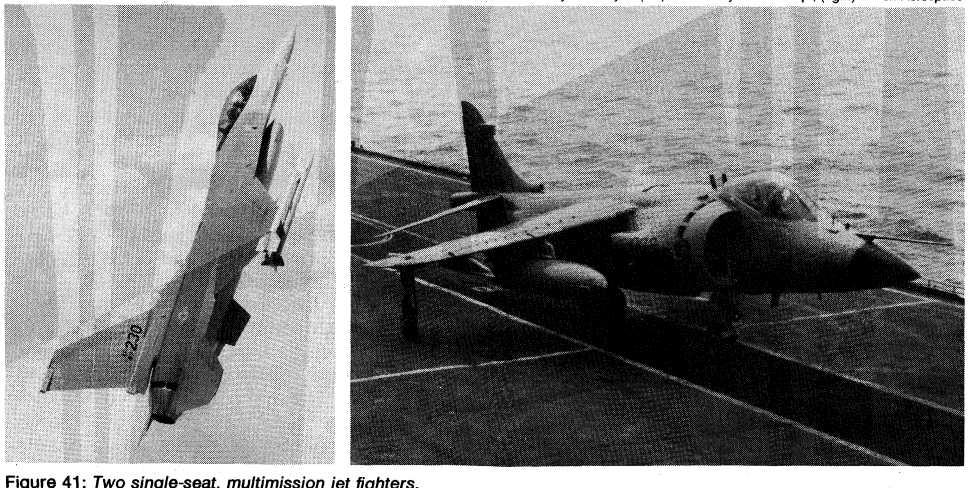


Figure 41: *Two single-seat, multimission jet fighters.*
(Left) U.S. F-16 Fighting Falcon armed with two air-to-air missiles: the AIM-9 Sidewinder at the wingtip and the AIM-120 AMRAAM inboard. (Right) British Sea Harrier, capable of vertical/short takeoff and landing.

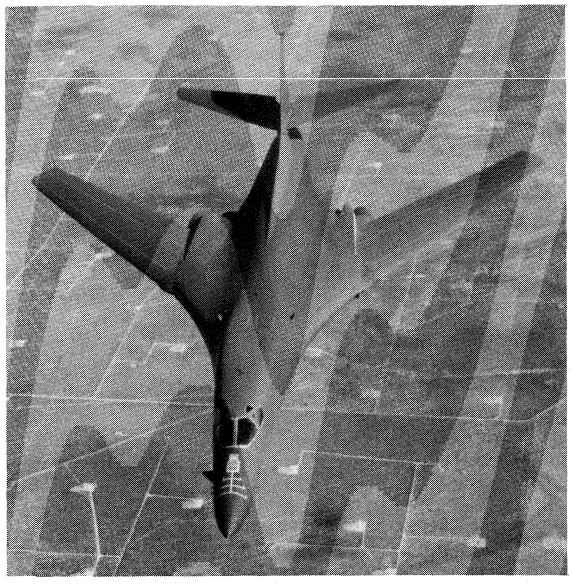


Figure 42: *The B-1B, strategic bomber of the U.S. Air Force.* Powered by four turbofan engines, the B-1B was designed for low-level, transonic penetration of radar defenses. The variable wings can swing forward for landing and takeoff.
© Mi Seitelman/IDI

Stratofortress. This huge bomber, 153 feet long and with a wing span of 185 feet, remained the principal long-range nuclear weapons carrier of the United States for 30 years. During the Vietnam War it dropped conventional bombs on both tactical and strategic missions, and in the 1980s it received a new lease on life by being fitted with air-launched cruise missiles, which permitted it to threaten targets from beyond the range of air-defense systems.

The British "V-bombers," introduced in the 1950s, comprised the Vickers Valiant, Handley Page Victor, and Avro Vulcan. These served as the backbone of Britain's nuclear deterrent until superseded by Polaris-missile-equipped nuclear submarines in the 1970s. The Vulcan, the first jet bomber to use the delta-wing configuration, remained in service long enough to drop conventional bombs in the Falkland Islands War.

The first Soviet jet bombers with strategic potential were the twin-engined Tupolev Tu-16 Badger (deployed in 1954) and the larger and less successful four-engined Myasishchev M-4 Bison (deployed in 1956). In 1956 the Soviets also fielded the only turboprop strategic bomber to see service, the Tu-95 Bear. A large, swept-wing aircraft powered by four huge turboprop engines with contrarotating propellers, the Tu-95 proved to have excellent performance. Like the B-52, it was adapted to maritime and cruise missile patrol after it had become obsolete as a strategic bomber.

The aircraft mentioned above were capable of only subsonic speeds. The first operational supersonic bomber was the delta-winged Convair B-58 Hustler of the United States. Placed in active service in 1960, this bomber carried its nuclear weapon and most of its fuel in a huge, jettisonable pod beneath the fuselage.

Low-level penetration. The B-58 had a service life of only three years, because in the early 1960s it became apparent that surface-to-air missiles could shoot down aircraft even at previously safe altitudes of over 50,000 feet. In response, bombers sought protection from early-warning radar by flying at low levels, and a new generation of high-performance bombers came into service that took complete advantage of the propulsion, aerodynamic, and electronic advances of the postwar era. The first of these was the U.S. General Dynamics F-111, the first operational aircraft to use a variable-sweep wing. Variable geometry was originally intended to allow the F-111 to combine the missions of low-altitude bomber and high-altitude fleet-defense fighter, but the fighter version was eventually abandoned. After a poor showing in Indochina in 1968, the F-111 became a successful high-speed, low-altitude, all-weather penetrator. As such, it joined with considerable effect in the final stages of the U.S. aerial offensive on North Vietnam, and it was also assigned to NATO as a tactical-range nuclear weapons carrier. The Soviet Su-24 Fencer was similar to the F-111.

The B-1 and Tu-26

Larger strategic bombers using variable geometry to achieve high performance at low altitudes included the Soviet Tu-26 Backfire, the U.S. Rockwell International B-1, and the Tu-160 Blackjack. These bombers, supplementing the older purely subsonic aircraft, formed an important part of the U.S. and Soviet nuclear forces after their deployment in 1975, 1985, and 1988, respectively. In common with all first-line combat aircraft, they were equipped with sophisticated electronic countermeasure (ECM) equipment designed to jam or deceive enemy radars. They could deliver free-fall conventional or nuclear bombs, air-to-surface missiles, and cruise missiles. The B-1B version (see Figure 42) could achieve supersonic flight only in short bursts at high altitude, while the Soviet bombers were capable of supersonic "dash" at low level and could fly at twice the speed of sound at high altitude.

Stealth. *The first operational craft.* The existence of a Stealth program, designed to produce aircraft that were effectively immune to radar detection at normal combat ranges, was announced by the U.S. government in 1980. The first aircraft employing this technology, the single-seat Lockheed F-117A ground-attack fighter, became operational in 1983. The second was the Northrop B-2 strategic bomber, which first flew in 1989. Both aircraft had unconventional shapes that were designed primarily to reduce radar reflection. The B-2 was of a flying-wing design that made it only slightly longer than a fighter yet gave it a wingspan approaching that of the B-52, while the F-117A had a short, pyramid-shaped fuselage and sharply swept wings.

Stealth technology. Ever since radar-directed defenses began taking a toll of bomber formations in World War II, aircraft designers and military aviators had sought ways to avoid radar detection. Many materials of the early jet age were known to absorb radar energy rather than reflect it, but they were heavy and not strong enough for structural use. It was not until after the 1960s and '70s, with the development of such materials as carbon-fibre composites and high-strength plastics (which possessed structural strength as well as being transparent or translucent to radar), that radar signature reduction for piloted combat aircraft became possible.

Reducing radar reflectivity

Reducing radar signature also required controlling shape, particularly by avoiding right angles, sharp curves, and large surfaces. In order to direct radar energy in the least revealing directions, the external shape of a stealth aircraft was either a series of complex, large-radius, curved surfaces (as on the B-2) or a large number of small, flat, carefully oriented planes (as on the F-117A). Fuel and ordnance were carried internally, and engine intakes and exhausts were set flush or low to the surface. To avoid interception of radar emissions, stealth aircraft had to rely on inertial guidance or other nonemitting navigational systems. Other possibilities included laser radar, which scanned the ground ahead of the craft with a thin, almost undetectable laser beam.

To escape detection in the infrared spectrum, stealth aircraft were not equipped with large, heat-producing afterburner engines. This rendered them incapable of supersonic flight. Also, the shapes and structures optimal for stealth aircraft were often at odds with aerodynamic and operational requirements. Since all weaponry had to be carried internally, ordnance loads were less than for equivalent conventional aircraft, and sophisticated artificial stabilization and control systems were needed to give stealth aircraft satisfactory flying characteristics. Unlike the fighter, the B-2 had no vertical fin stabilizers, relying on flaps on the trailing edge of its notched wing to control roll, pitch, and yaw.

Other military aircraft. *Transport.* The success of the C-47 and C-54 in World War II inspired the development of specialized military freighters with nose- and tail-loading features, roller conveyors on the floor, and built-in winches. These permitted the quick loading of vehi-

Figure 43: *Soviet Mi-24 Hind attack helicopter in the Czechoslovak Air Force.*
The weapon operator is seated under a separate canopy below the pilot. A Gatling-type machine gun projects from the "chin" directly beneath the weapon operator. Each auxiliary wing has two rails for guided antitank missiles and two pods for unguided antitank or antipersonnel rockets.
CTK from Eastfoto

cles and large equipment as well as their air-dropping by parachute. Military transports ranged from small V/STOL liaison aircraft and modified versions of civilian transports to huge craft such as the Lockheed C-5 Galaxy, designed in the 1960s to carry two M-60 tanks, 16 three-quarter-ton trucks, or 245 troops. After its introduction in 1969, the C-5 was the largest aircraft in the world for almost two decades, until it was surpassed by the Soviet Antonov An-225. With a cargo bay 21 feet wide, 14.5 feet high, and 140 feet long, the An-225 was designed to carry a payload of as much as 440,800 pounds.

Reconnaissance. Reconnaissance aircraft also carried ECM devices and relied heavily on electronic and infrared sensors to supplement their film cameras. Their tasks were to locate and photograph targets, using radar and conventional photographic techniques, and to probe enemy electronic defense systems to discover and evaluate the types of radio and radar equipment that were in use. They did this by offshore patrols, just outside territorial limits and, more rarely, by overflights. The best-known American type used for overflights in the 1950s and '60s was the Lockheed U-2, which could supplement its long endurance at very high altitude by shutting down its engine and gliding for a time. Successor to the U-2 until its retirement in 1990 was the Lockheed SR-71 Blackbird, which could cruise at speeds above Mach 3 and at altitudes as high as 100,000 feet for very long periods.

The U-2

Airborne early warning. Carrier-based early-warning aircraft had a large radar to detect aircraft or ships; some could also control interceptor fighters defending the fleet. This kind of airborne warning and control system (AWACS) airplane appeared in land-based air forces to detect low-flying enemy raiders and direct interceptors toward them. The first aircraft of this type was a Soviet turboprop, the Tu-126 Moss, which was succeeded in the 1980s by the jet-powered Ilyushin Il-76 Mainstay. These craft, like the U.S. E-3 Sentry (a converted Boeing 707), carried a large, saucer-shaped radar on their fuselage. Britain's early-warning aircraft was the British Aerospace Nimrod.

Helicopters. The helicopter had its first significant impact on military operations during the Korean War, but it came of age in Vietnam. Helicopters fielded air-mobile infantry units, evacuated casualties, hauled artillery and ammunition, rescued downed aviators, and served as ground-attack craft.

Helicopters became serious operational machines only after American manufacturers fitted them with gas-turbine engines. Turbine engines were much less sensitive than piston engines to high temperatures and low atmospheric density, and they had far greater power-to-weight ratios and occupied considerably less space.

Assault and attack helicopters. The mainstay of U.S. Army assault units in Vietnam was the Bell UH-1 Iroquois, popularly known as the Huey. As early as 1962, army aviators added turret-mounted automatic 40-millimetre grenade launchers, skid-mounted rocket pads, and remotely trainable 7.62-millimetre machine guns. These experiments, which proved effective in supporting helicopter assault operations, led to the AH-1G HueyCobra, deployed in 1967 as the first purpose-built helicopter gunship. With its pilot seated behind and above the gunner, the HueyCobra pioneered the tandem, stepped-up cockpit configuration of future attack helicopters.

After the Vietnam War, the lead in gunship design passed to the Soviet Union, which, in the Afghan War of the 1980s, fielded the Mil Mi-24 Hind, the fastest and possibly most capable helicopter gunship of its time (see Figure 43). A primary role of the Hind was to attack armoured vehicles; to this end, it mounted guided antitank missiles on stub wings projecting from the fuselage. In addition to the two-man cockpit configuration of the HueyCobra, it had a small passenger and cargo bay that gave it a limited troop-transport capability.

The Mi-24 Hind

The successor to the HueyCobra was the McDonnell Douglas AH-64 Apache, a heavily armoured antiarmour helicopter with less speed and range than the Hind but with sophisticated navigation, ECM, and fire-control systems. The Apache became operational in 1986. Soon afterward the Soviets produced the Mi-28 Havoc, a refinement of the Hind that, with no passenger bay, was purely a gunship.

Naval helicopters. Helicopters were used extensively in antisubmarine roles, "dipping" sonar sensors into the water to locate their targets and launching self-homing torpedoes to destroy them. Ship-borne helicopters also served as firing platforms for antiship missiles and were used to carry warning and surveillance radars, typically sharing information with their mother ships. By firing heat-producing or chaff flares to confuse infrared and radar homing systems, naval helicopters could serve as decoys for antiship missiles.

Remotely piloted vehicles. The first remotely piloted vehicles (RPVs) were small, pilotless aircraft controlled by command radio transmission. Most of these fell into one of two categories: extremely high-performance drones used to test new systems; and small, relatively inexpensive drones used for training. Both were typically reusable, being recovered by radio-controlled landing or, more commonly, by parachute. Target drones were commonly fitted with radar reflectors to stimulate the radar return of enemy aircraft, and it soon occurred to strategists to use them as decoys to assist bombers in penetrating enemy defenses. That modified target drones might be effective platforms for communications relay and for sensor and reconnaissance systems also became evident. The Ryan QM-34 Firebee, a photoreconnaissance modification of a standard U.S. target drone, saw extensive service in Vietnam. A swept-wing, turbojet-powered subsonic vehicle less than half the size of a jet fighter, the Firebee penetrated heavily defended areas at low altitudes with impunity by virtue of its small radar cross section and brought back strikingly clear imagery. Indicative of later development was the Boeing Compass Cope, a long-winged, subsonic, turbofan-powered drone capable of long flights at extremely high altitudes. (J.W.R.T./J.F.G.)

Warning and detection systems

Because military tactics from time immemorial have stressed the value of surprise—through timing, location of attack, route, and weight and character of arms—defenders have sought to construct warning systems to cope with all these tactics. Many types of warning systems exist. Long-term, or political, warning systems employ diplomatic, political, technological, and economic indicators to forecast hostilities. The defender may react by strengthening his defenses, by negotiating treaties or concessions, or by taking other action. Political warning, equivocal and incapable of disclosing fully an attacker's intention, often results in an unevaluated and neglected situation.

Medium-term, or strategic, warning, usually involving a time span of a few days or weeks, is a notification or judgment that hostilities may be imminent. Short-term, or tactical, warning, often hours or minutes in advance, is a notification that the enemy has initiated hostilities.

Warning and detecting are separate functions. The sensors or detection devices perceive either the attack, the possibilities of an attack, the nearness of the enemy, his location, his size, his activities, his weapon capability, or some changes in his political, economic, technical, or military posture.

Warning systems include detection devices but also imply the judgments, decisions, and actions that follow receipt of the sensor's information. Warning encompasses communications, analysis of information, decisions, and appropriate actions. Visual observation still remains important, supplemented by telescopes, cameras, heat-sensing devices, low-light-level devices, radar, acoustic, seismic, chemical, and nuclear detection devices. The product, or output, of these sensors is complicated and voluminous and requires computers to condense and summarize the data for the decision maker. Often, the most expensive portion and weakest link of the warning system is not the sensor but the communication and evaluation systems. Technology of all types is required in modern warning systems.

HISTORY

History abounds with examples of successful military surprises; examples of effective warning are difficult to find. Military training emphasized the value of surprise, stratagem, and deception, but the value of warning was long neglected. Flank and rear guards, to protect marching columns, patrols and scouts to locate the enemy, and sentries to guard camps, were of course used in war from earliest times. Animals were sometimes employed to detect the approach of an enemy; dogs and horses were especially favoured, though, according to the ancient historian Livy, the Romans used geese to detect the night attack of the Gauls on Rome in the 4th century BC. High ground, favourable for observation, was often supplemented by watchtowers, such as those placed along the Great Wall of China and on Hadrian's Wall in Britain.

Balloon reconnaissance and aerial photography

The observation balloon was an important technological advance. First used in warfare by the French in the late 18th century, primarily for offensive reconnaissance on the battlefield, its defensive possibilities were demonstrated in the American Civil War; in May 1863 a balloon of the army of the Potomac detected Lee's army moving from its camp across the Rappahannock to commence the Gettysburg campaign. Aerial photography had already been pioneered by the French and used in the War of Italian Independence (1859).

A balloon observer in the Spanish-American War of 1898 is credited with discovering an alternate route up San Juan Hill during the battle there. A few other successes are ascribed to such observation before the balloon was supplemented by the far more valuable airplane in World War I. Nevertheless, the balloon never fulfilled its potential as a warning device.

In sea warfare, warning and detection were equally neglected. As far back as the Minoan civilization of Crete, patrol ships were used, but mainly for offensive purposes. In later centuries, raised quarterdecks and lookout posts atop sailing masts were provided, but the beginnings of serious maritime detection technology did not come until the advent of the submarine.

The first observation airplanes

Binoculars, telescopes, the telegraph, and telephone were well established military equipment by 1914; the airplane, first used by the Italians in the Italo-Turkish War of 1911, showed its potential as an observation device at the Battle of the Marne. Radio communications provided the means to make air observations immediately available. Aerial combat became inevitable as each side tried to deny the other its aerial reconnaissance.

Searchlights, first used in the Russo-Japanese War (1904), saw large-scale use in World War I to detect dirigibles and aircraft on night bombardment missions. Flares were used to illuminate the battlefield between trenches to detect raiding parties. Listening devices, using directional horns to detect and locate enemy aircraft, were also used with limited success.

Despite the novelties of World War I, World War II produced far more technological innovation. Radar made obsolete the slow and inaccurate older listening devices. Radio communications made great strides, particularly in the very high frequency range. The combination of radar and interference-free very high frequency communications was pivotal in permitting the RAF to resist Hitler's aerial attack and win the Battle of Britain.

Notwithstanding radar sophistication, ground spotters played an important role in filling the gaps between radar coverage. Their messages, forwarded to a plotting centre, were assembled to trace the progress of intruders (tracking).

The advent of nuclear weapons (1945), especially when coupled later with the speed and range of intercontinental missiles, gave new dimensions to the value of surprise for the attacker. Long-term warning was suddenly of paramount importance. Not only did all forms of unequivocal warning become indispensable but the warning had to be made credible to an aggressor; that is, an assurance had to be given that the retaliatory weapons would not all be destroyed by a first strike. Bomber aircraft were kept in the air to avoid destruction on the ground and attempts were made to provide a degree of protection for the civilian population through shelters.

Practically all aspects of science and technology have been introduced into today's warfare and warning systems: airplanes, helicopters, submarines, earth satellites, television, lasers, and magnetic, acoustic, seismic, infrared, nuclear, and chemical detectors.

ELECTROMAGNETIC SENSORS

Modern detector technology. *The visible region.* Binoculars and telescopes have changed very little. Where vibration and motion create interference, gyroscopically stabilized optics are used in surface vehicles, ships, and aircraft.

Newer in character are the image intensifiers used for nighttime detection (see Figure 44). These devices receive the moonlight or starlight reflected from targets on a sensitive screen, amplify the image electronically, and present it at much higher light level on a small cathode-ray tube similar to that used in a television receiver. Typical of these devices is the starlight scope, resembling an oversized telescopic sight, with which riflemen can aim at night at 1,000–1,300 feet range. Artillery, tanks, helicopters, and aircraft use similar, larger devices having longer range. In aircraft the direct-viewing device is replaced by a cathode-ray tube in the instrument panel; this version is called low-light-level television.

The starlight scope

Ordinary searchlights can often be used at night even in combat situations; but, to avoid drawing fire, invisible light, in the ultraviolet or near infrared range, can be used with appropriate viewing devices. Conventional photography, used in aerial reconnaissance and essential to long-term warning, must have high resolution despite temperature and vibration interference. To cover wide

By courtesy of the United States Army

Figure 44: The night observation device, which provides a high degree of night vision capability to detect, locate, and identify the enemy. Using image intensification techniques, the low-light-level illumination of the night sky (*i.e.*, moonlight) reflected from the object and its background form a clearly defined image to the observer.

areas, panoramic cameras, scanning from side to side, record high-quality images. Frame cameras are also used, especially for mapping. At night, flares or flashing lights on aircraft are used.

Infrared. Infrared sensors on the ground, or in aircraft or spacecraft, can detect such hot spots as motor-vehicle engines, hot jet engines, missile exhausts, even campfires. They have good location accuracy and high sensitivity to signals, without registering such false targets as sun reflections.

In the very near infrared region, infrared imaging detectors use specially sensitized photographic film to reveal forms hidden by camouflage. More important are the detectors used in the far infrared region; objects at room temperature radiate sufficient energy for detection at ranges of several miles. Infrared imagery can have longer range than image intensifiers and can operate without starlight. When the humidity is high, the effective range is reduced.

The sniperscope, an early device that used infrared illumination and an infrared viewer, has been largely replaced by the image intensifier and by laser illuminators.

Radar. Radar is used by ground forces for many purposes: in portable sizes, for infiltration detection; in intermediate sizes, for mortar and artillery shell tracking; and in large sizes, for early warning, search, and control of air-defense weapons (interceptors and surface-to-air missiles).

Radar is used in fighter aircraft for finding enemy aircraft and controlling air-to-air missiles, rockets, and guns. It is used in bombers to find surface targets, fixed or moving, and to navigate and avoid obstacles. It is used in large aircraft as an airborne warning and control system, searching the skies over great distances for enemy aircraft, tracking them, and controlling interceptors. It also is used to search the seas for surface vessels or surfaced submarines. Radar also can be used in spacecraft to locate patterns of activity.

In all applications of radar, clutter in the form of reflections from surface objects or the terrain, or the disturbed sea, competes with reflection from the targets and must be cancelled by appropriate circuitry. Side-looking radars are used to obtain higher resolution than conventional radar, improving the ability to recognize surface targets.

Conventional radar operates at microwave and ultrahigh frequencies that propagate in straight lines like light rays; consequently, they cannot ordinarily detect objects beyond the Earth's horizon. Because high-frequency waves reflect from the ionosphere, over-the-horizon radars have been built to operate in this region.

Over-the-horizon radar

Radio sensors. Radio receivers can be used to detect and locate enemy radio. Enemy radars can be located in much the same way. Messages can be intercepted. This form of warning has been combated by radio silence and by spoofing, the transmission of signals intended to deceive. In 1967 the Israelis transmitted voluminous radio messages from empty airfields to hide the fact that aircraft had been moved to other locations.

Radio direction finders can be used to locate nuclear bursts, because the explosion generates a large amount of energy in the radio frequency region.

Acoustic techniques. While electromagnetic waves do not propagate well under water, acoustic waves do and can be used to detect submerged submarines. These detection systems, called sonar, may intercept propeller or other noise generated by the submarine or may send out sounds and receive echoes from the submarine hull. Sonar devices can be operated aboard surface ships, aboard submarines, on floating sonobuoys, or suspended by cables from helicopters and dunked in the ocean.

Sonar systems

Sonar systems are limited in range by attenuation (weakening) of the sound energy in water, bending caused by temperature differences in water layers, and extraneous noises, including reflections from the sea bottom.

Acoustic receivers are also used on land in sensors deployed near trails to detect the presence of personnel or vehicles along roads. The sounds are sent by radio to listening posts. Acoustic sensors are also used in monitoring nuclear explosions.

Seismic detectors. Seismic detectors (see Figure 45)—as well as underground acoustic detectors called geophones—are used in sensors for infiltration and vehicle detection.

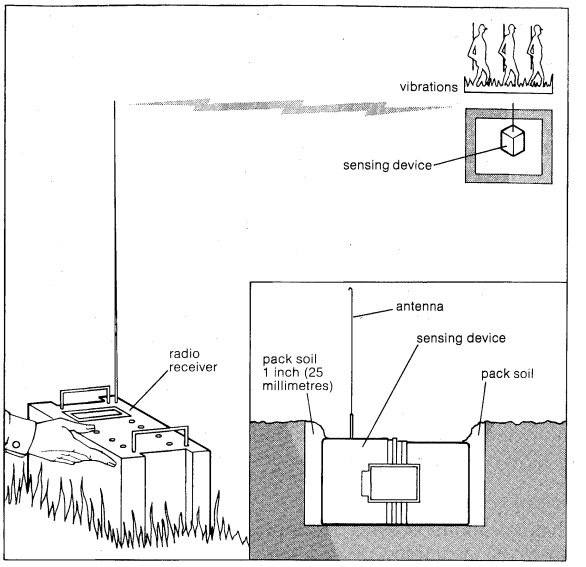


Figure 45: *Distant sensing and identification.* The acoustic sensing device is buried in the ground as shown. It detects movement of people and vehicles by sensing vibrations they cause in the earth and transmits a code to a special radio receiver.

Both types must be used, since either alone can yield false signals caused by the movement of animals.

Magnetic detectors. Sensitive magnetic detectors (magnetometers), flown in aircraft over the sea, are used to detect submarines because the large metallic mass of the submarine hull disturbs the earth's magnetic field. Buried on land, magnetometers are used to detect the passage of vehicles.

Nuclear detectors. Underground nuclear explosions are detected by sensitive seismometers. To increase the sensitivity and reject natural earth tremors, seismometers are often used in large arrays extending for hundreds of miles.

For atmospheric or space explosions, radio-pulse receivers and light flash and acoustic detectors are used, as well as devices to measure fallout. Aircraft and rockets can be used to collect radioactive debris, while high-altitude satellites carry detectors for gamma rays and other emissions.

Chemical sensors. Concealed chemical sensors, sensitive to minute amounts of body products, are capable of detecting personnel from short distances.

Future developments. Certain trends can be seen in sensor development for future warning systems. Infrared detectors of higher sensitivity and resolution are being developed. Higher-powered and smaller lasers will aid night warning systems. These and other lines of research, centred on lighter weight and more efficient optics and on more efficient detectors, should result in much cheaper systems with resolution approaching visual sensors. Perhaps most important are improvements in the resolution and brightness of the display—the chief limitation of most night viewing systems.

Photography has already reached an advanced state of technology, yet improvements in resolution are being actively pursued. Lightweight optics, more sensitive and fine-grained film, film that can be developed quickly by heat, and better compensation for the motion of the aircraft are some of the areas where photography can be improved.

Developments in large ground radars centre around the phased array radar having electronically steered beams. The beams must be computer-controlled. Moving target discrimination and Doppler processing are built out of digital circuitry as used in digital computers. This permits sensitive discrimination and rapid response.

The advent of the transistor and solid-state microcircuits is making small radars for infantrymen and tank operators possible. The miniature components and high reliability of these devices makes extremely complex and sophisticated circuitry possible.

Airborne radar is in its early stages of development. Side-looking Doppler-processing radar has already yielded high

resolution, but not quite as good as conventional photography. Developments in progress indicate that soon images comparable to photographic images will be obtained from airborne radar. Imperfections now common in radar imagery should be removed as a result of present research.

A great deal of effort in several countries has reduced the vulnerability of radars to electronic countermeasures; at the same time, however, similar improvements in electronic jamming and deception have taken place.

Nuclear propulsion enables submarines to remain submerged and escape detection by radar. This, plus its increased speed, makes the nuclear submarine a formidable threat. To combat this, sonar sensors for detection of submarines are now being formed into arrays. This increases the sensitivity and rejects extraneous noise, especially important in regions of turbulence.

The search for more sensitive systems of detection will go on. Measurement of the temperature change in the water in which a submarine lies and the magnetic anomaly observable when it is under the water are two directions in which study is being pursued. A range of such measurements may become possible. Testing of laser beams for underwater recognition capability has been proceeding for some time. The problem is extremely difficult, water being a medium quite different from air, and much work will be needed to overcome the obstacles.

Underwater laser beams

The subject is closely linked with the study of undersea conditions generally; that is, oceanography. American efforts dwarf those of any other Western nation, though France, a pioneer in undersea exploration, is active. Underwater acoustic navigation enables ships to be used for missile or satellite tracking. Underwater communication over very long distances is essential for the control of nuclear submarines, to make the most of their almost unlimited radius of operation.

WARNING SYSTEMS

Modern systems. *Air defense systems.* Radar and identification friend or foe (IFF) equipment constitute the forward elements of complex systems that have appeared throughout the world. Examples include the semiautomatic ground environment (SAGE), augmented by a mobile backup intercept control system called BUIC in the United States, NATO air defense ground environment (NADGE) in Europe, a similar system in Japan, and various land-mobile, airborne, and ship command and control systems. Little information concerning the Soviet systems is available, but they are known to be extensive, automated, and capable.

Air-defense systems require computers and communication nets to process the radar data. Position reports from the radars are formed into tracks of each detected aircraft. Height-finding radars add the third dimension. The IFF information, together with known flight plans, is correlated; clutter, false returns from clouds, and any electronic countermeasures are rejected. Decisions are made on whether to counter the attack with interceptors or surface-to-air missiles. The counterattack is controlled by guiding a missile or directing an intercept.

To avoid excessive centralization of equipment that would make the system vulnerable to nuclear attack, the computers and communication facilities are widely dispersed and supplemented by mobile facilities.

In addition to large conventional radars, small distributed radars (called gap fillers) are used to detect low-flying aircraft penetrating gaps in large radar coverage. Over-the-horizon radars and AWACS (airborne warning and control systems) are even more promising. The latter consist of large radar and computation, display, and control systems, housed in large aircraft. First introduced for naval defense, they have become potentially effective over land with new developments in clutter-rejection circuitry.

Large aircraft with powerful radars connected to sophisticated computer and display equipment can survive a nuclear attack and have a low-altitude surveillance capability. Their use, delayed because of problems caused by interference from land clutter, is growing.

A unique air-defense system is the U.S.-Canadian Distant Early Warning system stretching across the northern portion of North America. The radars are used strictly for early warning; no control of missiles or interceptors is provided. Elaborate communications to control centres to the south are part of the system.

The DEW line

Air-defense systems spread the warning to the civil population by sirens and radio alerts. Extensive communication nets are built for this purpose. Air-defense systems also select and assign the defensive weapons to particular threats. If interceptors are used, a control centre is assigned to send control information by digitally encoded radio messages.

If surface-to-air missiles are used, the target is designated to the missile control system, which has its own target-tracking and missile-control radar. Practically all surface-to-air missile systems have some autonomous capability of warning and target acquisition. Examples of these systems are the American Nike Hercules and Hawk, the British Thunderbird, Bloodhound, and Rapier, the French-German Roland, and the Italian Indigo. In sea warfare, such missiles as the U.S. Terrier and Talos, the British Sea Dart, and the French Masurca have autonomous radar capability.

At sea, air defense also uses large radars on ships, but more use is made of airborne radar and control systems. The weight and size of long-range radars restricts their installation to the larger ships; airborne radar over the ocean does not have severe land clutter to contend with, making it simpler than overland systems; the horizon limits are at a greater range; and the aircraft can patrol a large area. As in land defenses, extensive computer and display complexes, and communications between the ships, are used. In the U.S. Navy the Airborne Tactical Data System, consisting of airborne radar, computers, and memory and data links, is connected with the Naval Tactical Data System, located in fleet headquarters, which processes, organizes, and displays information of the overall picture of the tactical situation.

Ballistic missile warning. In the second half of the 20th century, warning against ballistic missiles with nuclear warheads has taken precedence over all other warning systems. Large ground radars, operating in the very high frequency (VHF) or ultrahigh frequency (UHF) range, are used. The radars search the skies and track detected objects. Computers calculate trajectory to determine if the target is a missile or an Earth-orbiting object. Depending on the trajectory, the number of objects, and other criteria, alerts, tentative warnings, or all-out warning signals are transmitted to command centres.

Surface-based radars have one serious flaw: they can detect an object only after it appears above the Earth's horizon. For earlier warning, over-the-horizon radars or satellite-borne infrared detectors can be used.

There are two types of over-the-horizon radars, operating in the high frequency range, which can reflect from the ionosphere. One system, called forward scatter, transmits from one location and receives the signal several thousand miles away on the other side of the launch point. The back-scatter system receives the signal from the same location as the transmitter, as is done in conventional radar. Both systems detect variations in the received signal due to fluctuations in the ionosphere caused by the missile's exhaust plume as it traverses the ionosphere.

Ballistic missile defense. Ballistic missile defense systems have their own warning and acquisition radar systems. These large radars are more sophisticated than the warning radars because they must form accurate tracks for the engagement radars. Decoy objects and lightweight metallic reflectors called chaff must be identified and rejected. To do this, the radars must be able to measure the velocity of all the objects, because lightweight objects decelerate more rapidly than heavy objects due to atmospheric drag and friction.

Space surveillance. Closely allied to warning systems are space-object detection and tracking systems. It is likely that only the United States and the Soviets have developed and operate these systems. A variety of very large radars are used, although the newer installations are phased-array radars that have stationary antennas with electronically steerable multiple beams. The scanning is more rapid than that by a mechanically rotated antenna, and

several objects can be tracked simultaneously. The radars used for ballistic missile early warning are connected into spacetrack nets.

Tracking by telescope

To supplement radars, telescopes have been designed for accurate tracking of comparatively low earth satellites. Telescopes, which can have cameras, have been adapted with varying degrees of success to pick up high-altitude satellites and extremely faint objects. The range depends on the size of the target, its reflectivity, and the solar aspect angle (angular position of the sun in the sky). Telescopes are not detection devices, but they can track objects if they are pointed in the correct direction by the ground radar net.

Detection of nuclear explosions. In 1963 a treaty banning nuclear weapon tests in the atmosphere, in outer space, and underwater was signed. Each signatory nation was to provide monitoring. A direct consequence was the development and construction of a wide variety of devices to monitor nuclear explosions.

Underground explosions, still permitted under the treaty, are monitored by seismometers, instruments that measure minute ground motions. Because of the high sensitivity required to measure at great distances the ground vibrations caused by nuclear explosions, the seismometers record many extraneous motions from natural sources; these are called noise. To reduce noise, a large number of seismometers arranged in arrays is used to reinforce the desired signal and exclude unwanted signals. Elaborate data processing, with the help of recorders and computers, further refines the output. Despite these measures, there is a limit to the sensitivity of underground and underwater systems, so that very small nuclear explosions at great distance from the receiving sites may not be detected or may be wrongly identified as a small earthquake.

Sensing an explosion in space

Detection of explosions in the atmosphere and in space depends upon measuring the products of an explosion. Acoustic sensors are used to measure the sound waves created by the blast, aircraft and rockets to collect possibly radioactive debris samples, flash detectors to detect the light flash as well as the radio pulse generated by the explosion, and a number of radio-detection techniques to measure the considerable disturbance of the ionosphere. None of the techniques is adequate by itself, since each is disturbed by various background signals. Analyzed together, however, they yield positive results.

To detect explosions in space, high-altitude satellites are used. They carry detectors of X-ray emissions, gamma rays, and neutrons, all of which are generated by a nuclear explosion. They can be detected because there is essentially no atmosphere in space to absorb the emissions.

Infiltration and base defense systems. The growth of insurgency warfare has made necessary the development of a variety of sensors to detect vehicles and personnel in the jungle along trails or on roads. Acoustic, seismic, magnetic, infrared, radar, and Doppler radar (radars that detect movement by shift in frequency of received signal) are the sensors.

The sensors are connected to processing centres where the progress of an infiltrating column or truck convoy can be monitored. This process eliminates many false detections due to random noise or animals. Because the sensors are widespread and the processing quite sophisticated, the systems have become known as the instrumented battlefield or electronic barrier.

Aerial reconnaissance. Aerial reconnaissance has grown in importance; it now encompasses all phases of warning. Visual observation from the air furnishes short-term information and warning. Direct receiving and image-recording infrared equipment in night reconnaissance, high resolution radar in bad weather, and conventional photography all contribute to medium and long-term warning by observing tactical preparations or discerning new military capabilities.

Manned aircraft are used more frequently than other platforms for these sensors. Unmanned aircraft, however, flying at low and high altitudes; helicopters, including small unmanned helicopters; and space vehicles are all used for various reconnaissance missions.

Photography from rockets was first undertaken in 1906. A model for military reconnaissance was built in 1912, but by this time photography from airplanes had been shown to be feasible. After the launching of the first Soviet satellite, Sputnik 1, in 1957, the potential of observations from space vehicles became obvious and various applications were developed.

Detection from a satellite

Satellite platforms can carry a variety of sensors. Cameras in space can collect images on photographic film, infrared images, or television-type signals. Radars can be carried aloft for operation at night or through clouds that could otherwise obscure the images. Infrared sensors can be used to detect missiles, or space warnings. Sensors to detect nuclear explosions can also be used to monitor possible violations of the nuclear test treaty.

To be useful, the sensors must have high resolution. The large distances involved make this difficult. Cameras must have telescopic optics and must be quite large and heavy. As the ability to lift larger weights to orbital altitudes increases, the capabilities of the sensors will improve. Infrared sensors also need heavy equipment. Radar sensors are limited not only in resolution (generally much poorer than optical sensors) but by electrical power limitations, since quite powerful radar transmitters are necessary.

Photographic resolution of about one second of arc is achievable today. At 200 miles (320 kilometres) altitude, this would be equivalent to a resolution of 10 feet (three metres); that is, an object 10 feet in diameter could be clearly distinguished. Vibration and high speed reduce this resolution considerably.

Antisubmarine systems. The limited range of both active (echo-ranging) and passive (listening) sonar makes the use of many sensors necessary in submarine detection. To guard a shore, a line of sensors can be set on the ocean floor. In the broad ocean area, however, the sensors on ships and submarines leave vast spaces uncovered. To fill these gaps, sonobuoys, floating buoys with sonar sensors and radio transmitters, are used. The signals from the sonobuoys are received by patrolling aircraft; these then track the submarines.

Naval vessels use helicopters for submarine detection and warning. Each carries a sonar sensor at the end of a cable, lowering it into the water to detect submarines. Such sensors are called dunked sonar sensors. (Ha.D.)

Military communication

Military communication, or signaling, has long played an important role in warfare. It provides the means for transmitting information from reconnaissance and other units in contact with the enemy and the means for exercising command by transmitting the orders and instructions of commanders to their subordinates. It comprises all means of transmitting messages, orders, and reports, both in the field and at sea and between headquarters and distant installations or ships.

EARLY DEVELOPMENT

Messengers have been employed in war since ancient times and still constitute a valuable means of communication. Alexander, Hannibal, and Caesar each developed an elaborate system of relays by which messages were carried from one messenger post to another by mounted messengers traveling at top speed. They were thus able to maintain contact with their homelands during their far-flung campaigns and to transmit messages with surprising speed. Genghis Khan at the close of the 12th century not only emulated his military predecessors by establishing an extensive system of messenger posts from Europe to his Mongol capital but also utilized homing pigeons as messengers. As he advanced upon his conquests he established pigeon relay posts across Asia and much of eastern Europe. He was thus able to use these messengers to transmit instructions to his capital for the governing of his distant dominions. Before the end of the 18th century European armies used the visual telegraph system devised by Claude Chappe, employing semaphore towers or poles with movable arms. The Prussian army in 1833 assigned such visual telegraph duties to engineer troops.

Genghis Khan's homing pigeons

At the same time that these elementary methods of signal

communication were being evolved on land, a comparable development was going on at sea. Early signaling between naval vessels was by prearranged messages transmitted by flags, lights, or the movement of a sail. Codes were developed in the 16th century that were based upon the number and position of signal flags or lights or on the number of cannon shots. In the 17th century the British admiral Sir William Penn and others developed regular codes for naval communication; and toward the close of the 18th century, Admiral Richard Kempenfelt developed a plan of flag signaling similar to that now in use. Later Sir Home Popham increased the effectiveness of ship-to-ship communication by improved methods of flag signaling.

THE ADVENT OF ELECTRICAL SIGNALING

Despite the early pioneering efforts on land and sea the real development of signal communication in war did not come until after invention of the electric telegraph by Samuel F.B. Morse. In his successful demonstration of electric communication between Washington, D.C., and Baltimore in 1844, he provided a completely new means of rapid signal communication. The development of the Morse Code of dots and dashes used with key and sounder was soon used to augment the various means of visual signaling. Vice Admiral Philip Colomb's flash signaling, adopted in the British navy in 1867, was an adaptation of the Morse code to lights. The first application of the telegraph in time of war was made by the British in the Crimean War in 1854, but its capabilities were not well understood, and it was not widely used. Three years later, in the Indian Mutiny, the newly established telegraph, which was controlled by the British, was a deciding factor.

Telegraph in the American Civil War

In the American Civil War (1861–65), wide use was made of the electric telegraph. In addition to its employment in spanning long distances under the civilian-manned military telegraph organization, mobile field service was provided in the Union army by wagon trains equipped with insulated wire and lightweight poles for the rapid laying of telegraph lines. Immediately before and during the Civil War visual signaling also received added impetus through development of a system, applying the Morse code of dots and dashes, that spelled out messages with flags by day and lights or torches by night. Another development for light signaling placed a movable shutter, controlled by a key, in front of a strong light. An operator, opening and closing the shutter, could produce short and long flashes to spell out messages in Morse code.

Simultaneously, the Prussian and French armies also organized mobile telegraph trains. During the short, decisive Prussian campaign against Austria in 1866, field telegraph enabled Count Helmuth von Moltke, the Prussian commander, to exercise command over his distant armies. Soon afterward the British organized their first field telegraph trains in the Royal Engineers.

Another instrument was added to the techniques for visual signaling through the development of the heliograph. It employed two adjustable mirrors so arranged that a beam of light from the sun could be reflected in any direction. The beam was interrupted by a key-operated shutter that permitted the formation of the dots and dashes of the Morse code. Where climatic conditions were favourable this instrument found much use, notably by the British army in India and the U.S. Army in the American Southwest. Because consistency and regularity of sunshine were important, the heliograph was never widely adopted throughout the armies of continental Europe.

The invention of the telephone in 1876 was not followed immediately by its adoption and adaptation for military use. This was probably due to the fact that the compelling stimulation of war was not present and to the fact that the development of long-distance telephone communication was not achieved for many years. The telephone was used by the U.S. Army in the Spanish-American War, by the British in the South African (Boer) War, and by the Japanese in the Russo-Japanese War. This military use was not extensive, and it made little material contribution to the development of voice telephony. Before the outbreak of World War I, military adaptation of the telephone did take place, but its period of growth had not yet arrived.

Near the close of the 19th century, a new means of military signal communication made its appearance—the wireless telegraph, or radio. The major powers throughout the world were quick to see the wonderful possibilities for military and naval signaling. Development was rapid and continuous, and, by 1914, it was adopted and in extensive use by all the armies and navies of the world. It soon became apparent that wireless telegraphy was not an unmixed blessing to armies and navies, because it lacked secrecy and messages could be heard by the enemy as well as by friendly forces. This led to the development of extensive and complicated codes and ciphers as necessary adjuncts to military signaling. The struggle between the cryptographer and the cryptanalyst expanded greatly with the adoption of radio and continued to be a major factor affecting its military use.

FROM WORLD WAR I TO 1940

The onset of World War I found the opposing armies equipped to a varying degree with modern means of signal communication but with little appreciation of the enormous load that signal systems must carry to maintain control of the huge forces that were set in motion. The organization and efficiency of the armies varied greatly. At one end of the scale was Great Britain, with a small but highly developed signal service; and at the other end stood Russia, with a signal service inferior to that of the Union Army at the close of the American Civil War. The fact that commanders could not control, coordinate, and direct huge modern armies without efficient signal communication quickly became apparent to both the Allies and the Central Powers. The Germans, despite years of concentration on the Schlieffen Plan, failed to provide adequately for communication between higher headquarters and the rapidly marching armies of the right wing driving through Belgium and northern France. This resulted in a lack of coordination between these armies, which caused a miscarriage of the plan, a forced halt in the German advance, and the subsequent withdrawal north of the Marne. On the Allied side, the debacle of the Russian forces in East Prussia—a crushing defeat at the hands of General Paul von Hindenburg in the Battle of Tannenberg—was largely due to an almost total lack of signal communication.

As the war progressed there was a growing appreciation of the need for improved electrical communications of much greater capacity for the larger units and of the need within regiments for electrical communications, which had heretofore been regarded as unessential and impractical. Field telephones and switchboards were soon developed, and those already in existence were improved. An intricate system of telephone lines involving thousands of miles of wire soon appeared on each side. Pole lines with many crossarms and circuits came into being in the rear of the opposing armies, and buried cables and wires were laid in the elaborate trench systems leading to the forwardmost outposts. The main arteries running from the rear to the forward trenches were crossed by lateral cable routes roughly parallel to the front.

The telephone in World War I

Thus, there grew an immense gridwork of deep buried cables, particularly on the German side and in the British sectors of the Allied side, with underground junction boxes and test points every few hundred yards. The French used deep buried cable to some extent but generally preferred to string their telephone lines on wooden supports set against the walls of deep open trenches. Thus electrical communication in the form of the telephone and telegraph gradually extended to the smaller units until front-line platoons were frequently kept in touch with their company headquarters through these mediums.

Despite efforts to protect the wire lines, they were frequently cut at critical times as the result of the intense artillery fire. This led all the belligerents to develop and use radio (wireless) as an alternate means of communication. Prewar radio sets were too heavy and bulky to be taken into the trenches, and they also required large and highly visible aerials. Radio engineers of the belligerent nations soon developed smaller and more portable sets powered by storage batteries and using low, inconspicuous aerials. Although radio equipment came to be issued to

the headquarters of all units, including battalions, the ease of enemy interception, the requirements for cryptographing or encoding messages, and the inherent unreliability of these early systems caused them to be regarded as strictly auxiliary to the wire system and reserved for emergency use when the wire lines were cut. Visual signaling returned to the battlefield in World War I with the use of electric signal lamps. Pyrotechnics, rockets, Very pistols, and flares had a wide use for transmitting prearranged signals. Messenger service came to be highly developed, and motorcycle, bicycle, and automobile messenger service was employed. Homing pigeons were used extensively as one-way messengers from front to rear and acquitted themselves extremely well. Dogs were also used as messengers and, in the German army, reached a high degree of efficiency.

A new element in warfare, the airplane, introduced in World War I, immediately posed a problem in communication. During most of the war, communication between ground and air was difficult and elementary. To make his reports the pilot had to land or drop messages, and he received instructions while in the air from strips of white and black cloth called "panels" laid out in an open field according to prearranged designs. Extensive efforts were made to use radiotelegraph and radiotelephone between the airplanes and ground headquarters. The closing stages of the war saw many planes equipped with radio, but the service was never satisfactory or reliable and had little influence on military operations.

Radio-equipped aircraft

During World War I, wireless telegraph communication was employed extensively by the navies of the world and had a major influence on the character of naval warfare. High-powered shore and ship stations made wireless communication over long distances possible.

One of the war lessons learned by most of the major nations was the compelling need for scientific research and development of equipment and techniques for military purposes. Although the amount of funds devoted to military development during the period from World War I to World War II was relatively small, the modest expenditures served to establish a bond between industry, science, and the armed forces of the major nations.

Of great importance in postwar radio communication was the pioneering by amateurs and by industry and science in the use of very high frequencies. These developments opened up to the armed services the possibilities of portable short-range equipment for mobile and portable tactical use by armies, navies, and air forces. Military work in these fields was carried out actively in Germany, Great Britain, and the United States. As early as 1938 Germany had completed the design and manufacture of a complete line of portable and mobile radio equipment for its army and air force.

Between World Wars I and II the printing telegraph, commonly known as the teleprinter or teletypewriter machine, came into civilian use and was incorporated in military wire-communication systems, but military networks were not extensive. Before World War II, military radioteleprinter circuits were nonexistent.

Another major communication advance that had its origin and early growth during the period between World Wars I and II was frequency-modulated (FM) radio. Developed during the late 1920s and early 1930s by Edwin H. Armstrong, an inventor and a major in the U.S. Army Signal Corps during World War I, this new method of modulation offered heretofore unattainable reduction of the effect of ignition and other noises encountered in radios used in vehicles. It was first adapted for military use by the U.S. Army, which, prior to World War II, had under development tank, vehicular, and man-pack frequency-modulated radio transmitters and receivers.

FM radio

On the eve of World War II, all nations employed generally similar methods for military signaling. The messenger systems included foot, mounted, motorcycle, automobile, airplane, homing pigeon, and the messenger dog. Visual agencies included flags, lights, panels for signaling airplanes, and pyrotechnics. The electrical agencies embraced wire systems providing telephone and telegraph service, including the printing telegraph. Both radiotelephony and radiotelegraphy were in wide use, but radiotelephony had not as yet proved reliable and satisfactory for tactical military communication. The navies of the world entered World War II with highly developed radio communication systems, both telegraph and telephone, and with development under way of many electronic navigational aids. Blinker-light signaling was still used. The use of telephone systems and loud-speaking voice amplifiers on naval vessels had also come into common use. Air forces employed wire and radio communication to link up their bases and landing fields and had developed airborne long-range, medium-range, and short-range radio equipment for air-to-ground and air-to-air communication.

WORLD WAR II AND AFTER

In communications electronics, World War II was in one sense similar to World War I: the most extravagant prewar estimates of military requirements soon proved to represent only a fraction of the actual demand. The need for all kinds of communication equipment and for improved quality and quantity of communications pyramided beyond the immediate capabilities of industry. An increase in manufacturing plant became vital, and research and development in the communications–electronics field was unprecedented. The early German blitzkrieg, with tank and armoured formations, placed a new order of importance on reliable radio communication.

The development of the air, infantry, artillery, and armoured team created new requirements for split-second communication by radio among all members. Portable radio sets were provided as far down in the military echelons as the platoon. In every tank there was at least one radio and in some command tanks as many as three. Multiconductor cables were provided wire communications; they could be reeled out rapidly and as many as four conversations could take place on them simultaneously through the use of carrier telephony. The Germans were the first to use this type of military long-range cable, and their example was followed promptly by both the British and the U.S. forces. High-powered mobile radio sets became common at division and regimental level. With these sets telegraph communication could be conducted at distances of more than 100 miles (160 kilometres) with vehicles in normal motion on the road. Major telephone switchboards of much greater capacity were needed. These were developed, manufactured, and issued for use at all tactical headquarters to satisfy the need for the greatly increased number of telephone channels required to coordinate the movements of field units whose mobility had been expanded many times.

Mobile communications

Radio relay, born of the necessity for mobility, became the outstanding communication development of World War II. Sets employing frequency modulation and carrier techniques were developed and used, as were also radio relay sets that used radar pulse transmission and reception techniques and multiplex time-division methods for obtaining many voice channels from one radio carrier. Radio relay telephone and teletypewriter circuits spanned the English Channel for the Normandy landing and later furnished important communication service for General George S. Patton, after his breakout from the Normandy beachhead.

The need for communication between the homelands and many far-flung theatres of war gave rise to the need for improved long-range overseas communication systems. A system of radioteletypewriter relaying was devised, by which a radioteletypewriter operator in Washington, London, or other capitals could transmit directly by teleprinter to the commander in any theatre of war. In addition, a system of torn-tape relay centres was established so that tributaries could forward messages through the major centres and retransmit quickly by transferring a perforated tape message from the receiving to the transmitting positions. In addition a system of holding teletypewriter conferences was developed. These conferences, called "telecons," enabled a commander or his staff at each end to view on a screen the incoming teletypewriter messages as fast as the characters were received. Questions and answers could be passed rapidly back and forth over the thousands of miles

separating the Pentagon in Washington, D.C., for example, from the supreme Allied headquarters in Europe or General Douglas MacArthur's headquarters in the Far East.

Electronic detection, navigation, and guidance

During the latter years of the war, new and improved communication and electronic devices came forth from research and development in ever-increasing numbers. A new long-range electronic navigation device, known as loran, used for both naval vessels and aircraft, was developed, as were short-range navigational systems, called shoran. Combinations of radar and communications for the landing of aircraft in zero visibility were perfected. One such system was the GCA, or ground-controlled approach system. Combinations of radio direction-finding, radar, and communications systems were developed and used for ground control of intercept aircraft—the system called GCI (ground-controlled intercept). Radio-controlled guidance of falling bombs enabled an operator in a bomber to direct a bomb to the target. Electronic countermeasures made their appearance in the form of jamming transmitters to jam radio channels and radar, navigation, and other military electronics.

The military services learned well from their wartime experiences the importance of scientific research and development in all fields, including communications electronics. Advances were made in the communication capacity of wire and radio relay systems and in improved electronic aids for navigation. Measures to provide more comprehensive and more reliable communication and electronic equipment continued to be stressed in the armies, navies, and air forces of the major powers.

After mid-century, accordingly, military efforts in all the many facets of signal communication continued to intensify almost as extravagantly as during World War II. Two major additions in the U.S. Army were television and "electronic brain" equipment. The latter, in many forms of digital and analog computers and of such data-processing devices as punch-card machines, were applied increasingly to personnel record handling and to depot and supply operations interconnected over wide areas by signal-communication networks.

Television proved a valuable training aid in military schools, where mass instruction, especially in manual skills, was needed and where instructors were few. A single instructor could teach many small classes simultaneously, each grouped before a TV set where they could watch demonstrations closely. Two-way communication permitted the instructor to call and question any student in any classroom and enabled any student to put questions to the instructor. Portable television equipment in the field proved valuable for sending back to headquarters, by antenna radiation or coaxial cable, a picture of any scene of operations such as a river crossing. Equally valuable was a television camera in the hands of a forward scout or in a reconnaissance aircraft, whether piloted or remotely controlled, to scan enemy territory.

Thus signal communication, combining in itself the powers of photography, television, radar, and other instruments using the electromagnetic radiation spectrum, moved into such new areas of military electronics as battle area surveillance and electronic warfare devices to interfere with, or jam, enemy transmitters. In the U.S. Army, battle area surveillance radically augmented conventional reconnaissance methods. An electronically controlled target acquisition system, to discover enemy troops or transport on the ground or in the air, was being developed using optical, sonic, photographic, infrared, and radar equipment. The aggregate of information gathered by these devices over a wide enemy front can be assembled electronically and displayed at headquarters where the combat commander can quickly estimate the situation and make tactical decisions. (G.I.Bk./G.R.T.)

BIBLIOGRAPHY

Military technology before the modern era. ARTHER FERRILL, *The Origins of War: From the Stone Age to Alexander the Great* (1985), offers a scholarly survey of evidence for prehistoric war. For a glimpse of the vestiges of prehistoric tradition in modern times, see the article "Ancient World of a War-Torn Tribe: Willigiman-Wallalua People," *Life*, 53(13):73–91 (Sept. 28, 1962), which contains dramatic photographs of New Guinea tribesmen skirmishing with javelins. JAMES D. MUHLY, "How Iron Technology Changed the Ancient World and Gave the Philistines a Military Edge," *The Biblical Archaeology Review*, 8(6):40–54 (November/December 1982), analyzes early metalworking techniques and their role in the beginnings of military technology. YIGAEL YADIN, *The Art of Warfare in Biblical Lands in the Light of Archaeological Study*, 2 vol. (1963; originally published in Hebrew, 1963), applies an operational perspective to the development of military technology and tactics in ancient and early classical time, but some of the author's conclusions need to be modified in light of subsequent archaeological discoveries. P.A.L. GREENHALGH, *Early Greek Warfare: Horsemen and Chariots in the Homeric and Archaic Ages* (1973), offers a revisionist analysis of the use of horses in war. J.G. LANDELS, *Engineering in the Ancient World* (1978), presents a comparative study of human and animal muscle, wind, and falling water as sources of power and has separate chapters on land and sea transport and catapult design. A.M. SNODGRASS, *Arms and Armour of the Greeks* (1967), is a brief but scholarly work on the subject. F.E. ADCOCK, *The Greek and Macedonian Art of War* (1957, reprinted 1967); and E.W. MARSDEN, *Greek and Roman Artillery: Historical Development* (1969), remain standard treatments. WERNER SOEDEL and VERNARD FOLEY, "Ancient Catapults," *Scientific American*, 240(3):150–160 (March 1979), is an illustrated technical analysis. EDWARD N. LUTTWAK, *The Grand Strategy of the Roman Empire From the First Century A.D. to the Third* (1976), devotes considerable attention to technical aspects of Rome's military power. J.F.C. FULLER, *The Generalship of Alexander the Great* (1958, reissued 1981), remains a standard historical treatment of the organization and use of the army; it can be supplemented by the brilliant examination in DONALD W. ENGELS, *Alexander the Great and the Logistics of the Macedonian Army* (1978). Brief historical surveys of equipment and dress, based on artistic interpretations of archaeological evidence, are presented in the works from the "Men-at-Arms" series: TERENCE WISE, *Ancient Armies of the Middle East* (1981), and *Armies of the Carthaginian Wars, 265–146 BC* (1982); NICK SEKUNDA, *The Ancient Greeks: Armies of Classical Greece, 5th and 4th Centuries BC* (1986), and *The Army of Alexander the Great* (1984); and MICHAEL SIMKINS, *The Roman Army From Caesar to Trajan*, rev. ed. (1984).

MICHEL BUR, "The Social Influence of the Motte-and-Bailey Castle," *Scientific American*, 248(5):132–140 (May 1983), describes the early beginnings of the European feudal system of war; and JOHN BEELER, *Warfare in Feudal Europe, 730–1200* (1971), summarizes the subject. BERNARD S. BACHRACH, "Animals and Warfare in Early Medieval Europe," in the symposium *L'Uomo di fronte al mondo animale nell'alto Medioevo*, vol. 1, pp. 707–751 (1985), analyzes textual, graphic, and archaeological evidence on the development of the technology of mounted shock action in western Europe; the same author's *Merovingian Military Organization, 481–751* (1972) examines the social context within which the new technology developed. A.V.B. NORMAN, *The Medieval Soldier* (1971); and A.V.B. NORMAN and DON POTTINGER, *English Weapons & Warfare, 449–1660* (1979, reissued 1985), address weaponry in detail. CHARLES FFOULKES, *Armour & Weapons* (1909, reprinted 1973), remains a classic treatment of the development of personal armour in Europe.

V.J. PARRY and M.E. YAPP (eds.), *War, Technology, and Society in the Middle East* (1975), contains a number of essays dealing with various aspects of warfare and military technology during the age of cavalry. DAVID C. NICOLLE, *Arms and Armour of the Crusading Era, 1050–1350*, 2 vol. (1988), is a comprehensive reference work. FRANK A. KIERMAN, JR., and JOHN K. FAIRBANK (eds.), *Chinese Ways in Warfare* (1974), is a collection of scholarly essays. S.R. TURNBULL, *The Mongols* (1980); and DAVID C. NICOLLE, *The Armies of Islam, 7th–11th Centuries* (1982), are two works from the "Men-at-Arms" series. RALPH PAYNE-GALLWEY, *A Summary of the History, Construction, and Effects in Warfare of the Projectile-Throwing Engines of the Ancients with a Treatise on the Structure, Power, and Management of Turkish and Other Oriental Bows of Mediaeval and Later Times* (1907, reprinted 1973), though old, contains valuable reports of practical experiments in the reconstruction of mechanical artillery and informed speculation on the capabilities of the composite bow. W.F. PATERSON, "The Archers of Islam," *Journal of the Economic and Social History of the Orient*, 9(1–2):69–87 (November 1966), offers technical and tactical analysis of medieval archery. PAUL E. KLOPSTEG, *Turkish Archery and the Composite Bow: A Review of an Old Chapter in the Chronicles of Archery, and a Modern Interpretation*, 3rd enlarged ed. (1987), contains the most complete technical description of composite bow construction, based on an 1847 Ottoman text written during the final revival of Turkish archery. RALPH PAYNE-GALLWEY, *The Crossbow, Mediaeval and Modern, Military and Sporting: Its Construction, History & Management, with a Treatise on the Balista and Catapult*

of the Ancients, and an Appendix on the Catapult, Balista & the Turkish Bow, 2nd ed. (1958, reprinted 1981), remains the basic text on the crossbow; ROBERT HARDY, *Longbow: A Social and Military History* (1976, reissued 1986), is a definitive treatment by a famous archer, with appendixes on design and ballistics by technical experts. The above-mentioned "Men-at-Arms" series includes DOUGLAS MILLER, *The Landsknechts* (1976) and *The Swiss at War, 1300–1500* (1979); DAVID C. NICOLLE, *Armies of the Ottoman Turks, 1300–1774* (1983); and S.R. TURNBULL, *Samurai Armies, 1550–1615* (1979). Though not primarily concerned with warfare or military technology, NORMAN ITZKOWITZ, *Ottoman Empire and Islamic Tradition* (1972, reprinted 1980); and PETER DUUS, *Feudalism in Japan,* 2nd ed. (1976), provide the institutional and cultural context for military developments.

DAVID AYALON, *Gunpowder and Firearms in the Mamluk Kingdom: A Challenge to a Mediaeval Society,* 2nd ed. (1978), studies a military elite that failed to adapt to gunpowder. CHARLES OMAN, *A History of the Art of War in the Middle Ages,* 2nd rev. ed., 2 vol. (1924, reissued 1959), and *A History of the Art of War in the Sixteenth Century* (1937, reprinted 1979), provide encyclopedic coverage and analytical treatment. F.L. TAYLOR, *The Art of War in Italy, 1494–1529* (1921, reprinted 1973), is a useful supplement. JACOB DE GHEYN, *The Exercise of Armes: A Seventeenth Century Military Manual,* new ed., edited by DAVID J. BLACKMORE (1986; originally published in Dutch, 1607), is one of the earliest published detailed books of musket drill. PHILIP J. HAYTHORNTHWAITE, *Weapons and Equipment of the Napoleonic Wars* (1979), provides comprehensive coverage of the subject. KENNETH MACKSEY, *The Guinness History of Land Warfare* (1973; also published as *The History of Land Warfare,* 1974), is a useful reference source. CHRISTOPHER DUFFY, *The Military Experience in the Age of Reason* (1987), is a comprehensive treatment of European methods of war. GEOFFREY PARKER, *The Army of Flanders and the Spanish Road, 1567–1659: The Logistics of Spanish Victory and Defeat in the Low Countries' Wars* (1972), examines human and tactical aspects of the war of the Netherlands, and his *The Military Revolution: Military Innovation and the Rise of the West, 1500–1800* (1988), uses current scholarship in an analysis of the impact of gunpowder on warfare and politics on a global scale.

CHRISTOPHER DUFFY, *Fire and Stone: The Science of Fortress Warfare, 1660–1860* (1975), and *Siege Warfare,* vol. 1: *The Fortress in the Early Modern World, 1494–1660,* and vol. 2: *The Fortress in the Age of Vauban and Frederick the Great, 1680–1789* (1979–85), treat the development of methods of siegecraft and fortification and survey positional warfare. SÉBASTIEN LE PRESTRE DE VAUBAN, *A Manual of Siegecraft and Fortification,* trans. and ed. by GEORGE A. ROTHROCK (1968; originally published in French, 1740), is the famous text with informative notes, appendixes, and introduction; and SIMON PEPPER and NICHOLAS ADAMS, *Firearms & Fortification: Military Architecture and Siege Warfare in Sixteenth-Century Siena* (1986), studies the technology and tactics of early modern positional warfare in detail. PETER YOUNG and WILFRID EMBERTON, *Sieges of the Great Civil War, 1642–1646* (1978), offers siege narratives of the English Civil War; and HORST DE LA CROIX, *Military Considerations in City Planning: Fortifications* (1972), surveys urban fortifications from the earliest times to the mid-19th century.

For the transition to modern warfare, see WILLIAM H. MCNEILL, *The Pursuit of Power: Technology, Armed Force, and Society Since A.D.1000* (1982), a broad overview of the political and economic effect of developments in military technology; HEW STRACHAN, *European Armies and the Conduct of War* (1983), which summarizes developments from the age of Frederick the Great onward; and MARTIN VAN CREVELD, *Technology and War: From 2000 B.C. to the Present* (1989). (J.F.G.)

Modern weapons and weapon systems: On small arms, basic reference works offering encyclopedic treatment include CLAUDE BLAIR (ed.), *Pollard's History of Firearms* (1983); JAROSLAV LUGS, *Firearms Past and Present: A Complete Review of Firearms Systems and Their Histories,* 2 vol., trans. from Czech (1973); and HAROLD L. PETERSON (ed.), *Encyclopedia of Firearms* (1964). Histories of small-arms development in various countries include HOWARD L. BLACKMORE, *British Military Firearms, 1650–1850* (1961); M.L. BROWN, *Firearms in Colonial America: The Impact on History and Technology, 1492–1792* (1980); JOHN WALTER, *The German Rifle: A Comprehensive Illustrated History of the Standard Bolt-Action Designs, 1871–1945* (1979); and EDWARD C. EZELL, *The AK-47 Story: Evolution of the Kalashnikov Weapons* (1986), including a history of shoulder-weapons development in Russia and the Soviet Union from 1800 to the 1980s, and *The Great Rifle Controversy: Search for the Ultimate Infantry Weapon from World War II Through Vietnam and Beyond* (1984), on U.S. and NATO small-arms development. Other works by Ezell are *Handguns of the World: Military Revolvers and Self-Loaders from 1870 to 1945* (1979); and *Small Arms of the World: A Basic Manual of Small Arms,* 12th rev. ed. (1983), with chapters that summarize the development of rifles, handguns, submachine guns, machine guns, and special-purpose weapons from 1945 to the 1980s. MERRITT ROE SMITH, *Harpers Ferry Armory and the New Technology: The Challenge of Change* (1977), discusses the American System of manufacture in its human and technological contexts. JOHN ELLIS, *The Social History of the Machine Gun* (1975, reprinted 1986), is an essay on the social impact of the machine gun. (E.C.E.)

On artillery, a full history of gun design and development in the second half of the 19th century is presented in the official British handbook *Treatise on Service Ordnance,* 7th ed. (1908); an overview of artillery knowledge of that period, with useful historical background, is offered in E.W. LLOYD and A.G. HADCOCK, *Artillery: Its Progress and Present Position* (1893). The design and development of the guns, carriages, ammunition, and tactics, with descriptions of current weapons, are covered in H.A. BETHELL, *Modern Guns and Gunnery, 1910: A Practical Manual for Officers of the Horse, Field, and Mountain Artillery,* new ed. (1910), and *Modern Artillery in the Field: A Description of the Artillery of the Field Army, and the Principles and Methods of Its Employment* (1911). THOMAS J. HAYES, *Elements of Ordnance: A Textbook for Use of Cadets of the United States Military Academy* (1938), is a later textbook covering the design and construction of guns, mountings, ammunition, and fire-control equipment in ample detail. IAN V. HOGG, *A History of Artillery* (1974), is a profusely illustrated history providing considerable technical detail on the development of arms from the 12th century to the present day. *Jane's Armour and Artillery* (annual) lists all current tanks and artillery weapons, with descriptions, specifications, and illustrations. (I.V.H.)

On missile systems, see KENNETH W. GATLAND, *Development of the Guided Missile,* 2nd ed. (1954), a somewhat technical examination of early efforts. Other historical works include HOMER E. NEWELL, *Guide to Rockets, Missiles, and Satellites,* 2nd rev. ed. (1961); ERIC BURGESS, *Long-Range Ballistic Missiles* (1961); E. MICHAEL DEL PAPA, *From SNARK to SRAM: A Pictorial History of Strategic Air Command Missiles* (1976); and NORMAN POLMAR, *Strategic Weapons: An Introduction,* rev. ed. (1982); and KENNETH P. WERRELL, *The Evolution of the Cruise Missile* (1985). Characteristics of strategic weapons are covered in MICHAEL J.H. TAYLOR, *Missiles of the World,* 3rd ed. (1980); and PHILIP BIRTLES and PAUL BEAVER, *Missile Systems* (1985). Some insight into future developments is suggested in MAX WALMER, *An Illustrated Guide to Strategic Weapons* (1988); and UNITED STATES. DEPARTMENT OF DEFENSE, *Soviet Military Power: An Assessment of the Threat, 1988* (1988). The latest developments in the field are covered in *Jane's Weapon Systems* (annual) and *The International Countermeasures Handbook* (annual). (S.O.F./J.F.G.)

The history of nuclear weapons is the subject of voluminous literature. RICHARD RHODES, *The Making of the Atomic Bomb* (1986), is an excellent work on the U.S. effort to develop nuclear weapons. It can be supplemented by the official histories: VINCENT C. JONES, *Manhattan, the Army and the Atomic Bomb* (1985); DAVID HAWKINS, EDITH C. TRUSLOW, and RALPH CARLISLE SMITH, *Manhattan District History-Project Y, the Los Alamos Project,* 2 vol. (1961, reprinted as *Project Y, the Los Alamos Story,* in 1 vol. with a new introduction, 1983); and RICHARD G. HEWLETT, OSCAR E. ANDERSON, JR., and FRANCIS DUNCAN, *A History of the United States Atomic Energy Commission,* 2 vol. (1962–69); continued by RICHARD G. HEWLETT and JACK M. HALL, *Atoms for Peace and War, 1953–1961* (1989). For the development of thermonuclear weapons, see HERBERT F. YORK, *The Advisors: Oppenheimer, Teller, and the Superbomb* (1976); and HANS A. BETHE, "Comments on the History of the H-Bomb," *Los Alamos Science,* 3(3):43–53 (Fall 1982). Technical data are compiled in THOMAS B. COCHRAN, WILLIAM M. ARKIN, and MILTON M. HOENIG, *U.S. Nuclear Forces and Capabilities* (1984); and THOMAS B. COCHRAN *et al., U.S. Nuclear Warhead Production* (1987), and *U.S. Nuclear Warhead Facility Profiles* (1987).

The British project is discussed in the official histories of the U.K. Atomic Energy Authority: MARGARET GOWING, *Britain and Atomic Energy, 1939–1945* (1964), and *Independence and Deterrence: Britain and Atomic Energy, 1945–1952,* 2 vol. (1974). There is not much published about the Soviet program, but see DAVID HOLLOWAY, *The Soviet Union and the Arms Race,* 2nd ed. (1984); and THOMAS B. COCHRAN, *Soviet Nuclear Weapons* (1989). No official history is available for the French project. BERTRAND GOLDSCHMIDT, *Les Rivalités atomiques, 1939–1966* (1967), is a semiofficial account by a participant. The Chinese project is covered in JOHN WILSON LEWIS and XUE LITAI, *China Builds the Bomb* (1988). DAVID IRVING, *The German Atomic Bomb* (1968, reprinted 1983), covers the German program; and ROBERT K. WILCOX, *Japan's Secret War* (1985),

examines Japanese work on the atomic bomb. Proliferation developments are followed in LEONARD S. SPECTOR, *Nuclear Proliferation Today* (1984), *The New Nuclear Nations* (1985), *Going Nuclear* (1987), and *The Undeclared Bomb* (1988).

(T.B.C./R.S.N.)

Chemical and biological weapons are the subject of FREDERICK J. BROWN, *Chemical Warfare: A Study in Restraints* (1968, reprinted 1981); L.F. HABER, *The Poisonous Cloud: Chemical Warfare in the First World War* (1986); UNITED NATIONS, *Chemical and Bacteriological (Biological) Weapons and the Effects of Their Possible Use: Report of the Secretary General* (1969); WORLD HEALTH ORGANIZATION, *Health Aspects of Chemical and Biological Weapons* (1970); STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE, *The Problem of Chemical and Biological Warfare: A Study of the Historical, Technical, Military, Legal, and Political Aspects of CBW, and Possible Disarmament Measures,* 6 vol. (1971–75); and MATTHEW MESELSON and JULIAN PERRY ROBINSON, "Chemical Warfare and Chemical Disarmament," *Scientific American,* 242(4):38–47 (April 1980).

(Ma.Me./J.P.P.R.)

Tanks and armoured vehicles are discussed in RICHARD M. OGORKIEWICZ, *Armour: A History of Mechanized Forces* (1960, reissued as *Armoured Forces: A History of Armoured Forces and Their Vehicles* 1970), and *Design and Development of Fighting Vehicles* (1968), on the technical aspects. The history of various types of vehicles from World War I to the end of World War II is presented in DUNCAN CROW (ed.), *AFVs of World War One* (1970), *British AFVs, 1919–40* (1970), *British and Commonwealth AFVs, 1940–46* (1971), *American AFVs of World War II* (1972), and *Armored Fighting Vehicles of Germany: World War II* (1973). PETER CHAMBERLAIN and HILARY L. DOYLE, *Encyclopedia of German Tanks of World War Two: A Complete Illustrated Directory of German Battle Tanks, Armoured Cars, Self-Propelled Guns, and Semi-tracked Vehicles, 1933–1945* (1978), provides a comprehensive pictorial record. DUNCAN CROW (ed.), *Modern Battle Tanks* (1978), focuses on the period between 1950 and 1970. For particular types of tanks, see WALTER J. SPIELBERGER, *From Half-Track to Leopard 2: The Complete Illustrated History of the Krauss-Maffei Ordnance Department* (1979; originally published in German, 1979); JOHN MILSOM, *Russian Tanks, 1900–1970: The Complete Illustrated History of Soviet Armoured Theory and Design* (1970); and R.P. HUNNICUT, *Sherman: A History of the American Medium Tank* (1978), *Patton: A History of the American Main Battle Tank* (1984), and *Firepower: A History of the American Heavy Tank* (1988). For current information on developments in technology, see *Jane's Armour and Artillery* (annual) and *Jane's Armoured Fighting Vehicle Systems* (annual). (R.M.O.)

Naval ships and craft are the subject of voluminous literature. For early history, WILLIAM LEDYARD RODGERS, *Naval Warfare Under Oars, 4th to 16th Centuries* (1939, reprinted 1967), and *Greek and Roman Naval Warfare* (1937, reprinted 1973), are older but still the most comprehensive sources on the classical era. J.S. MORRISON and J.F. COATES, *The Athenian Trireme: The History and Reconstruction of an Ancient Greek Warship* (1986), treats in detail design evolution and technological developments. JOHN H. PRYOR, *Geography, Technology, and War: Studies in the Maritime History of the Mediterranean, 649–1571* (1988), explores both military and commercial navigation in the region. JOHN FRANCIS GUILMARTIN, JR., *Gunpowder and Galleys: Changing Technology and Mediterranean Warfare at Sea in the Sixteenth Century* (1974), examines the way in which gunpowder changed the nature of warfare on land as well as at sea. PHILIP COWBURN, *The Warship in History* (1965); RICHARD HOUGH, *Fighting Ships* (1969); and DONALD MACINTYRE and BASIL W. BATHE, *The Man-of-War: A History of the Combat Vessel,* rev. ed. (1974), discuss and illustrate the general evolution of the warship from early to modern times. (J.F.G.)

The modern era is covered in PETER HODGES, *The Big Gun: Battleship Main Armament, 1860–1945* (1981); NORMAN FRIEDMAN, *Battleship Design and Development, 1905–1945* (1978), and *Modern Warship: Design and Development* (1979); and DAVID R. FRIEDEN (ed.), *Principles of Naval Weapons Systems* (1985). Historical characteristics of the major navies of the world are found in ROGER CHESNEAU and EUGENE M. KOLESNIK (eds.), *Conway's All the World's Fighting Ships, 1860–1905* (1979); ROBERT GARDINER and RANDAL GRAY (eds.), *Conway's All the World's Fighting Ships, 1906–1921* (1985); ROGER CHESNEAU (ed.), *Conway's All the World's Fighting Ships, 1922–1946* (1980); and ROBERT GARDINER (ed.), *Conway's All the World's Fighting Ships, 1947–1982,* 2 vol. (1983). For characteristics by country, see: (*United States*): NORMAN POLMAR, *Ships and Aircraft of the U.S. Fleet,* 14th ed. (1987); NORMAN FRIEDMAN, *U.S. Battleships* (1985), and *U.S. Cruisers* (1984); and PAUL H. SILVERSTONE, *U.S. Warships Since 1945* (1987); (*Great Britain*): D.K. BROWN, *A Century of Naval Construction: The History of the Royal Corps of Naval Constructors, 1883–1983* (1983); and ALAN RAVEN and JOHN ROBERTS, *British Cruisers of World War Two* (1980), and *British Battleships of World War Two* (1976). (*Other countries*): JOHN JORDAN, *Soviet Warships: The Soviet Surface Fleet, 1960 to the Present* (1983); NORMAN POLMAR, *Guide to the Soviet Navy,* 4th ed. (1986); HENRI LE MASSON, *The French Navy,* 2 vol. (1969); HANSGEORG JENTSCHURA, DIETER YUNG, and PETER MICKEL, *Warships of the Imperial Japanese Navy, 1869–1945* (1977; originally published in German, 1970); and ERICH GRÖNER (ed.), *Die deutschen Kriegsschiffe, 1815–1945* (1982–), with six out of projected seven volumes published by 1988.

The engineering characteristics of submarines are explored in ULRICH GABLER, *Submarine Design* (1986; originally published in German, 1964); NORMAN FRIEDMAN, *Submarine Design and Development* (1984); EBERHARD RÖSSLER, *The U-Boat: The Evolution and Technical History of German Submarines* (1981; originally published in German, 1975); and JOHN D. ALDEN, *The Fleet Submarine in the U.S. Navy: A Design and Construction History* (1979). ERMINIO BAGNASCO, *Submarines of World War Two* (1977; originally published in Italian, 1973); DORR CARPENTER and NORMAN POLMAR, *Submarines of the Imperial Japanese Navy* (1986); and RICHARD COMPTON-HALL, *The Underwater War, 1939–1945* (1982), discuss underwater naval operations during this specific period. Antisubmarine warfare is the subject of RICHARD COMPTON-HALL, *Submarine Versus Submarine: The Tactics and Technology of Underwater Warfare* (1988).

On aircraft carriers, see DAVID BROWN, *Aircraft Carriers* (1977); ROGER CHESNEAU, *Aircraft Carriers of the World, 1914 to the Present: An Illustrated Encyclopedia* (1984); NORMAN FRIEDMAN, *Carrier Air Power* (1981), *U.S. Aircraft Carriers: An Illustrated Design History* (1983), and *British Carrier Aviation: The Evolution of the Ships and Their Aircraft* (1988). (N.F.)

For military aircraft, historical overviews are offered by JOHN W.R. TAYLOR (ed.), *The Lore of Flight* (1970, reissued 1986), a technically informed and well-illustrated work with separate sections on history, structures, engines, equipment, and flying and containing an encyclopedic index; and DAVID BROWN, CHRISTOPHER SHORES, and KENNETH MACKSEY, *The Guinness History of Air Warfare* (1976), a concise compilation of salient events and developments. Aircraft design is covered in JOHN D. ANDERSON, JR., *Introduction to Flight,* 3rd ed. (1989), an engineering overview of aircraft design and performance at a basic level, incorporating a technically accurate history of manned flight; and EDWARD H. HEINEMANN, ROSARIO RAUSA, and K.E. VAN EVERY, *Aircraft Design* (1985), a study of contemporary theory and practice. Aircraft propulsion is covered in HERSCHEL SMITH, *Aircraft Piston Engines* (1981), a technically informed history of power plants from the Wright era to the years immediately after World War II; and EDWARD W. CONSTANT II, *The Origins of the Turbojet Revolution* (1980), a scholarly history of the development of turbojet engines. For early military aircraft, see RICHARD P. HALLION, *Rise of the Fighter Aircraft, 1914–18* (1984), a technologically and tactically informed account; and JOHN H. MORROW, JR., *German Air Power in World War I* (1982), an account stressing interactions between strategy and tactics on the one hand and social and economic factors on the other. ERIC M. BROWN, *Duels in the Sky: World War II Naval Aircraft in Combat* (1988), is an analysis of combat capabilities, based on personal experience, by a military test pilot. ROBERT L. SHAW, *Fighter Combat,* 2nd ed. (1988), focuses on modern jet aircraft but incorporates a solid historical base. R.A. MASON and JOHN W.R. TAYLOR, *Aircraft, Strategy, and Operations of the Soviet Air Force* (1986), covers the development of military aeronautics in the Soviet Union. For treatments of radar-evading aircraft, see DOUG RICHARDSON, *Stealth Warplanes* (1989; also published as *Stealth*), a review of the origins of stealth technology with informed speculation on current and future developments. (J.F.G.)

Airborne warning and control systems are discussed in ARNOLD LEE TESSMER, *Politics of Compromise: NATO and AWACS* (1988); and MIKE HIRST, *Airborne Early Warning: Design, Development, and Operations* (1983). DANIEL FORD, *The Button: The Pentagon's Strategic Command and Control System* (1985); UNITED STATES. CONGRESS. COMMITTEE ON GOVERNMENT OPERATIONS, *Our Nation's Nuclear Warning System: Will It Work If We Need It?* (1986); and STEPHEN KIRBY and GORDON ROBSON (eds.), *The Militarisation of Space* (1987), study missile-attack warning systems. Use of electronics in military engineering is the subject of CHARLES M. DAVIS *et al., Fiberoptic Sensor Technology Handbook* (1982); and RICHARD G. WILEY, *Electronic Intelligence, the Interception of Radar Signals* (1985).

For information on military communications, see R.J. HOLBECHE (ed.), *Land Mobile Radio Systems* (1985); WILLIAM C.Y. LEE, *Mobile Communications Design Fundamentals* (1986); H. HODARA (ed.), *Optical Fibres, Integrated Optics, and Their Military Applications* (1977); RAY BONDS (ed.), *The Modern U.S. War Machine: An Encyclopedia of American Military Equipment and Strategy,* rev. ed. (1987); and, for a more general view, L.H. GANN, *The Defense of Western Europe* (1987).

(Ed.)

The Theory and Conduct of War

War in the popular sense is a conflict among political groups involving hostilities of considerable duration and magnitude. In the usage of social science certain qualifications are added. Sociologists usually apply the term to such conflicts only if they are initiated and conducted in accordance with socially recognized forms. They treat war as an institution recognized in custom or in law. Military writers usually confine the term to hostilities in which the contending groups are sufficiently equal in power to render the outcome uncertain for a time. Armed conflicts of powerful states with primitive peoples are usually called pacifications, military expeditions, or explorations; with small states, they are called interventions or reprisals; and with internal groups, rebellions or insurrections. Such incidents, if the resistance is sufficiently strong or protracted, may achieve a magnitude that entitles them to the name "war."

This article is in two major sections. The first treats the social sciences of war, such as law, economics, and theories of the causes and prevention of war. The second section presents the military arts and sciences, such as strategy, tactics, logistics, and intelligence.

War is treated in several other *Macropædia* articles. For the relation between diplomacy and war, see INTERNATIONAL RELATIONS, 20TH-CENTURY. For the weapons of war, see WAR, THE TECHNOLOGY OF. For a history of World Wars I and II, see WORLD WARS, THE. Other wars are covered in the history sections of articles on the relevant countries or regions—*e.g.*, for the American Civil War, see UNITED STATES OF AMERICA; for the Punic Wars, see GREEK AND ROMAN CIVILIZATIONS, ANCIENT.

For coverage of related topics in the *Macropædia* and *Micropædia*, see the *Propædia*, section 544, and the *Index*.

The article is divided into the following sections:

The institution of war 628
- Theory of war 628
  - Evolution of theories of war
  - The causes of war
  - The control of war
- Law of war 633
  - Roots of the international law of war
  - Commencing hostilities
  - Conducting hostilities
  - Cessation of hostilities
  - War crimes
- Military law 639
  - Historical development
  - Jurisdiction
  - Procedure
  - Legal trends
- Defense economics 642
  - Opportunities foregone: the cost of war
  - Defense expenditure: the cost of deterrence
  - Defense management: budgeting deterrence
  - War finance: when deterrence fails

The conduct of war 647
- Strategy 647
  - Fundamentals
  - Historical development through World War II
  - Strategy in the nuclear age
- Tactics of land warfare 659
  - Fundamentals
  - Historical development
- Naval tactics 670
  - Fundamentals
  - Historical development
- Air tactics 675
  - Through World War I
  - Through World War II
  - The jet age
- Logistics 678
  - Fundamentals
  - Historical development
- Intelligence 686
  - Levels of intelligence
  - Sources of intelligence
  - Types of intelligence
- Guerrilla warfare 689
  - History
  - Principles
  - Strategy and tactics
  - Counterguerrilla warfare

Bibliography 693

THE INSTITUTION OF WAR

Theory of war

In all ages war has been an important topic of analysis. In the latter part of the 20th century, in the aftermath of two world wars and in the shadow of nuclear, biological, and chemical holocaust, more has been written on the subject than ever before. Endeavours to understand the nature of war, to formulate some theory of its causes, conduct, and prevention, are of great importance, for theory shapes human expectations and determines human behaviour. The various schools of theorists are generally aware of the profound influence they can exercise upon life, and their writings usually include a strong normative element, for, when accepted by politicians, their ideas can assume the characteristics of self-fulfilling prophecies.

The analysis of war may be divided into several categories. Philosophical, political, economic, technological, legal, sociological, and psychological approaches are frequently distinguished. These distinctions indicate the varying focuses of interest and the different analytical categories employed by the theoretician, but most of the actual theories are mixed because war is an extremely complex social phenomenon that cannot be explained by any single factor or through any single approach.

EVOLUTION OF THEORIES OF WAR

Reflecting changes in the international system, theories of war have passed through several phases in the course of the past three centuries. After the ending of the wars of religion, about the middle of the 17th century, wars were fought for the interests of individual sovereigns and were limited both in their objectives and in their scope. The art of maneuver became decisive, and analysis of war was couched accordingly in terms of strategies. The situation changed fundamentally with the outbreak of the French Revolution, which increased the size of forces from small professional to large conscript armies and broadened the objectives of war to the ideals of the revolution, ideals that appealed to the masses who were subject to conscription.

War as a rational instrument

In the relative order of post-Napoleonic Europe the mainstream of theory returned to the idea of war as a rational, limited instrument of national policy. This approach was best articulated by the Prussian military theorist Carl von Clausewitz in his famous classic *On War* (1832–37).

World War I, which was "total" in character because it resulted in the mobilization of entire populations and economies for a prolonged period of time, did not fit into the Clausewitzian pattern of limited conflict, and it led to a renewal of other theories. These no longer regarded war

as a rational instrument of state policy. The theorists held that war, in its modern, total form, if still conceived as a national state instrument, should be undertaken only if the most vital interests of the state, touching upon its very survival, are concerned. Otherwise, warfare serves broad ideologies and not the more narrowly defined interests of a sovereign or a nation. Like the religious wars of the 17th century, war becomes part of "grand designs," such as the rising of the proletariat in communist eschatology or the Nazi doctrine of a master race.

Some theoreticians have gone even further, denying war any rational character whatsoever. To them war is a calamity and a social disaster, whether it is afflicted by one nation upon another or conceived of as afflicting humanity as a whole. The idea is not new—in the aftermath of the Napoleonic Wars it was articulated, for example, by Tolstoy in the concluding chapter of *War and Peace* (1865–69). In the second half of the 20th century it gained new currency in peace research, a contemporary form of theorizing that combines analysis of the origins of warfare with a strong normative element aiming at its prevention. Peace research concentrates on two areas: the analysis of the international system and the empirical study of the phenomenon of war.

World War II and the subsequent evolution of weapons of mass destruction made the task of understanding the nature of war even more urgent. On the one hand, war has become an intractable social phenomenon, the elimination of which seems to be an essential precondition for the survival of mankind. On the other hand, the use of war as an instrument of policy is calculated in an unprecedented manner by the nuclear superpowers, the United States and the Soviet Union. War also remains a stark but rational instrumentality in certain more limited conflicts, such as those between Israel and the Arab nations. Thinking about war is, consequently, becoming increasingly more differentiated because it has to answer questions related to very different types of conflict.

Definitions of war

Clausewitz cogently defines war as a rational instrument of foreign policy: "an act of violence intended to compel our opponent to fulfill our will." Modern definitions of war, such as "armed conflict between political units," generally disregard the narrow, legalistic definitions characteristic of the 19th century, which limited the concept to formally declared war between states. Such a definition includes civil wars but at the same time excludes such phenomena as insurrections, banditry, or piracy. Finally, war is generally understood to embrace only armed conflicts on a fairly large scale, usually excluding conflicts in which fewer than 50,000 combatants are involved.

THE CAUSES OF WAR

Contemporary theories of the causes of war divide roughly into two major schools. One attributes war to certain innate biological and psychological factors or drives, the other attributes it to certain social relations and institutions. Both schools include optimists and pessimists concerning the preventability of war.

Biological theories. Theories centring upon man's innate drives are developed by ethologists, who draw analogies from animal behaviour, and also by psychologists and psychoanalysts.

Ethology. Ethologists start with the persuasive argument that study of animal warfare may contribute toward an understanding of war as employed by man. The behaviour of monkeys and apes in captivity and of young children, for example, shows basic similarities. In both cases it is possible to observe that aggressive behaviour usually arises from several drives: rivalry for possession, the intrusion of a stranger, or frustration of an activity. The major conflict situations leading to aggression among animals, especially those concerning access of males to females and control of a territory for feeding and breeding, are usually associated with patterns of dominance.

Aggressive behavior in animals

The analogies of animal to human behaviour drawn by many ethologists, however, are severely questioned by their more restrained colleagues as well as by many social scientists. The term aggression, for example, is imprecisely and inconsistently used, often referring merely to the largely symbolic behaviour of animals involving such signals as grimaces.

Observed animal behaviour can be regarded as a possible important source of inspiration for hypotheses, but these must then be checked through the study of actual human behaviour. As this has not yet been adequately done, the hypotheses advanced have little foundation and are merely interesting ideas to be investigated. Further, human behaviour is not fixed to the extent that animal behaviour is, partly because man rapidly evolves different patterns of behaviour in response to environmental factors, such as geography, climate, and contact with other social groups. The variety of these behaviour patterns is such that they can be used on both sides of an argument concerning, for example, whether or not men have an innate tendency to be aggressive.

Two particularly interesting subjects studied by ethologists are the effects of overcrowding on animals and animal behaviour regarding territory. The study of overcrowding is incomplete, and the findings that normal behaviour patterns tend to break down in such conditions and that aggressive behaviour often becomes prominent are subject to the qualification that animal and human reactions to overcrowding may be different. Ethologists have also advanced plausible hypotheses concerning biological means of population control through reduced fertility that occurs when animal populations increase beyond the capacity of their environment. Whether such biological control mechanisms operate in human society, however, requires further investigation.

Findings concerning the "territorial imperative" in animals—that is, the demarcation and defense against intrusion of a fixed area for feeding and breeding—are even more subject to qualification when an analogy is drawn from them to human behaviour. The analogy between an animal territory and a territorial state is obviously extremely tenuous. In nature the territories of members of a species differ in extent but usually seem to be provided with adequate resources, and use of force in their defense is rarely necessary, as the customary menacing signals generally lead to the withdrawal of potential rivals. This scarcely compares with the sometimes catastrophic defense of the territory of a national state.

Psychology. One school of theorists has postulated that the major causes of war can be found in man's psychological nature. Such psychological approaches range from very general, often merely intuitive assertions regarding human nature, to complex analyses utilizing the concepts and techniques of modern psychology. The former category includes a wide range of ethical and philosophical teaching and insights, including the works of such figures as St. Augustine and the 17th-century Dutch philosopher Spinoza.

Psychological maladjustments

Modern writers utilizing psychological approaches emphasize the significance of psychological maladjustments or complexes and of false, stereotyped images held by decision makers of other countries and their leaders. Some psychologists posit an innate aggressiveness in man. Others concentrate upon public opinion and its influence, particularly in times of tension; others stress the importance of decision makers and the need for their careful selection and training. Most believe that an improved social adjustment of individuals would decrease frustration, insecurity, and fear and would reduce the likelihood of war. All of them believe in the importance of research and education. Still, the limitations of such approaches derive from their very generality. Also, whether the psychological premises are optimistic or pessimistic about the nature of man, one cannot ignore the impact upon human behaviour of social and political institutions that give man the opportunities to exercise his good or evil propensities and to impose restraints upon him.

Social theories. Whereas psychological explanations of war contain much that seems to be valid, they are insufficient because man behaves differently in different social contexts. Hence many thinkers have sought their explanations in these contexts, focusing either on the internal organization of states or on the international system within which these operate.

The most voluminous and influential theories attributing war to the nature of the state fall into two broad streams, which can be loosely called liberal and socialist.

Liberal analyses. The early or classical liberals of the 18th and 19th centuries distinguished three basic elements in their analysis—individuals, society, and the state—and regarded the state as the outcome of the interaction of the former two. They assumed that society is self-regulating and that the socioeconomic system is able to run smoothly with little interference from the government. Economy, decentralization, and freedom from governmental control were the classical liberal's main concerns, as shown particularly clearly in the writings of John Stuart Mill. They accepted the necessity of maintaining defense but postulated the existence of a basic harmony of interests among states, which would minimize the incidence of wars. Economic cooperation based upon an international division of labour and upon free trade would be in the interests of everybody—commerce would be the great panacea, the rational substitute for war.

In explanation of wars that did occur, however, liberals emphasized a variety of factors. First, they focused on autocratic governments, which were presumed to wage war against the wishes of peacefully inclined people. It thus became a major tenet of liberal political philosophy that war could be eliminated by introducing universal suffrage because the people would surely vote out of office any belligerently inclined government. From the early American pamphleteer Thomas Paine onward, a major school of liberals supported republicanism and stressed the peaceful impact of public opinion. Although they could not agree about actual policies, they stressed certain general ideas concerning relations between states, paralleling their laissez-faire ideas of the internal organization of the state with ideas of a minimum amount of international organization, use of force strictly limited to repelling aggression, the importance of public opinion and of democratically elected governments, and rational resolution of conflicts and disputes. Later in the course of the 19th century, however, and especially after World War I, liberals began to accept the conclusion that an unregulated international society did not automatically tend toward peace and advocated international organization as a corrective.

Socialist analyses. Whereas liberals concentrated on political structures, regarding them as of primary importance in determining the propensity of states to engage in war, socialists turned to the socioeconomic system of states as the primary factor. Early in the 20th century the two streams did, to some extent, converge, as evidenced by the fact that the English radical liberal John Hobson explained wars in terms later adopted by Lenin.

Marxian class conflict

Karl Marx attributed war not to the behaviour of states but to the class structure of society. To him wars occurred not as an often voluntary instrument of state policy but as the result of a clash of social forces. To Marx the state was merely a political superstructure; the primary, determining factor lies in the capitalist mode of production, which leads to the development of two antagonistic classes: the bourgeoisie and the proletariat. The bourgeoisie controls governmental machinery in its own interests. In its international relations, the capitalist state engages in wars because it is driven by the dynamism of its system—the constantly growing need for raw materials, markets, and supplies of cheap labour. The only way to avoid war is to remove its basic cause, by replacing capitalism with socialism, thus abolishing both class struggle and states. The Marxist doctrine, however, gave no clear guidance about the interim period before the millennium is reached; and the international solidarity of the proletariat proved a myth when war broke out in 1914, facing the European Social Democratic parties with the problem of adopting an attitude to the outbreak of the war. The Second International of working-class parties had repeatedly passed resolutions urging the working classes to bring pressure upon their respective governments to prevent war, but, once war had broken out, each individual party chose to regard it as defensive for its own state and to participate in the war effort. This was explained by Lenin as being due to a split in the organization of the proletariat that could be overcome only through the activity of a rigidly organized revolutionary vanguard.

Socialists in the West turned increasingly, although in varying degrees, to revisionist interpretations of Marxism and returned to their attempts to revise socioeconomic structures through evolutionary constitutional processes, seeing this as the only possible means of preventing wars. In the Soviet Union the socialist theory of war changed as the new communist regime responded to changes in circumstances. Soviet theoreticians have distinguished three major types of war: between capitalist states, between capitalist and socialist states, and colonial wars of liberation. The internecine wars among capitalist states are supposed to arise from capitalist competition and imperialist rivalries, such as those that led to the two world wars. They are desirable, for they weaken the capitalist camp. A war between capitalist and socialist states is one that clearly expresses the basic principle of class struggle and is, therefore, one for which the socialist states must prepare. Finally, wars of colonial liberation can be expected between subjugated people and their colonial masters.

The weakness of the theory is that the two major expected types of war, the intracapitalist and the capitalist–socialist, have not materialized as frequently as Soviet theoreticians predicted. Further, the theory failed to analyze adequately the situation in the Soviet Union and in the socialist camp. Even in communist countries, nationalism seems to have proved more powerful than socialism: "national liberation" movements have appeared and have had to be forcibly subdued in the Soviet Union, despite its communist regime. Also, war between socialist states is not unthinkable, as the doctrine indicates: only the colossal preponderance of Soviet forces prevented a full-scale war in 1956 against Hungary and in 1968 against Czechoslovakia; war between the Soviet Union and the People's Republic of China was a serious possibility for two decades after the Sino-Soviet split in 1962; and armed conflict erupted between China and Vietnam after the latter country became the most powerful in Southeast Asia. Finally, the theory does not provide for wars of liberation against socialist states, such as that conducted by the Afghan mujahideen against the Soviet Union from 1979 to 1989.

Nationalism. Many theories claim or imply that wars result ultimately from the allegiance of men to nations and from the intimate connection between the nation and a state. This link between the nation and the state is firmly established by the doctrine of national self-determination, which has become in the eyes of many the major basis of the legitimacy of states and the major factor in their establishment and breakup. It was the principle on which the political boundaries of eastern Europe and the Balkans were arranged after World War I and has been the principal slogan of the anticolonial movement of the 20th century, finding expression in Chapter I, article 1, of the Charter of the United Nations in the objective of "self-determination of peoples," as well as in the more specific provisions of Chapters XI and XII. It is this intimate link between nationalism and statehood that renders them both so dangerous. The rulers of a state are ultimately governed in their behaviour by what is loosely summed up as the "national interest," which occasionally clashes directly with the national interests of other states.

National self-determination

The ideal of the nation-state is never fully achieved. In no historical case does one find all members of a particular nation gathered within one state's boundaries. Conversely, many states contain sizable national minorities. This lack of full correlation has frequently given rise to dangerous tensions that can ultimately lead to war. A government inspired by nationalism may conduct a policy aiming at the assimilation of national minorities, as was the general tendency of central and eastern European governments in the interwar period; it may also attempt to reunite the members of the nation living outside its boundaries, as Adolf Hitler did. National groups that are not in control of a state may feel dissatisfied with its regime and claim self-determination in a separate state, as demonstrated in the attempt to carve Biafra out of Nigeria and the separation of Bangladesh from Pakistan.

There is no rational basis for deciding on the extent to which the self-determination principle should be applied in allowing national minorities to break away. As a rule, the majority group violently opposes the breakaway movement. Violent conflicts can ensue and, through foreign involvement, turn into international wars. No suitable method has been found for divorcing nationalism from the state and for meeting national demands through adequate social and cultural provisions within a larger unit. Such an attempt in the Austro-Hungarian Empire before its dissolution in World War I failed. Whether the Soviet Union will be permanently successful in containing its large proportion of national minorities remains to be seen.

Nationalism not only induces wars but, through the severity of its influence, makes compromise and acceptance of defeat more difficult. It thus tends to prolong the duration and to increase the severity of wars. Possibly, however, this is the characteristic only of new, immature nationalisms, for nationalism has ceased to be a major cause of conflict and war among the nations of western Europe.

Nationalism is but one form of ideology: in all ages people seem to develop beliefs and try to proselytize others. Even within particular ideological groups schisms result in conflicts as violent as those between totally opposed creeds, and heretics are often regarded as more dangerous and hostile than opponents. As long as individual states can identify themselves with explosive differences in beliefs, the probability of a war between states is increased, and its intensity is likely to be greater.

Special-interest groups. Whereas some theories of war regard the state as an undifferentiated whole and generalize about its behaviour, other theorists are more sociologically oriented and focus on the roles played within the state by various special-interest groups.

A distinction is made by these theorists between the great mass of people and those groupings directly involved or influential with government. The people, about whose attitudes adequate knowledge is lacking, are generally assumed to be taken up with their daily lives and to be in favour of peace. The influential groups, who are directly involved in external affairs and, hence, in wars, are the main subject of analysis. Warlike governments dragging peace-loving people into international conflict is a recurrent theme both of liberal and socialist analyses of war. Some writers have gone to the length of postulating a continuous conspiracy of the rulers against the ruled that can be traced to prehistoric times, when priests and warriors combined in the first state structures. Most writers, however, narrow the field and seek an answer to the question of why some governments are more prone to engage in war than others; and they generally find the answer in the influence of important interest groups that pursue particular and selfish ends.

The chief and most obvious of such groups is the military. Military prowess was a major qualification for political leadership in primitive societies; the search for military glory as well as for the spoils of victory seems to have been one of the major motivations for war. Once the military function became differentiated and separated from civilian ones, a tension between the two became one of the most important issues of politics. The plausible view has generally been held that the military strive for war, in which they attain greater resources and can satisfy their status seeking and sometimes, also, an aspiration for direct and full political power. In peacetime, the military are obviously less important, are denied resources, and are less likely to influence or attain political power directly. At the same time, a second, although usually subsidiary, consideration of the military as a causal agent in war holds that an officer corps is directly responsible for any fighting and is thus more aware of its potential dangers for its members and for the state as well. Although intent on keeping the state in a high state of preparedness, the military may be more cautious than civilians about engaging in war. It is often held, however, that increased military preparedness may result in increased tensions and thus indirectly lead to the outbreak of war.

Closely allied are theories about groups that profit from wars economically—capitalists and the financiers, especially those involved in industries catering to war. All of these play a central part as the villains of the piece in socialist and the liberal theories of war, and even those not subscribing to such theories do not deny the importance of military-industrial complexes in countries in which large sectors of the economy specialize in war supplies. But, although industrialists in all of the technologically advanced systems are undoubtedly influential in determining such factors as the level of armaments to be maintained, it is difficult to assume that their influence is or could be decisive when actual questions concerning war or peace are being decided by politicians.

Military-industrial complex

Finally, some scientists and technologists constitute a new, much smaller, but important group with special interests in war. To some extent one can generalize about them, although the group is heterogeneous, embracing as it does nuclear scientists, space researchers, biologists and geneticists, chemists, and engineers. If they are involved in defense work, they all share the interest of the military in securing more resources for their research: without their military applications, for example, neither nuclear nor space research would have gone ahead nearly as fast as it has. War, however, does not enhance the status and standing of scientists; on the contrary, they come under the close control of the military. They also usually have peaceful alternatives to military research, although these may not be very satisfactory or ample. Consequently, although modern war technology depends heavily upon scientists, and although many of them are employed by governments in work directly or indirectly concerned with this technology, scientists as a group are far from being wedded to war. On the contrary, many of them are deeply concerned with the mass destruction made possible by science and participate in international pacifist movements.

THE CONTROL OF WAR

The international environment within which states and the people within them operate is regarded by many theorists as the major factor determining the occurrence and the nature of wars. War remains possible as long as individual states seek to ensure self-preservation and to promote their individual interests and—in the absence of a reliable international agency to control the actions of other states—rely on their own efforts. It is no accident that reforms of the international system figure prominently in many prescriptions for the prevention of war. Whereas the reform of human propensities or of the state is bound to be a long, drawn-out affair, if it is at all possible, relatively straightforward partial reforms of the international system may produce significant restraints upon resorting to war, and a thorough reform could make war impossible.

Some theorists, being more optimistic about the nature of states, concentrate upon the removal of the fear and suspicion of other states, which is characteristic of the present as well as of all historical political systems; others, being less optimistic, think mainly of possible controls and restraints upon the behaviour of states. The underlying reasoning of both parties is generally similar. If individual states in competitive situations are governed by a short-term conception of their interests, acute conflicts between them will occur and will show a strong tendency to escalate. Thus, one state erects a tariff barrier to protect its industry against the competition of a trade partner, and the partner retaliates, the retaliatory interaction being repeated until the two countries find themselves in a trade war. Armaments races show a similar tendency to escalate, particularly so in an age of rapid technological change. The economic and the scientific efforts necessary to avoid falling behind rivals in the invention and development of rapidly improving weapons of mass destruction have already reached unprecedented heights.

Restraining national self-interest

And yet, neither trade wars nor arms races necessarily end in violent conflict; there seem to be operating some restraining and inhibiting factors that prevent an automatic escalation. Much of the theory of war concerns itself with the identification, improvement, and development of these restraining factors.

Diplomacy. The outcome of starkly competitive behaviour leading to wars is clearly against the interests of

states, and it is rational for them to seek more desirable outcomes. If competitive behaviour is dangerous, theorists seek for alternative methods of cooperative behaviour that would not jeopardize the interests of the state through exposing it to the possibly less cooperative behaviour of others. Some theorists concentrate upon improving the rationality of the decision-making of individual states through a better understanding of the international environment, through eliminating misperceptions and irrational fears, and through making clear the full possible costs of engaging in war and the full destructiveness of an all-out war, possible in our age.

Balance of power

The relative paucity of wars and their limited nature throughout the century following the Napoleonic Wars (1815–1914) have stirred great theoretical interest in the nature of the balance-of-power system of that period—that is, in the process by which the power of competing groups of states tended toward a condition of equilibrium. Contributing to the successful operation of the balance-of-power system of the 19th century were relatively slow technological change, great diversionary opportunities for industrial and colonial expansion, and the ideological and cultural homogeneity of Europe. Pursuit of a balance of power is a way of conducting foreign policy that is perhaps less prone to war than other types of policy because, instead of indiscriminately increasing their power, states increase it only moderately, so as not to provoke others; and instead of joining the strongest, they join the weaker side in order to ensure balance. States in a balance-of-power system must, however, be ready to abide by constraints upon their behaviour in order to ensure stability of the system.

The application to international relations of a branch of mathematics—game theory—that analyzes the strategy of conflict situations has provided a new tool of analysis. In state interaction, as in any game situation, one side's strategy generally depends upon that side's expectations of the other side's strategy. If all sides in a game are to maximize their chances of a satisfactory outcome, it is necessary that some rational rules of behaviour be conceptualized and agreed upon, and this idea of a set of rational rules can be applied to competing states in the international system. Game theorists distinguish antagonistic situations called zero-sum games in which one state's gain can be only at the expense of another state because the "payoff" is fixed. Even then a mutually acceptable distribution of gains can be rationally reached on the basis of the "minimax" principle—the party in a position of advantage satisfies itself with the minimum acceptable gain because it realizes that the other party, in a position of disadvantage, would yield on the basis of its possible minimum loss but would violently oppose a distribution even more to its detriment. In other situations called non-zero-sum games, the payoff is not constant but can be increased by a cooperative approach; the gain of one participant is not at the cost of another. The contestants, however, have to agree about the distribution of the gain, which is the product of their cooperation.

The theory of games is the foundation of theories of bargaining that analyze the behaviour of individual states in interaction. Diplomacy based upon such theories is less likely to lead to war. Policymakers pursuing such strategies will conduct conflicts of the zero-sum type so that war is avoided. More than that, with some skill, such situations can be transformed into the non-zero-sum type by introducing additional benefits accruing from cooperation in other interactions and also, more generally, by eliminating the likelihood of war and, consequently, by reducing the costs of preparing for one.

Regional integration. Because wars within states have been eliminated through the establishment of suitable political structures, such as central governments that hold a monopoly of coercive power, many theories concentrate upon the establishment of parallel structures within the international context. Regional integration (cooperation in economic, social, and political affairs, as, for example, within the European Economic Community) and the establishment of security communities, such as the North Atlantic Treaty Organization, have made much greater advances than attempts at the reform of the entire, global international system.

Because conflicts among neighbours tend to be frequent, regional integration is an important advance toward reducing the incidence of war. Even if it were to become generally successful, however, regional integration would simply shift the problem of war to a different level: there would be fewer possibilities of war because intraregional conflicts would be contained, but interregional conflicts could still give rise to wars of much greater scope and severity. The phenomenon of war must, therefore, be analyzed at the universal level.

International law. Some of the most influential thinking about war and the international system has come from specialists in international law. All of them postulate that there exists an international society of states that accepts the binding force of some norms of international behaviour. These norms are referred to as international law, although they differ fundamentally from municipal law because no sovereign exists who can enforce them. Most international lawyers realistically accept that international law is, consequently, among rather than above states. It is, according to legal doctrine, binding on states, but unenforceable.

The problem of enforcement

International law concerns itself largely with two aspects of war: its legality and its regulation. (For the regulation of warfare, see below *Law of war.*) As far as the legality of war is concerned, there has arisen in the 20th century a general consensus among states, expressed in several international treaties, including the Covenant of the League of Nations, the Kellogg–Briand Pact of 1928, and the Charter of the United Nations, that resort to armed force, except in certain circumstances, such as self-defense, is illegal. Such a legalistic approach to the prevention of war, however, remains futile in the absence of a means of enforcement. The enforcement provisions of the Covenant of the League of Nations and those of the United Nations Charter, which entail the application of military and economic sanctions, have never been applied successfully, owing to political disagreement among the major powers. This underlines the fact that legal norms, to be effective, must reflect an underlying political reality.

The United Nations. The United Nations is charged with the maintenance of international peace and security. The several approaches to peace outlined in its Charter and developed in its practice are based upon and clearly reflect the cumulative development of the relevant theories of war.

Drawing heavily upon the experience of the League of Nations, the Charter develops three interrelated approaches: first, pacific settlement of disputes, which would leave nations with nothing to fight about; second, collective security, which would confront aggressors with too much to fight against; and third, disarmament, which would deprive them of anything substantial with which to fight.

Peaceful settlement of disputes. Pacific settlement of disputes is based upon the assumption that war is primarily a technique for settling disputes, although it can, of course, also serve other purposes, such as allaying fears and seeking status. Further assumptions are that war frequently comes about because of the unawareness of decision makers of the possibility of settling disputes peacefully to the mutual advantage of both sides—an unawareness due to mere ignorance, pride, lack of imagination, or selfish and cynical leadership. It is thus possible that international organizations can contribute to the prevention of wars by devising and institutionalizing alternative, peaceful techniques for the settlement of disputes and by persuading the states to use them.

The scope of this approach is limited, for states are notoriously reluctant to abide by impartial findings on matters they regard as being of vital importance. Hence, what the procedures really offer is a means of slowing down the progression of a dispute toward war, giving reason a chance to prevail.

Collective action against aggressors

Collective security. Collective security is an approach to peace involving an agreement by which states agree to take collective action against any state defined as an aggressor. Leaving aside the problems of settling disputes or enforcing

law or satisfying justice, it concentrates upon forestalling violence by bringing to bear an overwhelmingly superior international force against any aggressor. Although collective security, in somewhat different forms, played a prominent part in the League of Nations Covenant and is embodied in the United Nations Charter, it has completely failed in both cases. Failing an international government capable of ultimately determining the issues, nations have not managed to agree on an unequivocal definition of aggression, have not in practice accepted the principle that aggression must be acted against independently of the identity of the perpetrator, and, therefore, have not established the international collective security force envisaged in the Charter.

Disarmament. Disarmament and limitation of armaments are based upon the theory that states are inclined to strive for dominance in arms over any potential rivals, and that this leads to arms races that tend to end in war. The major besetting sin of this theory is that it often tends to confuse cause with effect. Although arms races develop momentum of their own, they are themselves the result of political tensions leading to war. In short, it is the tensions that cause war, not the arms races. To hold otherwise is to mistake a symptom for a cause. Hence, reducing the levels of armaments does not necessarily reduce these tensions. Furthermore, it is the instability of strategic balances, rather than their level, that leads to war; agreements about disarmament or limitation of armaments may easily disturb the existing precarious balance and, therefore, be actually conducive to war.

Limiting conflict. As these major approaches to peace envisaged in its Charter have not proved very fruitful, the United Nations has developed two new procedures aiming at the limitation of wars. First, "preventive diplomacy," largely comprising the diplomatic initiatives of the secretary-general and the stationing of peacekeeping forces, has served to contain local conflicts and to prevent escalation, especially the involvement of the superpowers. Second, although the General Assembly's recommendations have no legal binding force, they have become increasingly influential, for by the mid-1970s the assembly was becoming an important agency for what has been called the collective legitimization of state policies. Resort to war becomes more costly when a state is faced with the prospects of a collective condemnation. This new restraint upon war does not, however, act upon conflicts that the assembly may favourably regard as wars of colonial liberation. Nor could the assembly's disapproval be relied upon to deter states from waging war in pursuit of an interest they deemed to be truly vital.

World government. Both the shortcomings and the limited practicability of all the approaches to the elimination of war through the reform of the international system have driven many thinkers to accept the idea that war can only be abolished by a full-scale world government.

Forcible abolition of war

No midway solution between the relative anarchy of independent, individual states and a world government with the full paraphernalia of legislative powers and of an overwhelming military force would provide a sufficiently stable international framework for the nations to feel that wars would not break out and thus stop them from behaviour that is often conducive to wars. In an age faced with the danger of a war escalating into a general extermination of mankind, the central importance of preserving peace is obvious and is generally accepted. But here the thinkers divide. Some press on from this analysis to the logical conclusion that mankind must, and, therefore, will establish a world government, and they advance ideas how best to proceed in this direction. Others regard the world government as completely utopian, no matter how logical and desirable it may be. Yet, in terms of actual policies, the adherents of the two schools do not necessarily divide. Whether they do or do not believe that world government is attainable, they do agree that the complex phenomenon of war represents a potential calamity of such a magnitude that all theorists must endeavour to understand it and to apply their understanding to the prevention and mitigation of war with all the means at their disposal.

(J.Fr./Ed.)

Law of war

The law of war has come to mean that body of international law relating to the conduct of war and to the protection of the victims of war. Its aim is to limit the suffering caused to combatants and, more particularly, to those who may be described as the victims of war—that is, noncombatant civilians and those no longer able to take part in hostilities. Thus, the wounded, the sick, the shipwrecked, and prisoners of war also require protection by law.

The laws of war have found it difficult to keep up with rapid changes wrought by the development of ever-newer weapons and more technologically advanced warfare, with their attendant damage to the natural environment. It therefore becomes important constantly to supplement (but not to abolish) earlier treaties. This section shows how the process of supplementation has been carried out.

The law of war has also been taken to include limitations placed upon states on their use of armed force. No system of law can prevent a state (or, indeed, an individual) from using force in self-defense, and the limitations of this concept are also discussed in this section.

ROOTS OF THE INTERNATIONAL LAW OF WAR

Law by treaty. In ancient times war was not subject to any control other than that exercised by the combatants themselves, and any limitations that they might have placed on their own actions on the battlefield would have been due to military necessity rather than any belief that to attack civilians or to kill prisoners of war was wrong—let alone illegal. The Viking invaders in the 11th century, for instance, knew no concept of sparing the civilian population from attack or pillage, and they did not generally protect and release captured enemy combatants. And there was no reason why they should: no treaties prohibiting brutal acts in battle had been negotiated between states, nor had there developed a uniform practice among states that considered themselves civilized to avoid such conduct. In order for such norms to develop, there had to come into existence a belief shared by a number of independent states that some limits should be placed on the methods and means of war among themselves—especially if wars were to be fought between Christian states. (Crusades against the infidel were not controlled by any similar concern.) In the Middle Ages in Europe the precepts of Christianity began to provide vague guidelines of conduct on the battlefield. In 1625 Hugo Grotius wrote *On the Law of War and Peace* (*De Jure Belli ac Pacis*), in which he explored the basic principles of the humanitarian treatment of the victims of war.

If civilians were to enjoy any protection, it would also become necessary clearly to distinguish them from the combatants. This could come only with the development of a professional army wearing a distinctive uniform and taking upon itself a code of chivalry. Certain actions would then become unchivalrous and would lead to heavy sanction from brother soldiers. Chivalry, however, did not protect the common soldier or the ordinary civilian, for whom notions of chivalry were considered inappropriate. Protection by rule of law for the lower orders had to await the acceptance of principles of humanity that took a distinctive form in the 19th century.

Until the 20th century there existed no principle of international law that limited the right of states to go to war. War was seen as an integral part of state sovereignty to be entered into for political reasons. There were, however, attempts to distinguish wars that were considered "just" from those which were "unjust." This was a Christian doctrine formulated by, among others, St. Augustine, but it was an extremely flexible one, enabling a state to describe its war as just at its own discretion. As a corollary, the enemy state would therefore be fighting an unjust war, and its soldiers could be treated in any manner by the state claiming to be fighting a just war. It was more than likely that all states involved in a single conflict would claim to be fighting for a just cause and would show an attendant lack of concern for the protection of those unable, through wounds or capture, to defend themselves.

The development of modern weapons that could cause unnecessary suffering to combatants, and the great strides made in battlefield medical care, led to a growing awareness that international cooperation was required to protect the wounded and sick. Henri Dunant, a Swiss citizen and founder of the Red Cross, was preeminent in leading a number of states to conclude the first Geneva Convention in 1864 to protect the wounded and sick. But the first attempt to codify the laws of war was drafted by Francis Lieber, a college professor in New York City. Promulgated to Union forces by President Abraham Lincoln during the American Civil War, the Lieber code was to have a profound effect on subsequent codifications of the laws of war. In 1868 the Declaration of St. Petersburg prohibited the use of explosive projectiles weighing less than 400 grams, while in 1899 two major treaties were concluded at The Hague, one concerning asphyxiating gases and another concerned with expanding bullets. The second Hague conference, in 1907, proved to be a milestone, producing 13 separate treaties. In 1925 the Geneva Gas Protocol was signed, prohibiting the use in war of asphyxiating, poisonous, or other gases and of bacteriological methods of warfare. This was followed in 1929 by two further Geneva Conventions, dealing with the wounded and sick and with prisoners of war. Following World War II yet another conference produced the four 1949 Geneva Conventions dealing, respectively, with the wounded and sick on land, with the wounded, sick, and shipwrecked at sea, with prisoners of war, and with civilians. Further treaties followed, including the 1954 Hague Convention on the Protection of Cultural Property, the 1977 United Nations Convention on Military or Any Other Hostile Use of Environmental Modification Techniques, and the two 1977 Protocols to the Geneva Conventions of 1949, extending the terms of the conventions to wars of national liberation and civil wars.

The first Geneva Convention

Law by custom. The laws of war are to be found not only in treaties entered into by states but also in customary international law, which is found in the actual practice of states and in the belief (called *opinio juris:* "opinion of the law") that that practice is in conformity with international law. Much of this customary international law has found its way into the various conventions described above. Therefore, it may properly be argued that, although a particular state is not a party to a certain treaty, it is nevertheless bound by the principle of customary international law codified in that treaty. Further, a treaty may have such wide acceptance that it can be said to reflect the practice of all states, and it may then bind all states as reflecting customary international law. As an example of this approach, the International Military Tribunal at Nürnberg in 1946 decided that the fourth Hague Convention of 1907, concerning the laws and customs of war on land, reflected customary international law; hence, its principles bound Germany even though some states, which were at war with Germany, were not parties to it.

The legally binding effect of custom

Some areas of the laws of war are not covered by treaty provisions, making it necessary to turn to other sources of international law. However, it may be that a particular point has never arisen before. In this case the Martens Clause, which first appeared in one of the 1899 Hague Conventions (and has been repeated in virtually every major treaty since), avoids any lacuna in the law by providing the following:

> Until a more complete code of the laws of war has been issued, the High Contracting Parties deem it expedient to declare that, in cases not included in the Regulations adopted by them, the inhabitants and the belligerents remain under the protection and the rule of the principles of the law of nations, as they result from the usages established among civilized peoples, from the laws of humanity, and the dictates of public conscience.

Judicial decisions are also a source of the international laws of war. The International Military Tribunals at Nürnberg and Tokyo following World War II laid down many general principles that became widely accepted, but, in fact, following that conflict a large number of other tribunals were conducted by individual states to try those charged with war crimes. In addition, a Japanese court, in the case of *Shimoda* v. *Japan* (1955), dealt with the legality in international law of the atomic bombing of Hiroshima and Nagasaki.

COMMENCING HOSTILITIES

The Covenant of the League of Nations in 1920 attempted to restrict, but not to prohibit, recourse to war. It provided that states should seek to settle their disputes peacefully by referring them to arbitration, judicial settlement, or to the Council of the League. The parties to the Covenant agreed that they would in no case resort to war until three months after the award by the arbitrators, the judicial decision, or the report by the council. It was not until the Kellogg–Briand Pact of 1928 that 63 states party to it renounced war as an instrument of national policy. This treaty was relied upon by the Nürnberg tribunal in establishing not only that there was an international crime of waging aggressive war but that international law also imposed individual liability.

Legally defining war. Two particular matters that were not referred to by either treaty were the meaning of the word war and the limits of any right of self-defense. The term war remained subjective, giving states liberty to withhold the term from their military adventures if they were so minded. (For example, in the fighting over Manchuria between Japan and China from 1937 to 1941, the Japanese refused to call the conflict a war.) As a concept, the term was left with little significance after the United Nations Charter of 1945, in article 2(4), prohibited "the threat or use of force against the territorial integrity or political independence of any State, or in any other manner inconsistent with the Purposes of the United Nations." Moreover, all the Geneva Conventions apply to armed conflicts, whether or not they are officially called wars. In the Falkland Islands conflict in 1982, for example, the United Nations Security Council (in Resolution 502) condemned the Argentine invasion of the islands as a breach of the peace, even though neither Argentina nor the United Kingdom had declared war. Upon capture by the enemy, combatants were entitled to the treatment prescribed by the third Geneva Convention of 1949.

Aggression. The Security Council of the UN is empowered by article 39 of the Charter to determine the existence of any threat to the peace, breach of the peace, or act of aggression. It may make recommendations or decide what measures (including the use of armed force) shall be taken. In practice, the Security Council often is unable to act because of the veto power possessed by its permanent members (the United States, the United Kingdom, the Soviet Union, France, and China), and it is unable to take action through the use of armed force because none of the agreements between individual states and the UN envisaged by the Charter were ever made.

In 1974, General Assembly Resolution 3314 defined and gave some examples of aggression. Article 3 gave, as examples, invasion or attack by armed forces of a state, military occupation, bombardment against the territory of another state, blockade of ports or coasts, action of a state in allowing its territory to be used for preparing an act of aggression against a third state, and the sending of armed bands, groups, irregulars, or mercenaries to carry out acts of armed force against another state. Other General Assembly resolutions, notably Resolution 2625 of 1970 (the Declaration on Principles of International Law Concerning Friendly Relations and Co-operation Among States in Accordance with the Charter of the United Nations), stress the prohibition on the use of force contained in article 2(4).

Vague definitions of armed force

Lawyers lament the imprecise definition of force as prohibited by the Charter. Three concepts appear to be used virtually interchangeably: force (and threat of force), aggression, and armed attack (this last to be found in article 51, concerned with self-defense). Article 2(4) further confuses the issue by prohibiting force "against the territorial integrity or political independence of any State, or in any other manner inconsistent with the Purposes of the United Nations." This has led to arguments—as in the Corfu Channel case between Britain and Albania in 1949 and in the attack by Israeli aircraft against an Iraqi

nuclear reactor in 1981—that although there had been a use of force in certain cases, that force was not directed against the territorial integrity or political independence of any state or against the purposes of the UN. In the Corfu Channel case, Britain insisted that it had acted only to clear Albanian mines from an international strait, and in the Iraqi case Israel argued that it had destroyed a facility that might acquire an ability to make nuclear weapons that would then threaten Israel. The International Court of Justice condemned the first action and the Security Council, the second.

It may well be that any use of armed force outside the territory of a state is a breach of article 2(4) of the UN Charter, and that the term force as used here also means aggression. Any state that uses force, therefore, will be required to show that it is doing so not out of aggression but in self-defense.

Self-defense. Article 51 of the Charter states the following: "Nothing in the present Charter shall impair the inherent right of individual or collective self-defense if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security." In *Nicaragua* v. *United States* (1986), the International Court of Justice ruled that this passage confirmed the existence of the right of self-defense under customary international law. In 1837 the *Caroline* affair, a dispute between the United States and Britain over the crossing into U.S. territory by British troops fighting Canadian rebels, led to a general acceptance that any state wishing to show that it had acted in self-defense would need to show an instant, overwhelming necessity of self-defense that left no choice of means and no moment for deliberation. In addition, the act of self-defense would need to be in proportion to the force used against it.

It should be noted that article 51 mentions a right of individual as well as collective self-defense. Following the invasion of the Falkland Islands, the British government claimed that the sending of its task force and the subsequent military action against Argentine forces was in conformity with the right of individual self-defense under article 51. On the other hand, the United States argued in 1966 that its military assistance to South Vietnam was justified as collective self-defense. The United States also tried to argue, in the case brought by Nicaragua before the International Court of Justice in 1986, that its military and paramilitary activities against that country were in collective self-defense with Costa Rica, El Salvador, and Honduras. The court decided, however, that there "was no rule permitting the exercise of collective self-defense in the absence of a request by the State which regards itself as the victim of an armed attack." (At the relevant time, none of these three Central American states considered that an armed attack had occurred against it.) In addition, the court considered that, in order for the right of collective self-defense to apply, the attacked state would have to request assistance from the state claiming to act in collective self-defense with it. Because this had not occurred, the United States could not justify its actions against Nicaragua as collective self-defense under article 51 of the Charter.

It is not clear whether anticipatory self-defense is permitted under the Charter. Read literally, article 51 requires an armed attack actually to have occurred before a state can act in self-defense. If, however, the "inherent right" expressed in article 51 allows customary international law to be considered, then it may be argued that a state does indeed have a right of anticipatory self-defense. The Nicaragua case seemed to suggest this interpretation, and it was used very effectively by Israel in June 1967 when it destroyed much of the Egyptian air force on the ground prior to an anticipated Egyptian attack on Israel. The United States justified its air attack on military targets within Libyan cities in 1986 by claiming that such action was taken to prevent terrorist attacks on Americans in the future. Moreover, it may be argued that a state claiming to be acting in self-defense can take into account the accumulation of hostile acts that have been committed against it in assessing the proportionality of its response.

Anticipatory self-defense and reprisal

There is here a very fine line dividing anticipatory self-defense, which may be legally permissible, from reprisal, the prime object of which is to punish an alleged wrongdoing and which is not legally permissible. The destruction by Israel of 13 civilian aircraft in Beirut, Lebanon, in 1968 was condemned by the UN Security Council as a reprisal, since the raid was in retaliation for the attack on an Israeli aircraft at Athens in which one Israeli citizen was killed. (Because the Security Council is not a court of law, it does not automatically follow that its condemnation of military action signals its illegality.)

A further problem with the definition of self-defense in article 51 is the extent to which a state may intervene by military force into the territory of another state in order to rescue its nationals who are threatened there. In 1976 members of the Israeli Defense Force entered, without permission, the territory of Uganda to rescue Israeli nationals who had been hijacked while traveling on a civilian airliner by a terrorist organization and who were being kept hostage at Entebbe airport near Kampala. There was some evidence that the Ugandan authorities had lent some assistance to the hijackers. The Entebbe raid was not condemned by the Security Council, and many writers on international law considered the raid justifiable. Of course, there were a number of distinctive features that made this a clear-cut legal case: The Israelis used minimal military force against a state that appeared to be assisting terrorists, and they left as soon as their citizens were under their control. In the Grenada incident in 1983, the United States sent armed forces to recover U.S. citizens from the island when the government had ceased to exist after its principal members were killed. U.S. forces remained on the island until elections were held, and then they were withdrawn. The Grenada invasion may be less clearly justifiable than the Entebbe raid, but, because the United States (along with the armed forces of other states in the region) took action at the request of Grenada's governor-general, the invasion may have been justifiable under international law (although not all international lawyers would agree).

International and internal conflicts. *Regional action.* Chapter VIII of the UN Charter permits the existence of regional arrangements or agencies for dealing with such matters of international peace and security as are appropriate for regional action. It goes on to provide, in article 53, that no enforcement action shall be taken under regional arrangements or by regional agencies without the authorization of the Security Council. Article 54 states that the Security Council shall be kept informed of all such activities. On a number of occasions, states have justified the use of force (or the threat of force) under this part of the Charter, despite the lack of prior authorization from the Security Council, by arguing that the measures they took did not amount to enforcement action and therefore did not require the authorization of the Security Council. Thus, the United States, after stopping ships on the high seas in 1962 to search them for missiles or missile parts intended for Cuba, argued that this was not enforcement action since the regional arrangement (in this case, the Organization of American States) had merely made a recommendation to member states and had not rendered a decision that had to be enforced. A similar argument was used following the Grenada incident: this action, the United States declared, was not directed against a government but was merely carried out to restore law and order to the island under the aegis of the Organization of Eastern Caribbean States.

War by proxy. Armed conflict need not be, and often is not, of the traditional type—that is, a conflict between regular armed forces in the territory of one or more states. *Nicaragua* v. *United States* showed that an armed attack (which would give the attacked state the right to act in self-defense) must be understood as "including not merely action by regular armed forces across an international border, but also the sending by or on behalf of a state of armed bands, groups, irregulars or mercenaries, which carry out acts of armed force against another state of such gravity as to amount to an actual armed attack conducted by regular forces, or its substantial involvement therein." Therefore, if a state sent an armed band into another state

to depose its rulers or to attack civilians of that state, then the sending state would have committed an armed attack, giving the attacked state the right to act in self-defense. As discussed above, the response must be proportionate to the aggression; in assessing this, the accumulation of events may be taken into account.

Civil war. The term civil war, although perhaps dated, is used here to mean a noninternational armed conflict. It therefore covers any internal conflict, whatever the motive for the fighting.

Foreign involvement in a civil war

It is often difficult to determine whether a conflict is truly internal or international, since other states may be involved to some extent. If it is indeed an international armed conflict, then an attacked state may seek the military assistance of any other state, which will then be acting in collective self-defense with it. (An example of this was the Vietnam War, although, it should be said, many states regarded it as a civil war.) Also, if the conflict has become international, then the 1949 Geneva Conventions and the whole of the body of the laws of war will apply to the combatants as well as to civilians caught up in the conflict. Should the war be a civil one (which can properly be described as an armed conflict), international law would point to the nonintervention of other states, and only article 3 of each of the 1949 Geneva Conventions would apply (protecting only those not taking an active part in the hostilities). Further protection is given (mainly to those who do not take part in the conflict) by the second Protocol of 1977, which applies to civil wars in which dissident armed forces, under responsible command, exercise such control over a part of the territory of a contracting state as to enable them to carry out sustained and concerted military operations and to implement the Protocol. For these reasons, the Protocol would not apply to the conflicts in Northern Ireland or Spain, in which neither the Irish Republican Army nor the Basque separatists controlled any territory, while it would apply in the conflict in El Salvador, in which rebels controlled sizable areas of the countryside.

War of national liberation. The first Protocol of 1977 provides that peoples fighting against colonial domination and alien occupation and against racist regimes in the exercise of their right of self-determination are to be treated as if they were engaged in an international armed conflict and not a civil war. There is considerable difficulty over the meaning of this phrase, and it may be difficult to apply in practice.

CONDUCTING HOSTILITIES

Treaties regulating the conduct of war

Although the Hague Conventions, concerning the conduct of hostilities, apply to the states that are party to them in the event of war, the various Geneva Conventions of 1949 (and the 1977 Protocols to them) come into operation where there is an armed conflict between two or more contracting parties even if a state of war is not recognized by one (or both) of them. They also apply to the occupation of another state's territory even if the occupation meets with no armed resistance. Since much of the Hague Conventions reflect customary international law, it can be assumed that these laws of war (or the *jus in bello*) also apply whether or not any declarations of war exist. In considering the legal conduct of a conflict, the laws of war take no account of its causes. This means that the combatants of the aggressor nation are owed the same rights as those of the attacked state.

The controls placed on the actual methods and means of war are to a large extent based on the Hague Conventions, but there are also a number of important provisions in the first Protocol of 1977, the 1954 Hague Convention on cultural property, and the 1981 Conventional Weapons Convention.

Lawful combatants. Those who may lawfully take part in hostilities are those who would be entitled to prisoner-of-war status if captured. Any other person taking part in a conflict may be treated as an unprivileged belligerent, or a franc-tireur, and he may be punished if captured. Article 4 of the third Geneva Convention of 1949 and article 43 of the first Protocol of 1977 provide that a lawful combatant is generally a member of the armed forces of a state. The term also includes members of the merchant marine and inhabitants of unoccupied territory who, on the approach of the enemy, spontaneously take up arms to resist the invading forces until the territory has been occupied.

A spy is in a unique position, since he is often a member of the armed forces of a state; but if he acts in disguise in the zone of operations of an enemy in order to obtain information to pass on to his own forces, he may be punished provided he has a trial.

A mercenary is not protected at all; he has the right to be neither a combatant nor a prisoner of war. A mercenary is defined in the first Protocol of 1977 (which neither the United Kingdom nor the United States has ratified) as a person who is specially recruited to take part in a conflict, who is motivated essentially by private gain, and who is paid substantially more than the ordinary armed forces of the state to which he has been recruited. He must not be a national of the recruiting state or a member of the armed forces of a party to the conflict.

Legally defining a guerrilla fighter

Guerrilla fighters are not solely a modern phenomenon, although during and after World War II they became a common feature of armed conflicts, especially those occurring in the developing world. The third Geneva Convention of 1949 required what is called an organized resistance movement to possess four characteristics before its members could be treated as prisoners of war upon capture. These were: (1) being commanded by a person responsible for his subordinates, (2) having a fixed and distinctive sign recognizable at a distance, (3) carrying arms openly, and (4) conducting operations in accordance with the laws and customs of war. In time, it became apparent that two of these four conditions were difficult for guerrilla fighters to meet. Were guerrillas to wear a fixed and distinctive sign recognizable at a distance or carry arms openly, they could hardly operate with any safety in occupied territory. The first Protocol of 1977 made a number of important changes that bind those states that are parties to it. For example, one of the major problems with recognizing guerrilla fighters as lawful combatants is that they may not, in fact, distinguish themselves from the civilian population—in which case, all civilians are placed at risk. Therefore, article 43 of the Protocol requires all combatants to distinguish themselves from the civilian population while they are engaged in an attack or in a military operation preparatory to an attack. However, even if a combatant does not do this, he will still be entitled to treatment as a lawful combatant if he carries his arms openly during each military engagement and during such time as he is visible to the adversary while engaged in a military deployment preceding the launching of an attack in which he is to participate.

A member of the armed forces of a party to a conflict will lose his status as a prisoner of war upon capture if he commits an act of hostility while wearing civilian clothes. In the case of *Osman Bin Mohammed* v. *Public Prosecutor* (1968), the Privy Council in London held that members of the Indonesian armed forces who had landed in Singapore during an armed conflict between Indonesia and Malaysia were not entitled to be treated as prisoners of war after having placed a bomb in a civilian building that caused the deaths of civilians. This loss of prisoner status will also apply, among the states that are parties to the first Protocol of 1977, if their combatants do not at least carry their arms openly, as described above.

Limits on the methods and means of war. *Weapons.* Article 22 of the Regulations Annexed to the Hague Convention of 1907 provides that "the right of belligerents to adopt means of injuring the enemy is not unlimited." This particular principle underpins much of the law in this area, and there are many examples of it. Article 23 of the same treaty, for instance, prohibits certain activities such as the employment of poison or poisoned weapons, killing or injuring enemy combatants treacherously, attacking those who have surrendered, or declaring that no quarter will be given. It also prohibits the employment of arms, projectiles, or material calculated to cause unnecessary suffering. One reason for this approach, as stated in the Declaration of St. Petersburg of 1868, is that "the only legitimate object which states should endeavour to

accomplish during war is to weaken the military forces of the enemy."

This principle explains, to some extent, the prohibition on the use of certain weapons. Hence, the use of chemical and bacteriological weapons was banned by the 1925 Geneva Protocol. By the Bacteriological Weapons Convention of 1972, states party to it agreed never in any circumstances to develop, produce, stockpile, retain, or acquire bacteriological or biological weapons or toxins. If a ban on chemical weapons came about, it would likely take the same form.

The legal status of nuclear weapons

The use of nuclear weapons against enemy combatants is not subject to any express prohibitions. A number of international lawyers, however, take the view that their use is implicitly prohibited by the principles stated above, because radiation effects can be considered not only a form of poison but also a weapon calculated to cause unnecessary suffering. The General Assembly of the United Nations condemned their use in Resolution 1653 of 1961, but the value of this resolution is considerably weakened by the fact that, of the nuclear-weapon states, only the Soviet Union voted for it. In *Shimoda* v. *Japan* (1983), a Japanese court held that the use of atomic weapons against Nagasaki and Hiroshima was contrary to international law, not merely because of the type of weapon used but because bombardment, by any means, of the civilian population of those two cities was contrary to the Hague Conventions of 1907.

Like nuclear weapons, incendiary weapons are not specifically banned unless used against the civilian population. It might be argued, however, that their use against enemy combatants (as opposed to military equipment) would infringe the 1925 Geneva Gas Protocol, since they could come within the prescription of "all analogous liquids, materials, or devices."

The Vietnam War illustrated the dangers that modern weapons can cause to the environment. The use in that conflict of chemical herbicides and other methods of deforestation, along with attempts to alter weather patterns, called the attention of the world to such activities. The result was the 1977 United Nations convention on environmental modification, which requires states not to engage in military or any other hostile use of environmental modification techniques having widespread, long-lasting, or severe effects. The first Protocol of 1977 also prohibits the employment of methods or means of warfare that are intended, or may be expected, to cause widespread, long-term, and severe damage to the natural environment. States are specifically directed by this protocol to consider whether any new weapons that they might develop would infringe any rules of international law.

On the seas, naval forces may attack enemy warships. The sinking of the Argentine warship *General Belgrano,* therefore, was not contrary to international law despite its being attacked outside the Total Exclusion Zone that the British government had declared around the Falkland Islands.

Civilians. According to customary international law, only members of the armed forces of a party to a conflict can take part in hostilities, and the law has always attempted to draw a clear distinction between the lawful combatant, who may be attacked, and the civilian, who may not.

Ban on attacking civilians

One of the Fundamental Rules of International Humanitarian Law Applicable in Armed Conflicts, which were prepared by the International Committee of the Red Cross in 1978, requires parties to a conflict to distinguish at all times "between the civilian population and combatants in order to spare civilian population and property. Neither the civilian population as such nor civilian persons shall be the object of attack. Attacks shall be directed solely against military objectives." Restrictions on the use of chemical or nuclear weapons against the civilian population have been discussed above. In addition, the 1981 Conventional Weapons Convention specifically prohibits the use of mines, booby traps, and other similar devices and incendiary weapons directed against the civilian population or used indiscriminately, and the first Protocol of 1977 imposes very detailed target restraints in order to protect civilians. For example, aerial bombardment engaged in for the sole purpose of terrorizing the civilian population is prohibited, and the use of aircraft to carry out such a role would therefore be illegal. Merchant ships may in limited circumstances be attacked, but they may not be sunk by a submarine without its first having placed passengers, crew, and ship's papers in a place of safety.

Neutrals. The fifth Hague Convention of 1907 declares that the territory of neutral powers is inviolable and that a neutral state has a duty to prevent a belligerent state from carrying the conflict to its territory. In particular, troops belonging to the army of a belligerent state who enter the territory of a neutral must be interned. Also, a neutral must act evenhandedly to all belligerent states; for this reason, the United Kingdom declared its neutrality in the war between Iran and Iraq (1980–88), refusing to sell either side military equipment that would have significantly enhanced its capability to prolong the conflict.

Neutral shipping may be stopped on the high seas (as occurred in the Iran–Iraq War when a British merchant vessel was stopped by an Iranian warship) to check on the carriage of contraband. In naval warfare, the 13th Hague Convention of 1907 bans belligerents from conducting military operations in the territorial waters of a neutral state, and neutrals themselves have duties imposed on them not to assist the warships of belligerent states.

Prohibited areas of combat. Military activities of any kind cannot be carried out on the Moon (the Moon Treaty of 1979), Antarctica (the Antarctic Treaty of 1959), or on the territory (including the airspace) or territorial waters of neutral states. In addition, nuclear weapons or other weapons of mass destruction cannot be orbited around the Earth (the Outer Space Treaty of 1967) or placed on the seabed (the Seabed Treaty of 1971).

Prisoners of war. The third Geneva Convention of 1949 provides the basic framework of protection accorded to a prisoner of war. He is protected from the moment he falls into the power of an enemy until his final release and repatriation. No form of coercion may be inflicted on him to secure information of any kind; he need only give his name, rank, date of birth, and serial number. When an Argentine army officer captured by British forces during the Falklands conflict was alleged to have been responsible for the disappearance of French and Swedish nationals in Argentina prior to the conflict, he could not be compelled to disclose information on the subject and was released.

The rights of prisoners of war

A prisoner of war is entitled to decent and humane treatment, to be evacuated from the combat zone, and to be granted rights and duties as similar as possible to those of the armed forces of the detaining power. No reprisals may be taken against prisoners of war; they may not be treated in a way contrary to the Convention even though an enemy state treats its prisoners of war in such a way. Officers may not be compelled to work, and other ranks may not be compelled to do dangerous or unhealthy work. Article 52 of the third Convention of 1949 goes on to provide that the removal of mines or similar devices shall be considered dangerous labour.

In order to ensure that prisoners of war are accorded the treatment laid down in the Conventions, states must ensure that a protecting power is appointed to act on their behalf. A protecting power is a neutral state acceptable to the state that holds prisoners of war. There were no protecting powers appointed during the Vietnam War or the Iran–Iraq War, but in the Falklands conflict Switzerland acted for the United Kingdom and Brazil for Argentina. A state may allow the International Committee of the Red Cross (ICRC) to act as a substitute protecting power. The ICRC has, in addition, a right to visit prisoner-of-war camps.

Protecting powers (or the ICRC) must be kept informed if a prisoner of war is to be tried (rather than being given disciplinary punishment) for an offense, in order, for instance, that the protecting power might find the accused a lawyer. If the death penalty is imposed, it cannot be carried out for at least six months after the judgment and after sentence has been communicated to the protecting power. A prisoner of war may be tried for an offense committed prior to capture (such as a war crime), but he

is entitled to retain his status as a prisoner of war even if convicted.

The use of weapons against prisoners of war attempting to escape constitutes an extreme measure and is to be preceded by warnings. The detaining power must hold an inquiry into the death of a prisoner of war and notify the protecting power. Such an incident occurred in the Falklands conflict, when a British soldier shot and killed an Argentine prisoner of war whom he believed was attempting to escape. The resultant inquiry exonerated the soldier, and a report was passed to the ICRC.

At the conclusion of hostilities prisoners of war are to be repatriated. Problems occurred at the conclusion of the Korean War when a number of North Koreans did not wish to return. A repatriation commission was established in 1953, and remaining prisoners of war were transferred to it. It has become more common to repatriate able-bodied prisoners of war before the end of hostilities. To a limited extent this occurred in the Iran–Iraq War, but it was a major feature of the Falklands conflict.

Occupation. World War II illustrated that civilians in occupied territory were largely unprotected by the laws of war. In consequence, the fourth Geneva Convention of 1949 provided detailed rules for their protection. A protected person is anyone who, at a given moment and in any manner whatsoever, finds himself, in case of a conflict or occupation, in the hands of a party to the conflict or occupying power of whom he is not a national. The inhabitants of occupied territory are, therefore, protected persons under the Convention; they are entitled to humane treatment and to respect for their person, honour, family rights, religion, manners, and customs. Article 34 of the fourth Convention specifically prohibits the taking of hostages and reprisals against them or their property. Article 49 prohibits the transfer of protected persons out of occupied territory unless, in a given area, the security of the population or imperative military reasons so demand. After the war of June 1967, Israel occupied territory in the West Bank, the Gaza Strip, and the Golan Heights, but it claimed that the fourth Convention did not apply to them. The United Nations took a different view in resolutions in 1988 when it specifically declared that the Convention was applicable to all the Palestinian and other Arab territories occupied by Israel since 1967. The resolutions went on to condemn a number of Israeli practices in these territories, such as the killing, wounding, and deportation of Palestinian civilians (who are protected persons under the fourth Convention), during uprisings against Israeli rule.

Legally protected persons under military occupation

The occupying state may make such laws for occupied territory as enable it to carry out its obligations under the Convention, to maintain the orderly government of the territory, and to ensure its safety. At the same time, it must respect other laws in force before the occupation. Requisitions for the needs of the occupying army may be taken, but only on payment, and foodstuffs and medical supplies may be requisitioned only if the needs of the civilian population have been taken into account. If the supply of such items is inadequate for the needs of the civilian population, then the occupying state will be under an obligation to bring them into the territory. The Nürnberg trial concluded that "the German armies were to be fed out of Soviet territory, even if many millions of people were to starve to death." It is this type of conduct that the fourth Convention attempts to prevent.

Protected persons who are not members of the armed forces and who use force against occupying forces are not entitled to special treatment, since they are not entitled to prisoner-of-war status upon capture. The occupying state may place them on trial for breach of either the ordinary laws of the territory or the laws it has imposed. However, if it is to sentence such a person to death, it must take into account that the protected person does not owe the occupier any duty of allegiance. Also, a state that occupies territory does not thereby obtain good title to it. Various UN resolutions confirm this; a General Assembly resolution in November 1988 reaffirmed that the "occupation by Israel of the Palestinian territories since 1967, including Jerusalem, in no way changes the legal status of those territories."

CESSATION OF HOSTILITIES

Hostilities may be suspended pending negotiation between the parties. Negotiation may, or may not, be preceded by the display of a white flag, which merely means that one side wishes to enter into communication with the other. The parties may then enter into an armistice, and, when all matters are agreed, a peace treaty may be concluded. Of course, it is possible to end hostilities without any treaty; neither the Falklands conflict nor the Iran–Iraq War ended in this way, although an agreement sponsored by the UN provided for the withdrawal of Soviet troops from Afghanistan in 1989.

It has been shown that the acquisition of territory as a result of a war of aggression does not give title to that territory under international law. In the same way, a treaty by which a victor (who has started a war of aggression) requires a vanquished state to cede to it territory would not be considered a valid transfer of sovereignty over the territory concerned.

WAR CRIMES

The term war crime has no definite meaning. It was commonly thought of as a violation of the laws of war committed by a combatant or even a civilian. In 1945 the charter of the Nürnberg tribunal gave that court jurisdiction to try crimes against the peace (which consisted of waging a war of aggression), war crimes (that is, violations of the laws and customs of war), and crimes against humanity (such as the murder and ill-treatment of civilians). Twenty-two persons were charged at Nürnberg and 25 at the Tokyo tribunal, but many more were tried by tribunals established by Allied governments in territory they occupied at the conclusion of World War II. The tribunals had a profound effect on the development of international law as it is concerned with the responsibility of both states and individuals for conduct leading to and during war. In particular, the tribunal confirmed that individuals could be held liable for a breach of international law: "Crimes against international law are committed by men, not by abstract entities, and only by punishing individuals who commit such crimes can the provisions of international law be enforced."

Importance of the Nürnberg and Tokyo tribunals

One problem associated with the tribunal was that of its jurisdiction. Did international law (upon which the framers of the charter relied) permit states to try the nationals of another state for committing crimes under international law in the territory of yet other states? The charter decided that it did, since it was concerned with offenses having no particular geographic location. A wider view of international law was taken in the case of *Attorney General of the Government of Israel* v. *Eichmann,* which was decided by the District Court of Jerusalem in 1961. Adolf Eichmann, head of the Jewish office of the Gestapo during World War II, was convicted of war crimes, crimes against the Jewish people, and crimes against humanity. Although the crimes were not committed on the territory of Israel (which at the time did not exist as a state), the court held that such acts could be tried by any state that had custody of the defendant. (Eichmann had, in fact, been abducted from Argentina by Israeli agents.)

The Nürnberg tribunal also had to consider arguments put forward by the defense. Many defendants pleaded that their actions were carried out on the orders of superiors. The framers of the charter realized that this was likely to be a major issue, and they added to the charter an article which stated that superior orders would not relieve a defendant of liability but could be considered in mitigation. For this reason no convictions were brought against those responsible for bombing Allied cities or for the waging of unrestricted submarine warfare.

The Nürnberg principles were affirmed by the United Nations in 1946. In 1948 the United Nations prepared a Convention on the Prevention and Punishment of the Crime of Genocide, and in 1968 it offered for signature a convention that removed the statute of limitations from war crimes and crimes against humanity.

Postwar conventions on war crimes

The four Geneva Conventions of 1949 take a different approach to trying those responsible for breaches of the laws of war during an armed conflict. Each Convention

lists a number of "grave breaches," which include willful killing, torture or inhuman treatment, and the causing of great suffering or serious injury to body or health. States party to the Conventions undertook to enact legislation to try those suspected of grave breaches and to search for such persons. The United Kingdom, for instance, enacted the Geneva Conventions Act of 1957, making it a criminal offense for any person to commit a grave breach of the Conventions anywhere in the world. The first Protocol of 1977 adds to the list of grave breaches, such as making the civilian population or individual civilians the object of attack, launching an indiscriminate attack affecting the civilian population, the perfidious use of the distinctive emblem of the Red Cross, and the transfer of protected persons from occupied territory (as discussed above in relation to Israel).

The Protocol also provides for the establishment of fact-finding commissions to inquire into any allegation of a grave breach. Allegations of war criminality were made in regard to actions committed during the Korean War, the Vietnam War (especially the killing of prisoners of war), and the Iran–Iraq War, but no conviction for a grave breach of the Geneva Conventions has been recorded. Individual members of the armed forces may instead be tried by court-martial for a breach of their domestic penal or military law. For example, in 1947 a British army medical officer was convicted by court-martial for the ill-treatment of German nationals held when the United Kingdom occupied parts of Germany, and in 1971 a U.S. army lieutenant was convicted of murder for his part in the massacre of villagers in South Vietnam. (P.J.R.)

Military law

All states require a code of laws and regulations for the raising, maintenance, and administration of their armed forces. All this may be considered the field of military law. The term, however, is generally confined to disciplinary military law, that part of the code that aims at and sanctions the maintenance of discipline in the armed forces. In the past this was also known by the name of martial law, a term that now has the meaning of military enforcement of order upon a civil population either in occupied territory or in time of disorder. Members of armed forces do not cease under modern conditions to have duties as citizens and as human beings. All systems of military law must aim to ensure that the soldier is in no way enabled to escape the obligations of his country's ordinary law or of international law as recognized in various conventions.

HISTORICAL DEVELOPMENT

The object of the disciplinary code is to ensure that the will of the commander is put into effect. Military law therefore traces its origins to the prerogative power of rulers. In Rome, just as a sector of civil law developed from the imperium of the magistrates, so did military law derive from the imperium of those same magistrates in their capacity as commanders of the military forces. The Roman historian Tacitus indicates that military justice in the 1st century AD was somewhat rough-and-ready and heavy-handed and varied much with the individual commander. But it became more formalized 400 years later in the Digest and Codex of the emperor Justinian. With the rise of the kingdoms of the Middle Ages, the maintenance of discipline was enforced by ordinances or articles of war issued by the sovereign or by a commander authorized by him at the beginning of each campaign. The earliest now extant are those of the English king Richard I in a charter of 1189 for the government of those going to the Holy Land.

Medieval articles of war

With mercenary armies drawn from many nations in the wars of the 16th and 17th centuries, each national contingent tended to apply the articles of the supreme commander according to its own rules of procedure. The articles of war of Maurice of Nassau, prince of Orange, and Gustav II Adolf had a considerable influence on the national commanders who served under them, when they came to command elsewhere. In the English Civil Wars, the ordinances of the royalist and the parliamentary commanders were thus in the most part literally the same and in the next reign formed the basis of Prince Rupert's code of 1672. The famed discipline of Cromwell's army was due not to any improved code but to the fact that the articles were rigorously enforced. On the continent of Europe, the articles of Gustav Adolf continued to be followed until supplanted by the codification of the 19th century, which established throughout those countries a generally similar system that, with revision and amendments, continues to this day.

With the introduction of a standing army in England in 1689, Parliament aimed to prevent this force coming under complete control of the sovereign by a series of mutiny acts, normally passed annually, to which the prerogative articles were subordinate. By a statute of 1717 the power to make articles was embodied in the act. In the United States in 1775 and again in 1806, articles of war were adopted that were modeled upon the mutiny acts and articles then in force in Great Britain. In the British army, the articles of war were replaced in 1881 by an annually renewed Army Act (reformed in 1955), although they continued in the Royal Navy until 1957. In the United States they were replaced by the Uniform Code of Military Justice in 1951.

JURISDICTION

Persons subject to military law. The jurisdiction of military law is not necessarily confined to offenses injurious to the discipline of the forces committed by members of those forces. It extends in various countries in varying degrees to all offenses committed by members of the forces and to offenses injurious to discipline committed by persons who are not members of those forces.

Military personnel. In countries in which an obligation to military service exists, soldiers who fail to answer their initial call-up or report for duty are liable to military jurisdiction for such offenses as desertion or self-mutilation either because the military code makes such offenses applicable to them as a class of civilians (as in Belgium, France, Italy, and Israel) or because under the act introducing national service they are deemed to be enlisted on the dispatch of a calling-up notice (as was the case in Great Britain when conscription was in force). They continue to be liable for such offenses, even if not otherwise subject to military law, during authorized absence from the conscripted service or temporary reserve service. Reservists are also subject (as in Italy) to military jurisdiction for such offenses as treason, communicating with foreign countries, and revelation of official secrets. In Belgium, released soldiers remain liable for rebellion or offenses against superiors committed within one year of their release.

Legal status of conscripts

Civilians. Civilians may become subject to military jurisdiction in any number of ways. In Italy and Turkey, for example, treason or rebellion can be dealt with under the military code, and in Norway breaches by a civilian of the Geneva Conventions of 1949 and their additional Protocols of 1977 are dealt with under military law. In other countries, civilians who instigate or participate in military crimes may themselves be triable under military law. In a number of countries, civilians within a war zone or theatre of active operations, or in conditions defined as a "state of siege," can come under military jurisdiction for offenses similar to those mentioned above—or even completely under military jurisdiction, as in Argentina.

In other countries, only civilians associated with the armed forces may be triable under service law. In Israel, for example, civilians who are employed by the army, or who have been provided with army weapons, are subject to military law, as are those held in army custody. Under British military law, civilians accompanying armed forces stationed in a foreign country (including families of soldiers as well as British civilians working for or with the services) are triable under offenses against the good order of the military community. In the United States, however, civilians—even those forming part of a service community abroad—cannot in peacetime be tried at all under the military process, though they may become subject to military jurisdiction in time of war. Austria and Spain

are among countries in which no civilian can be liable to military jurisdiction.

Prisoners of war. Also among those who fall under military jurisdiction are prisoners of war. Sometimes, as in France, Belgium, and Luxembourg, they are expressly included among those to whom the ordinary military law applies; elsewhere, special regulations concerning their behaviour and trial must be passed. Under the third Geneva Convention of 1949, prisoners of war must be tried by a military court, except where the laws of the belligerent expressly allow a member of the belligerent's armed forces to be tried by a civil court for the same offenses. Prisoners of war must not be sentenced to any penalties other than those which might be inflicted on members of the forces of the detaining power for the same act.

Differences between common-law and continental systems

Offenses against military law. The military law of the Anglo-American countries and of countries deriving their military law from them, such as India and other independent members of the British Commonwealth, differs from that of the majority of the continental countries in that the latter tend to divide military offenses into two classes: crimes that are the subject of judicial punishment and, second, breaches of discipline that are subject only to administrative action. The former group of countries (and a few others, including most communist countries) recognize no such distinction, regarding all military offenses as crimes. Apart from offenses of a peculiarly military nature, such as mutiny, insubordination, desertion, and misconduct in action or in performance of service duties, when an act committed by a soldier constitutes an offense in the civil code, it will frequently constitute an offense of which military law takes cognizance. In the Soviet Union and Belgium, for example, all civil offenses committed by soldiers, except very minor ones, are tried by military court. In France, Germany, Austria, and Scandinavia, in peacetime, all crimes, military or civil, are dealt with by civil courts. Great Britain, Canada, and other countries include as military crimes all actions committed by soldiers anywhere that would be offenses against the criminal law of their own country, although the most serious of these cannot be tried by a military court unless committed abroad, or in India at specified Frontier Posts. In the United States, because of the differences between the criminal law of different states, certain civil crimes are specifically made offenses against the military code. All countries have rules to prevent the double jeopardy of an offender being punished for one act by both civil and military jurisdiction. Generally, when civil jurisdiction may be exercised, this takes precedence over military jurisdiction.

PROCEDURE

Summary punishment. In both Anglo-American and continental systems, soldiers are subjected to penalties imposed summarily as well as to those imposed by courts. In the majority of countries, summary penalties can be inflicted only by officers not lower than the rank of captain, the commanding officer of the military unit being the principal source of discipline. The forms of punishment so inflicted are normally loss of privileges for a specified period, fines, or deprivation of liberty. Higher military commanders usually have power to deal summarily with officers (normally up to the rank of major), though in some countries these will not be liable to loss of liberty.

Appeal. Under the British and some other systems of military law, if a commanding officer has it in mind to award a punishment beyond a certain degree of severity (usually including deprivation of liberty), he must first offer the accused the option of being tried by a court-martial. Also, in Britain, the United States, and other common-law countries the accused may complain of unjust or unduly severe punishment to the commanding officer's superiors. Such complaints may proceed to the highest level, in effect achieving a review of the commanding officer's award. In other countries the soldier may appeal to a tribunal; in yet others, such as Norway and Sweden, he may have a right of appeal through the chain of military command up to a certain level (the brigade commander in Norway; in Sweden, the regimental commander) but, beyond that, to a tribunal (in Sweden, the county court).

Between summary action and trial. In the Anglo-American countries, offenses beyond a commanding officer's powers are dealt with by a service court (court-martial). In the continental countries, military crimes and similar offenses are also dealt with judicially. In the latter, however, there is an intermediate form of tribunal that deals with the more serious breaches of discipline and may impose punishments affecting a soldier's career, such as loss of rank or appointment, dismissal, or forfeiture of pension rights. In Germany, military courts of service, which also hear soldiers' complaints, may impose career sanctions. Also to be noted are the unit assemblies in the Soviet Union, which are general assemblies of all military members of a unit. All those attending are invited to comment on the behaviour under investigation. An elected tribunal, normally of five, decides whether the public examination is in itself sufficient educational corrective of the aberrant behaviour or whether the offender should be reported to a competent commander for disciplinary measures.

Examination by peers in the Soviet military

Court-martial. *Pretrial procedure.* Military courts follow judicial procedures no less formal than those of the higher civil courts. There is always some form of preliminary investigatory procedure that fills a role similar to that of the committal proceedings in the British legal system, the grand jury in the United States, and the juge d'instruction in continental systems. Under the British system, and those of Commonwealth and other countries deriving from it, it is the accused's commanding officer who is responsible for the conduct of this quasi-judicial investigation, having the evidence reduced to writing, considering it, and deciding whether it justifies his remanding the accused for trial by court-martial. Under other military legal systems, the preliminary investigation is likely to be in the hands of a military magistrate and set in motion by a military procurator, who corresponds to the official responsible in such countries for initiating civil prosecutions on the public behalf. In Israel, whose military judicial procedures otherwise derive from the British model, the responsibility for both the investigation and the decision to proceed to trial rests with a military advocate, the commanding officer being excluded altogether from the investigative process and forbidden to interfere with it.

Protecting the accused. Pending trial, all countries maintain a presumption of the accused's innocence. He must be allowed full facilities for preparing his defense, and there are normally safeguards provided to protect him from being held unjustifiably in arrest before trial. In some systems his arrest must be ordered and authorized by a magistrate, usually for a limited period only. Where the accused's commanding officer is empowered to authorize arrest, he is likely to be obliged to report the progress of the case at specified and frequent intervals to higher authority, so that the need to retain the accused in arrest can be constantly monitored.

Composition of the court. Courts-martial are generally composed, depending on the type of case, of between three and seven judges; these are usually military officers, though in some countries the membership of the court may include other ranks and even civilian judges. In the United States, for instance, the accused enlisted man may require that not less than one-third of the court be made up of enlisted men. In the Soviet Union the court includes popular assessors, elected from the ranks of the unit and holding office for two years. British military law provides for the court to include civilian crown servants when the accused is a civilian, one if the court is a district court-martial and two if the trial is by general court-martial.

The role of professional lawyers

The military courts of most countries embody at least one lawyer, who may be a legally qualified serving officer or a civilian and whose role may be either that of a participating member of the judicial tribunal (sometimes its president) or that of a legal adviser to a tribunal composed of lay military men. The judicial independence of the professional lawyers, where they serve as participating judges, is commonly safeguarded by their appointment on a fixed tenure of office. In Israel, for example, a legally qualified officer on a five- to seven-year tenure sits as president with two lay officers. The Belgian military court consists of a civilian judge on a three-year tenure sitting

with four serving officers. In Italy two permanent civilian judges sit with one military officer who is selected by lot for a two-month tour of duty as a member of the court. In France the military tribunal consists, in wartime, of two civil and three military judges (since 1983 French soldiers in peacetime have come entirely under civil jurisdiction). In the Soviet Union, the professional element is supplied by the presidents, vice presidents, and members of military courts, who are nominated by the Presidium of the Supreme Soviet; in courts of first instance it is one of these appointed professionals who sits with the two popularly elected assessors.

In those courts in which the lawyer sits as a legally qualified judge, he takes part with the other members of the tribunal in deliberating upon the court's finding, as is usually the custom in civil trials in their countries. The other mode of trial, in which the lawyer is advisory to a court-martial of laymen, is more common in countries accustomed to the Anglo-American mode of jury trial, where the professional judge, having instructed a lay fact-finding body (the jury) as to the principles of law they must apply, takes no part in their subsequent deliberations. In a similar manner, the legal adviser to the court-martial sums up the law and the facts in open court and then retires, leaving the members of the tribunal to their own discussions and returning only when they announce their finding. The adviser normally remains present during the court's subsequent discussions on sentence, but only as an adviser, having no vote. In Britain and the countries of the Commonwealth, this legal adviser to a court-martial is termed a judge advocate. The British judge advocate is almost always a member of the judicial staff of the judge advocate general, a civilian official responsible to the lord chancellor and, thus, entirely independent of the service authorities. Many Commonwealth countries also make use of a civilian judge advocate. In the United States the erstwhile legal adviser to the court-martial has been replaced by a military judge, who is a serving officer but is part of an independent military judiciary. When sitting with a court-martial, his functions remain advisory, much as already described; he has, however, also been given an alternative jurisdiction to sit, at the request of the accused, as the sole judge in the case, determining guilt or innocence and, in the event of a finding of guilty, passing sentence.

The judge advocate

Courts of varying competence. In some countries there are grades of courts-martial with varying competence as regards persons whom they may try or punishment they may impose. In the United States, Great Britain, and Canada, general courts-martial composed of not less than five officers with a legal adviser (military judge in the United States) may deal with all persons subject to military law and pass any sentence authorized by the code; special courts-martial (United States), district courts-martial (Britain), and disciplinary courts-martial (Canada) consist of at least three officers and have limited powers. In the Soviet Union there are inferior, intermediate, and superior courts. Although under the Anglo-American system, in cases of minor importance, prosecution and defense may be conducted by regimental officers of no legal qualifications, in the majority of countries, the prosecution will normally be in the hands of a legally qualified official, known variously as commissioner, fiscal general, auditor, or military procurator.

Counsel for the accused. A soldier being dealt with summarily, or by disciplinary procedure that is not regarded as judicial action, is not usually defended—though this right has been introduced in The Netherlands, is under examination in the United Kingdom, and is not forbidden in West Germany, should the soldier so desire. In trials before military courts, all countries allow the accused to be assisted in his defense by an advocate, and in some countries this is compulsory. All countries permit the employment of qualified civilian lawyers. In Greece and the Soviet Union, the defense may be conducted by the family or friends of the accused. In the Soviet Union, additionally, such persons as representatives of syndicates and other social organizations may defend.

The stage at which a defender may operate varies. Normally, he may assist immediately after the first interrogation, when an accused is informed of his rights. He then has rights of intervention during the process of instruction and must be present at such features of it as the interrogation of the accused. In other countries (as in Greece), the defender has no part in the instruction and appears only at the trial.

Appeal. *Appeal through the courts.* Under the Anglo-American system, a court-martial's finding of guilty and its sentence must normally be confirmed by the military commander who convened the trial or by an officer superior to him. They are also subject to further review at higher levels in the military chain of command. The convicted soldier is entitled to petition the confirming officer and, subsequently, any reviewing authority against either the finding or the sentence. In some systems there may be, instead of or in addition to this right, a right of appeal from the court-martial to a superior military court. In most countries there is either an immediate or an ultimate right of appeal to a court of civilian judges—in Continental countries a Court of Cassation and in Britain a Courts-Martial Appeal Court consisting in practice of judges of the Criminal Division of the Court of Appeal. In both the United States and Britain, there can in some circumstances be a final appeal to the highest court in the land—namely, in the United States the Supreme Court and in Britain the House of Lords. In Israel, too, the right of appeal from courts-martial can extend to the Supreme Court.

Rights of the convicted soldier

In general, appeal courts are concerned only with the legality of conviction, not with matters of sentence, and the supreme courts only with points of law. Normally, only the defense can appeal, but sometimes, as in the Soviet Union, the prosecution too can appeal either against the original finding and sentence of the court-martial or on a question of law.

The ombudsman. In a few countries, representations about conditions of service and applications for advice and help outside the normal service channels may be made through specific officials. In Norway a military ombudsman was introduced in 1952. This official sometimes raises questions on disciplinary and penal offenses. The first military ombudsman was probably in Sweden, established in 1915 to take note of the sentences of military courts, conditions in military prisons, and other matters of military administration. This office as such, however, was abolished in 1968, and the supervision of the military, including complaints by soldiers, became part of the responsibilities of one of four parliamentary ombudsmen. The route of appeal by way of an ombudsman or similar civilian official in those countries that have them (including, among others, Finland and West Germany) has developed into an effective means of protecting the rights of soldiers within the military system.

Wartime procedure. Almost all countries, including those that leave the soldier in peacetime to an exclusively civilian jurisdiction, make provision for trial in time of war or emergency by military courts composed wholly or predominantly of soldiers. Sweden is an exception; even in wartime, military offenders can be tried only by the ordinary county courts. Where the normal peacetime military court is wholly or mainly composed of soldiers, provision may be made for fewer or for less-senior officers than would be called to constitute a court in peacetime. Procedure may be simplified, sometimes including an abridgment of the rights of appeal and an empowerment of the military commander to override these rights when military discipline and morale call for the immediate and exemplary execution of the penalty—including, in some cases, a death sentence.

LEGAL TRENDS

Since the 1950s and '60s there has been a trend toward the civilianization of military courts. France has joined Norway, Sweden, and West Germany in placing its servicemen in peacetime exclusively under civil jurisdiction. In countries that have not gone as far as this, the military court of first instance has come to include a substantial number—sometimes a majority—of civil judges. Neces-

The civilianization of military courts

sarily, this is possible only in countries whose armed forces are stationed entirely within (or within easy reach of) their own frontiers.

The 1950 European Convention on Human Rights, along with various rulings on its applications to military trials, have led some European countries to overhaul or amend their military judicial processes. This has had the aim of bringing them into compliance, in particular, with articles 5 and 6 of the Convention, which provide that no one may be deprived of his liberty save by a competent court and that the accused may declare his right upon a "criminal charge" to a "fair and public hearing by an independent and impartial tribunal established by law."

Both the European human rights convention and the United Nations Covenant on Civil and Political Rights expressly recognize that the right of free association may lawfully be restricted in the armed forces. Nevertheless, some countries (notably West Germany and The Netherlands) permit soldiers to form unions in order to safeguard and improve their working and economic conditions—though not to the extent of engaging in "industrial actions" such as strikes. Other countries allow servicemen to belong to unions appropriate to their particular trades.

Political activity by servicemen on behalf of a particular faction or cause would clearly be detrimental to the needs of discipline in an armed force. In general, there is an increasing tendency to allow soldiers far greater freedom during off-duty hours, so that life in barracks has come to resemble far more closely that of the civilian holding an "eight-till-five" job. There remains, however, the need for the military commander to exercise sufficient control over the private lives of his subordinates to ensure the efficiency, discipline, and good order of the force for which he is responsible. Any orders reasonably calculated to maintain these objectives are normally lawful, even though they may prohibit the soldier from some activity that may not be unlawful for a civilian. Few armed forces are, for example, prepared to retain practicing homosexuals in their ranks—not because of any moral disapproval but because the development of homosexual relationships within the closed community of a fighting force is all too likely to be subversive of discipline and effectiveness. For similar reasons—that is, the efficiency of the individual soldier and the force to which he belongs—the use of drugs is likely to be severely discountenanced.

(W.E.S./J.S.-Sm.)

Defense economics

OPPORTUNITIES FOREGONE: THE COST OF WAR

There is no such thing as an inexpensive war. First, there is the human cost in loss of life and in the physical and psychological maiming of healthy people. While the personal cost of such loss is immeasurable, the economic cost to society can be estimated. This measure was first proposed by a French economist, Jean-Baptiste Say, in 1803. He asserted the principle that war costs more than its direct expenses, for it also costs what its casualties (military and civilian) would have earned throughout their lifetimes if they had never participated in war.

Measuring social costs

Second, war has economic costs arising from the destruction of buildings, productive farmlands and forests, public services such as waterworks, electricity-generating and distribution systems, roads, bridges, harbours, and airfields, and all manner of personal and corporate property such as homes, possessions, factories, machinery, vehicles, and aircraft. War, therefore, destroys physical capital that has been created by previous economic activity.

Reconstruction after war is a particular economic burden because the finance, imported capital goods, and labour used in reconstruction merely restore the losses a country has sustained, rather than adding to the stock of capital available to its economy. Thus, even if it manages to restore all its physical losses, it uses scarce resources that would otherwise have been available for extending and improving economic activity. As most wars since 1945 have occurred in the Third World, some of the world's poorest countries have suffered the most from the economic losses of war.

War also costs a great deal in goods and services to create the weapons of war and to supply the people engaged in the war effort. The diversion of these goods and services—which range from the metals and chemicals transformed into weapons to the food, clothing, and shelter for the armed forces—reduces current civilian consumption, which lowers the population's living standards. Metal used to make a tank cannot be used to build bridges, fuel used to transport military supplies cannot be used on school buses, cement used to construct ammunition dumps cannot be used in house construction. This constitutes the opportunity cost of war—that is, the extent to which the economy foregoes the opportunity to commit these resources to alternative peaceful uses.

The opportunity cost of war is also felt in the future. In addition to allocating resources to consumption (the satisfaction of current needs), an economy allocates resources to investment (the new factories and machinery that produce tomorrow's goods and services). Resources diverted to war cannot be used to create new productive capacity for future consumption, and this reduces the living standards of the population below what they otherwise would have been in the future.

The total costs

In summary, the total costs of war include the cost of the foregone use of the economic resources used up in the conflict. These include the cost of the foregone lifetime earnings of those killed in the war, the cost of lifetime medical care for those permanently incapacitated by the war, the cost of replacing the physical capital destroyed or damaged by the war, the cost of supplying the armed forces with the weapons of war, the cost of sustaining the armed forces and those in support functions (including their pay and pensions), and the losses to the economy caused by the diversion of resources from peaceful investment in future economic capacity.

DEFENSE EXPENDITURE: THE COST OF DETERRENCE

As war is expensive, countries aim to avoid its costs and remain independent within sovereign borders. In the absence of a universally binding and verifiable agreement to abolish war, the best option is to deter those countries prone, by their history or by the policies of their governments, to resolve disputes by resorting to war. Deterrence has two aspects. First, by allocating resources for a minimum level of military capability, a nation ensures that it can resist an attack by a potential aggressor and severely damage the aggressor's economy and territory. In this way the costs to the aggressor of initiating a war will far exceed any likely gains. Second, by making credible its willingness to use military force, should it prove necessary to do so, the nation aims to leave potential aggressors in no doubt of the consequences they will suffer if they are tempted to launch an attack.

Deterrence, while expensive, is incomparably less expensive than war. The study of its expense constitutes the subject matter of defense economics.

Measuring the burden. Adam Smith, the founder of economics as a discipline in the social sciences, was the first economist to theorize about the economics of war. In his major work, *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776), Smith considered a perennial problem of defense management, namely, the increasing expense of war-fighting equipment. He noted that changing technology raised the costs of war—for example, that the musket was a more expensive item to acquire than its predecessor, the javelin. (In the same way, a modern jet fighter is much more expensive than its propeller-driven predecessor.)

The d/GDP ratio

The rising cost of weapon technology does not mean that defense costs (d) necessarily rise as a proportion of gross domestic product (GDP; the sum of all expenditures made in one year). The d/GDP ratio is a measure of the military burden, and evidence suggests that this burden has not risen through time (in high-income economies it has been falling for most of the post-World War II decades). Although the unit costs of specific weapons rise as technology adds to their capabilities, high-cost solutions to one form of a military threat (for example, the use of expensive tanks to defend against a massed tank attack)

usually become vulnerable to low-cost alternatives (such as the relatively cheap antitank missile and precision-guided munitions), which either alter the nature of the threat or make redundant the high-cost solution.

In a developed economy, the annual costs of defense procurement and logistics typically take up more than half of the defense budget, the rest being spent on personnel. In the underdeveloped economies, the balance is reversed: most of the annual costs (70–90 percent) are spent on personnel, with the remainder spent on procurement and logistics. This difference reflects the gap in available war-fighting technology between the developed and the underdeveloped worlds. The bulk of the world's defense spending is accounted for by the high-income economies (the United States, Europe, and the Soviet Union), primarily because of the cost of high-technology weapon systems. Yet most wars are fought in low-income countries between relatively poorly equipped armed forces. Moreover, the inability of low-income countries to maintain sophisticated weapons to the operational standards of their manufacturers fully explains the many logistical problems the armed forces of poor countries have faced in their wars. Importing sophisticated weapon systems does not guarantee a sophisticated defense capability if the support system (fuel, spares, ammunition, repairs, and overhaul procedures) is either less than satisfactory or less than adequately funded. Defense capability is inseparably linked to the cost of maintenance.

Defense is a public good; that is, once deterrence is achieved, all citizens benefit from the avoidance of war and no citizen can be excluded from enjoying the benefits. People who could not be excluded from a public benefit would, if given the choice, rationally choose not to contribute toward its cost. In other words, they could "free ride" on the contributions of others. For this reason, defense in all countries is paid for by taxation, a burden that is borne by all citizens, and in all countries the military force considered necessary for deterrence is under the direct and exclusive control of the government.

Comparing burdens. *Settling on a standard.* International comparisons of how governments arrange their defense spending are fraught with conceptual discrepancies. The defense burden of a country is measured by the d/GDP ratio, which indicates how much of the nation's resources are being allocated to defense each year, but different estimates of both d and GDP are possible, each giving a different d/GDP ratio. Capitalist economies, which use the GDP, measure economic activity differently from communist economies, which use a net material product (NMP) system. The NMP excludes many expenditures, including state administration and defense, normally included under GDP. This complicates comparisons between these systems.

Capitalist and communist measuring systems

Defense expenditures themselves are subject to controversy. The North Atlantic Treaty Organization (NATO) has agreed on a measure of defense activity to which it adheres when making comparisons of its members' defense burdens, but other countries follow different conventions. Some, largely low-income countries, exclude internal security expenditures, which can be relatively high, thus lowering their official d/GDP ratio. Others, such as the Soviet Union, exclude defense-related research and development, frontier guards, and paramilitary reserves, thereby reducing the nominal defense expenditure by up to 30 percent.

Even if agreement could be reached on what constitutes defense expenditure, this would still leave countries with a measure denominated in their domestic currencies. For meaningful comparisons of the absolute amounts spent on defense, every country's defense expenditures would have to be reduced to a common currency. But the act of converting each currency into, for example, U.S. dollars could lead to distortions, because official exchange rates reflect official policies and not existing realities. Thus, two countries with similar amounts in dollars spent on defense, and therefore in balance in their defense capabilities, could face a growing imbalance in their dollar-based defense expenditures purely because one of their currencies has changed its exchange rate with the U.S. dollar.

Comparisons of the absolute amounts each country spends on defense are prone to error and must always be used with caution. Nevertheless, because each country measures its defense spending and its GDP in its own currency, the d/GDP ratio is an acceptable measure of a country's defense burden. Ratios can be compared across countries and in different time periods. The d/GDP ratio rises rapidly during a major war—in Britain in 1944 the d/GDP ratio reached 60 percent—and it falls in periods of prolonged peace. A country raising its d/GDP ratio signals that it is concerned with security, in turn causing concern among countries likely to be affected.

Defense burdens worldwide. Security expenditures for both external defense and internal law and order account for major shares of government expenditures. In many low-income countries, these expenditures often exceed 20–30 percent of the state budget and more than 10 percent of the country's GDP. The higher-income countries, while spending higher absolute amounts on defense, tend to spend smaller proportions of state expenditure (under 15 percent) and smaller proportions of GDP (under 5 percent). Given the perilous security situation in the lower-income regions of the world, these discrepancies are understandable (if also regrettable in view of their other pressing needs).

Defense spending in rich and poor countries

By looking at the actual defense burden of an individual country and comparing it with the norm for similar economies, analysts can infer localized circumstances that may be influencing the government's perceptions of security. For instance, a poor country with a very low d/GDP, or a rich country with a very high d/GDP, is behaving differently from the norm for similar economies. An economist would seek explanations for this in the perception of a threat indicated by the public statements of the government concerned. In the absence of such statements, or where public statements are contrary to the behaviour of the government (for example, if it is raising its d/GDP but publicly proclaiming its peaceful intentions and denying that it is threatened by, or is threatening, any other country), the economist would rely on the d/GDP as an indicator of true intentions. It is likely that the intelligence services of neighbouring countries would draw similar conclusions from the economic data.

Within the higher-income countries there are notable differences in the amounts spent on defense. The United States and Britain have spent relatively high proportions (5 to 10 percent) of their GDP on defense since 1955, compared with Japan, which has spent less than 1 percent of GDP over the same period. Germany and France also have tended to spend a smaller proportion of their GDPs than Britain on defense (though the absolute amounts have been similar, since they have larger GDPs than Britain).

The Japanese case is interesting because of the differences in economic achievement between Japan and the big defense spenders. Many economists believe that there is a connection between the amount of GDP spent on defense and a country's economic growth, investment, and living standards. Japan's limit of d/GDP to under 1 percent was a result of its defeat in World War II. There is no doubt that it benefited enormously from limited defense spending (particularly while it could free ride under the military protection of the United States), since resources not allocated to defense went into economic investment, to the direct benefit of civilian employment and output. However, at the same time Japan also spent much less (about half as much) of its GDP on general government expenditure (g) than the United States and western Europe. Whether it was the low d/GDP (1 percent) or the low g/GDP (9 percent) that allowed the resources for Japanese economic successes to be mobilized is arguable. A low g/GDP ratio implies a lower level of taxation than a high g/GDP. This releases a higher flow of savings into the economy, enabling higher investment ratios to be maintained. High rates of investment, which are associated with higher growth rates of GDP, characterized the Japanese post-war economy and are a more likely explanation for the Japanese economic success than its low d/GDP.

Effects of defense spending on economic growth

The Soviet Union has long spent a high proportion of its national resources on defense. Estimates vary, but the consensus among Western economists is that Soviet

d/GDP for much of the postwar period was around 14 percent. (The official Soviet d/GDP was 6 percent.) If defense spending competes with economic growth in the capitalist economies, contributing to inflation, low investment, and lower living standards, then it must have a devastating impact on poorer economies such as the Soviet Union. The need to compete with the United States at all technological levels across the weapons spectrum has been met at the cost of heavy distortions in the rest of the Soviet economy. This has compelled the Soviet Union to review its priorities and to consider whether its security is best assured by continually raising the military ante with the West or by living at some lower level of military tension with a reduced offensive military capability.

DEFENSE MANAGEMENT: BUDGETING DETERRENCE

"How much defense is enough defense?" is the great unanswerable question of defense economics. Those charged with preparing a defense capability tend to be more cautious about the level of capability than those who eventually have to pay for it. In fact, the very success of deterrence—a high probability of nonattack throughout a long period of peace—tends to reduce the amount of defense spending that the electorate considers necessary to achieve deterrence. Judging the appropriate level of military preparedness is not a science; it is a mixture of intelligent response to credible threats and of judicious, cautious preparation "just in case" this or that should arise. The managers of the armed forces tend to increase the contingencies they wish to prepare for, while skeptical taxpayers tend to question whether certain preparations are absolutely essential. In democracies this tension forms the permanent agenda of the defense debate.

Tension between managers and taxpayers

Stocks and flows. Defense expenditures are made on an annual basis, the government allocating so much of its total budget to personnel costs, so much to the procurement of weapon systems, and so much to general support. The pay and allowances of defense personnel are consumed within the year; that is, they spend their wages, allowances, and pensions on consumer goods and, in so doing, add to total demand in the economy. Procurement, on the other hand, is somewhat different. A tank lasts much longer than the single year in which it is purchased. Because it is supposed to last as long as it takes to become obsolescent, the tank becomes part of the country's permanent defense capability. That defense capability is, in economic terms, a stock, while the annual expenditure is called a flow.

Even if, for some reason, a defense budget is reduced in a single year, a country's defense capability need not be reduced. The government can still draw on the stock paid for by previous defense budgets, which is manifested in its tanks, aircraft, ships, communications systems, trained personnel, and expertise in military affairs. Clearly, if the defense budget continues to be reduced every year, there will come a point at which the country's defense capability will decline through attrition as items of equipment become obsolete or beyond repair.

The analogy is with a bath that is filling with water while the plughole is open. As water pours into the bath, water also drains from the plughole. It is the difference between the rates at which water flows in and out that determines whether the bath fills or empties. If the flows in and out are equal, the water level will remain constant. Likewise with defense capability: if the additions (flow in) to the stock of weapons matches the attrition (flow out) of the stock from all causes, then the country's defense capability will remain constant.

Measuring a threat: the example of NATO. Budgeting a nation's defense capability is complicated, however, because defense capability is not determined unilaterally; it depends on the capability of the potential aggressor. The gap in military capability between any two countries is known as the threat, and estimates of the threat constitute the major input into defense planning.

If a potential aggressor develops an advanced weapon system that effectively counters a weapon stocked by the defender, it will eventually threaten to overwhelm the latter's defenses. Likewise, a growth in the stock of weapons deployed by a potential aggressor can eventually have a similar effect in threatening to overwhelm the defender's smaller stocks. If the defending country does not invest in overcoming each new threat to its capability—by technology, new types of weapons, increasing the stock of current weapons, or all three options simultaneously—it will risk a reduction in the probability of nonattack—that is, its deterrence capability will be compromised.

Failing to invest in defense

The economics of nuclear deterrence. Estimates of the threat of a Soviet invasion across the German border determined the nature of NATO's response for more than 40 years. While NATO planners considered their own forces to be technologically superior to the Soviet forces, they were nevertheless mindful that the Soviet Union had a decisive quantitative superiority in conventional forces (more tanks, armoured vehicles, artillery, combat aircraft, and troops). The threat of a land-based invasion by Soviet forces, which the planners considered to be virtually unstoppable, led directly to the decision to deploy nuclear weapons as the ultimate deterrent against an invasion of western Europe.

Nobody could survive a major nuclear war in Europe. The damage to the Soviet Union from an American nuclear strike would be matched only by the damage to the United States from a Soviet nuclear strike. Because each country has maintained sufficient nuclear forces to respond in kind to a first strike by the other, a nuclear exchange would be suicidal for both. Whatever the rhetoric, therefore, both countries have a strong interest in preventing war of any kind from breaking out on the continent of Europe. Literally, they are hostage to each other's behaviour, making Europe an unsafe place to start a war. This doctrine, known as "mutual assured destruction," was given the appropriate acronym MAD.

The consequences of MAD led NATO to adopt a policy known as "flexible response." Rather than an all-or-nothing nuclear exchange, this envisaged a staged escalation of NATO's response to a Soviet invasion, based on containing the initial thrust of the Soviet forces and warning them of the consequences of further encroachment on NATO's territory. To underline the credibility of the threat of nuclear retaliation, NATO commanders were issued battlefield nuclear weapons, which NATO governments might or might not release for immediate use, with or without warning. Uncertainty about NATO's policy of probable first use of nuclear weapons was regarded as sufficient to make Europe an unsafe place for the Soviet Union to risk the consequences of a conventional war. As long as the risk of the horrendous consequences of a nuclear war exceeded the prospects of potential gain from launching an attack, the probability of nonattack on western Europe by the Soviet Union remained at an acceptable level.

The economics of conventional deterrence. The possession of nuclear weapons by some NATO countries (the United States, Britain, and France) did not obviate the need for expenditure on conventional armed forces. To abandon conventional forces would risk having to use nuclear weapons as soon as the first Soviet forces crossed the German border or some naval incident occurred in any part of the world. This escalation from a small incident to the end of the world in one short step was unacceptable; hence, NATO countries invested resources in conventional capabilities in addition to nuclear weapons. These conventional forces aim to blunt a Soviet attack and give time for political processes to influence the Soviet government's decisions.

Matching conventional forces to Soviet conventional capabilities had to take note of two facts: First, the Soviet Union had overwhelming superiority in conventional forces. Military doctrine holds that concentrating superior force at a single point can overwhelm the defense, and the Soviet Union had the capability to achieve such a strategic advantage at a time and place of its choosing. Second, while NATO had advantages in military technology, there was a constant effort by the Soviet Union to close the technological gap. Also, there is some point at which a quantitative advantage acquires a qualitative dimension, and this advantage cannot be neutralized solely by relying on a technological gap between the weapon systems themselves.

Closing the technological and quantitative gap

Thus the paradox of NATO defense spending. The alliance was constantly trying to widen the technological gap to compensate for its disadvantage in numbers, while at the same time it was required to maintain large quantities of its existing systems to redress the ever-widening gap in numbers that the Soviet Union was believed to be creating across the German border. Whether to develop ever-new weapon systems to combat a closing of the technological gap by the Soviet Union as well as the sheer numbers of Soviet systems, or to concentrate on supplying the armed forces with duplicate copies of existing designs, has long been NATO's quandary.

Choosing weapon systems. Since the 1960s there have been several attempts to impose some rationality upon defense planning. The complexity of weapons development has few parallels in civilian development. Working close to technological frontiers (sometimes having to think beyond them) under management systems imbued with a public-sector rather than a commercial ethos, under government budgetary constraints and shifting political priorities, and subject to ever-changing estimates of the threat the system is designed to counter, has produced an expensive and time-consuming procurement system. Lead times of 18 to 25 years from initial concept to in-service production are not unknown in defense procurement.

The need for more rational choices in weapon programs and in the deployment of scarce resources increased as the defense budgets grew in absolute size (though falling as a proportion of a growing GDP). The cost of errors in choice increases as the cost of a single weapon platform escalates, so that, with a new weapon system costing $10–40 billion, it is crucial not to find that it is not needed by the time it is in service, that the technology cannot be made to work, or that it has been made obsolete by new developments.

The Polaris submarine-launched ballistic missile program, begun by the United States in 1956, was the first highly complex system that required new management techniques to be brought to successful completion. One technique—called program evaluation and review technique (PERT)—found civilian applications after it was invented by the U.S. Navy to build Polaris on time and under budget. Similarly, earlier techniques such as cost–benefit analysis (invented to cope with submarine hunting problems during World War II) and input–output analysis (a technique developed by the U.S. Air Force for identifying the critical parts of an economy to develop or damage) rapidly spread into civilian use and into most academic management programs.

Rational economic choice of military options

The first attempts to bring rational choice to the management of a defense budget coincided with the U.S. involvement in Vietnam. Terms such as systems analysis, as well as planning, programming, and budgeting systems (PPBS) and functional costing, became common in defense management. Much of the intellectual capital invested in these techniques came from economists, whose discipline in costing options and marginal analysis provided them (if not always the defense managers they advised) a set of tools appropriate to the task.

When decisions are made solely by the lobbying of special interests—such as the navy, air force, and army—the result is likely to be a constant compromise under which programs remain in the budget because of political considerations. Defense analysts attempt to force the military lobbyists to set specific objectives for their programs and to accept criteria by which the military value of the programs can be judged. Like many management fashions, PPBS and its associated techniques did not survive in their earliest forms, but they did establish the belief that analyzing, evaluating, and choosing rationally was superior to lobbying by bureaucrats and, sometimes, by corrupting commercial interests.

This can be illustrated by the technique of functional costing. Ordinarily, most budgets are a listing of expenditures under various main headings—personnel, equipment, and supplies—and the total is approved through the political process. This type of budget is called an accountability budget because it accounts for defense expenditure, but it cannot inform the defense planner (or the taxpayer) how efficiently the defense department has spent the budget. Under functional costing, the objectives of a proposed military program are shown along with the costs of all the resources needed to fulfill each objective, irrespective of which armed service contributes to the activity.

For example, under functional costing there is no navy budget that costs everything spent by the navy. There is instead a maritime defense budget, a deep-sea navy budget, a coastal defense budget, and so on. These budgets may include costed contributions from units of the navy, air force, and army, plus an assessment of the costs of support functions used to carry out the activities. If, for instance, it is proposed to add a longer-range aircraft to the maritime defense role, this can be costed, and, depending upon the importance of extending maritime defense compared to other objectives, an informed decision can be made on whether to allocate the incremental funds to the upgraded aircraft or to some other project supporting some other military activity. Deciding between marginal increases or decreases in expenditures across different functions, all of them linked to specified objectives, is an improvement over buying aircraft simply because it is the turn of the air force to get a big project approved by the government.

Conscript or volunteer. Modern armed forces use either a voluntary recruitment scheme or a form of conscription to supply the people needed to staff the military. Each scheme has economic consequences.

Conscription involves a period of compulsory military service for all eligible males, usually triggered by their date of birth. (In some countries, such as Israel, females are also required to undergo military service, though usually in a support rather than a combat role.) Conscription provides a pool of recruits at a low cost per head. The conscripts receive extremely low wages, well below what they would earn as civilians. This difference in earnings is a direct monetary loss to them and a loss to society, which loses the output they would produce if they remained civilians. Conscription offers a net saving only to the defense budget, although what is saved in personnel costs is largely spent in increased training costs. Conscript armies require much larger training programs than volunteer armies because the service life of a conscript (two or three years) is shorter than a volunteer's term of engagement (three to 15 years). Each new age group of conscripts has to be trained, diverting full-time soldiers from other duties as well as adding to overall costs.

Wages earned versus opportunities lost

Volunteer armies cost more per head because their wages must be comparable in some degree to civilian wages. While a national emergency can induce people to volunteer, a peacetime recruit is influenced by the alternative incomes that can be earned as a civilian. Some people volunteer whatever the wages, and some volunteer because they are unemployed as civilians, but most evidence indicates that volunteer rates will fall if military wages fall too far below civilian wages. This is particularly true for volunteer officers, who take with them critical skills when they leave the armed forces for well-paid jobs as civilians.

While volunteer armies cost more per head, they can cost less in total because they do not have to be as large as conscript armies. Although conscription is common across the three branches of the armed services, the proportion of regular volunteers to conscripts is smaller in the army and larger by far in the navy and air force. Ships and aircraft require more-skilled and better-educated personnel than infantry divisions, and the navy and air force in most countries tend to use conscripts only in less-skilled roles, reserving the command roles (pilots, captains, engineers, navigators) for volunteers. This pattern can be seen in many Latin-American military forces. In Israel, where conscription covers practically the entire population, the problem of retaining skilled recruits is met by extending the periods of military service through the civilian lifetimes of the recruits.

All personnel policies are vulnerable to demography. The proportion of a nation's population that is made up of young people eligible by fitness and intelligence for military service sets a limit on how many can be conscripted or induced to volunteer. Conscript armies, which are cost-effective when there is a large pool of young people from

which to choose, are particularly threatened by demographic changes that reduce the pool of potential recruits. As the birth rate appears to fall in higher-income economies over time, the prospect for mass conscript armies looks bleak. Switching to an all-volunteer force is a short-term alternative, although the decline in the recruitable age group will force up military wages as the armed forces compete with civilian employers for the same age group. Substituting technology for labour is another short-term solution. But it too has limitations, not the least of them the problem of recruiting from a shrinking age group a sufficiently educated and skilled labour pool to operate sophisticated military equipment.

The economics of conscription in rich countries

WAR FINANCE: WHEN DETERRENCE FAILS

War is too serious, and too expensive, to be left to the whim of chance, yet throughout history many governments have been willing to engage in war if it suited their interests as they perceived them, and many have also been dragged into wars when cooler calculations might have encouraged them to remain at peace. It was out of the need to raise the finance to conduct wars that the earliest systems for collecting public finance developed.

Taxation. The practice of taxing the population to pay for war has a long history. In early nomadic societies, wars could be fought with little expense other than time and casualties. Nomadic horsemen engaged in war as an extension of their normal activities as herdsmen. If successful, the warriors plundered the defeated, who were either killed, sold, or scattered. With more-settled agricultural societies, wars could be fought between planting and harvest, the armies living off of the land or the pay of the king. In the feudal system each man had an obligation to fight if required by the lord of the manor, to whom the vassal owed his livelihood. Weapons were often the personal possessions of the warriors and were fashioned by themselves, their forebears, or craftsmen. As weapons improved in quality and ingenuity, special efforts to produce them had to be made—for which their producers were paid out of public funds, as were the soldiers to whom they were distributed. These expenses necessitated the collection of special funds, and thus the government turned to levies on the population to provide the resources.

Where taxation alone was not sufficient to pay for a war, the king could resort to selling relief from feudal obligations (usually to prosperous cities), and to borrowing from rich individuals (who risked confiscation if they refused and not being paid back if they agreed). Most political crises in European history up to the 19th century arose from disputes over public finance, usually to pay for wars. The English Civil Wars of 1642–51 were a typical example of a bitter dispute between king and parliament over which had the powers of taxation.

By the time of the French Revolution in 1789, warfare had ceased to be a localized affair with a few thousand soldiers engaged in a single decisive battle. Warfare had become a massive undertaking, often lasting months and years and involving armies of hundreds of thousands. The expense of war reached levels unheard of previously. The Royal Navy, for instance, employed 200,000 men in the 1790s alone. Out of the need to find new sources of revenue for the long war between Britain and France from 1793 to 1815, the British government introduced a temporary income tax to be levied on all persons earning above a high minimum income. The advantages of the system soon became evident, and income taxes were widely adopted as permanent sources of government revenue.

The income tax

Borrowing. The British government also discovered another source of revenue, the national debt. Whereas previous borrowings by monarchs were a great risk to the lender, under the national debt scheme the government agreed to guarantee regular payment of interest to all persons who lent to it, either in perpetuity or for a fixed term. Holders of government bonds were also permitted to sell them, passing the right to the guaranteed income to the buyer. Again, the system was so successful that it was soon copied by other governments, not all of them as scrupulous in their repayments as they promised. War bonds featured strongly in the two world wars, and it was regarded as patriotic to use personal savings to purchase them—though most of the borrowing came from institutions.

The new sources of war finance enabled the increased expense of war to be met—and, in the opinion of some economists, made it more likely that a government would embark on a war for less than good reason. If war had to be financed out of the current consumption of the population, natural limits would be set on war being undertaken lightly, but borrowing removed these limits. Though the population would be saddled with interest payments for decades to come, this cost was small and of long duration compared to the immediate cut in living standards that would have to be made to pay for a war in full.

War's consequences: inflation and recession. A major modern war can divert from 40 to 60 percent of a country's GDP, and one object of war finance is to release within the economy the resources needed for the war effort without causing inflation. If the government merely prints money to pay for the resources it requires, it will bid up prices in competition with civilians. The alternative is to reduce civilian consumption by imposing taxes at levels sufficient to force consumers to forego bidding for goods and services. The income from taxes can then be applied by the government to bid for the resources released by its program.

Taxation acts both to raise necessary finance and simultaneously to reduce aggregate demand, which releases the resources needed for the war effort. The war effort represents a substantial expansion of production, for which producers receive wages and profits. When these same producers attempt to spend their incomes, they face a diminished quantity of civilian goods available for purchase. They must either face rapidly rising prices (caused by excess money chasing insufficient goods), or they must restrain—or be restrained—from spending. A war economy therefore imposes higher taxes on wages and profits to reduce demand. War bonds and taxes provide finance for the war effort and reduce demand for civilian goods and services. To conduct a major war without such an austerity program risks inflation.

Reducing demand in wartime

If inflation is a risk during a war, recession is another risk at the end of it. The massive expansion in production to provide resources for the war effort, if suddenly contracted by the cancellation of all defense contracts, throws large numbers of people out of work. The unemployed reduce their consumer spending, causing further cuts in aggregate demand, which throws yet more people out of work.

World War I was followed by recession. Because much of the war damage along the Western Front was confined to the vast, static battlefields across France, where the destruction was mainly human and the cost was mainly in war materials, there was no need for a massive reconstruction program. Also, the damage on the Eastern Front was swept behind the newly formed Soviet state, which for ideological reasons eschewed capitalist reconstruction. After the war factories closed down, removing a flow of wages and profits into the economy, the demobilization of troops put surplus labour into an economy that was already in recession.

Recession was averted at the end of World War II by reconstruction of the cities and economies of western Europe and Japan. Reconstruction rapidly transformed the war economies into mass consumer economies supported by the pent-up demand that had been frustrated by the lack of civilian goods and by high taxes during the war. In Europe's case there was a transfer of capital from the United States through the Marshall Plan. As the war economies were dismantled, economic growth surged, and those countries that did best economically were those that dismantled their highly regulated, government-controlled war economies quickest (West Germany, for example). Those countries that were least successful in dismantling their regulated economies were the slowest to recover. Among these were Britain, which increased state intervention from 1946 to 1951, and the Soviet-controlled economies of eastern Europe, which went even further down the road of government-managed economies.

(G.Ke.)

THE CONDUCT OF WAR

Strategy

FUNDAMENTALS

Strategy, narrowly defined, means "the art of the general" (from the Greek *stratēgos*). In a strictly military sense, the term first gained currency at the end of the 18th century, when warfare was still relatively simple and limited. In its military aspect, the term had to do with stratagems by which a general sought to deceive an enemy, with plans he made for a campaign, and with the way he moved and disposed his forces in war. Often defined as the art of projecting and directing campaigns, military strategy came to preempt almost the whole field of generalship, short of the battlefield itself. It also came to include the planning of naval warfare. To tactics military jargon reserved the art of executing plans and handling forces in battle.

The expanding definition of strategy

The term strategy has expanded far beyond its original military meaning. As society and warfare have steadily grown more complex, military factors have become more and more inseparable from the nonmilitary in the conduct of war and in programs designed to secure peace. Nations have found it necessary to adjust and correlate political, economic, technological, and psychological factors, along with military elements, in the management of their national policies. The demarcation between strategy as a purely military phenomenon and national strategy of the broader variety became blurred in the 19th century, particularly in wartime. The distinction became even less clear in the 20th century when nations became more interdependent and the line between war and peace less clearly definable. As a result, the appearance of the term grand strategy (or higher strategy), meaning the art of employing all the resources of a nation or coalition of nations to achieve the objects of war (and peace), steadily became more popular in the literature of warfare and statecraft of the 20th century.

This broadened scope of strategy has tended to blur distinctions customarily drawn by earlier writers between strategy and statesmanship and between garden varieties and higher, or "grand," forms of strategy. Though there is still no agreed definition of the precise meaning of the term strategy, few students of the subject any longer accept the earlier narrow definition. Also, few contest that strategy, whether in its narrow or broad sense, will, by the very nature of its shifting bases, continue to be a changing art.

The search for principles. The starting point of all strategic planning and action is national policy. Once the national aims are set forth by the leaders of the state, the commander sets about drawing up his plans. He must take many matters into account; for example, factors of space and time, the state of his own forces, the enemy's capabilities and intentions, and reactions at home and abroad to his projected moves. The strategist deals in many uncertainties and imponderables. Indeed, the art of the strategist is the art of the "calculated risk."

The debate over the principles of strategy

The growing complexity of modern warfare has led some students to take a fresh look at the principles that have traditionally guided military strategists in war. It has long been a favourite occupation of military theorists to seek to distill from the great mass of military experience simple but all-pervasive truths—lists of principles—to guide commanders. Usually they have derived such principles from a study of campaigns of the great captains of history; occasionally outstanding practitioners have set them down on the basis of personal experience. As far back as 400 BC Sun-tzu, a Chinese general, set forth 13 principles. The axioms range from American Civil War General Nathan Bedford Forrest's simple admonition about getting there first with the most men to Napoleon's 115 maxims. The stress varies from list to list. For example, the followers and interpreters of the 19th-century theorist Carl von Clausewitz believed that the battle was all and that defeat of the enemy's armed forces was the correct objective and path to victory. Exponents of "the strategy of indirect approach," on the other hand, sought victory by indirect methods.

Though there is no complete agreement on the number of principles, most lists include the following: the objective, the offensive, cooperation (unity of command), mass (concentration), economy of force, maneuver, surprise, security, and simplicity. The British have added one called "administration"; the Soviets, another, translated as "annihilation." Despite debate over their precise number and meaning, the principles of war are widely taught, and most military students accept them as basic concepts.

The permanence or impermanence of strategic principles

The individual authors of the lists have almost uniformly claimed the principles to be immutable. They have argued that success in military strategy in the past has been the result of adhering to them and that the advantages of the offensive, the concentration of force, the effort to achieve surprise, the proper movement of forces, and their security from attack, sabotage, or subversion are in the province of modern as well as ancient warfare. Some authorities have even argued that since war is not the concern of soldiers only, the "principles" deserve a wider application throughout government—in grand as well as military strategy.

Other authorities have argued that the claim of immutability cannot be accepted literally, that there is little agreement as to what the principles are and mean, that they overlap, that they are fluid and require constant reexamination, that they are not comparable with scientific laws since no two military situations are ever completely alike, that the so-called principles are not really principles at all but merely methods and commonsense procedures adopted by great commanders of the past, and that changes in the conditions of war alter their relative importance.

The debate over principles was renewed with the coming of the nuclear era. Some theorists argued that the new weapons had destroyed whatever value the principles once had; others contended that the principles were as valid as ever, even more so. To some extent this was a debate over semantics. Defenders pointed out that each age must make its own applications of the "fundamental truths" of strategy. Opponents argued that there can be no set rules for the art; the so-called principles must by no means be interpreted as pat formulas for victory to be followed blindly and rigidly; the only sound guide in war and strategy is flexibility.

Relation between strategy and tactics. In the theory of warfare, strategy and tactics have generally been put into separate categories. The two fields have traditionally been defined in terms of different dimensions: strategy dealing with wide spaces, long periods of time, and large movements of forces, tactics dealing with the opposite. Strategy is usually understood to be the prelude to the battlefield, and tactics the action on the battlefield itself. As a result, much of the literature and theory of strategy has in the past been preoccupied with the proper approach to the battlefield, the leading of troops up to the time of contact with the enemy. This situation explains the attention to strategic maneuver—aimed at putting one's army into the most favourable position to engage the enemy and compelling the enemy to engage at a disadvantage and depriving him of freedom of movement. Indeed, early writers on strategy dealt heavily in the so-called "geometrical strategy"—the angles formed by lines of movement and supply of opposing armies.

Strategy as the prelude to battle

Despite distinctions in theory, strategy and tactics cannot always be separated in practice. In fact, the language of strategic maneuver (for example, "envelopment," "penetration," "encirclement") is also largely the language of tactics. Movement begets action, and action results in new movement. The one merges into the other. Strategy gives tactics its mission and wherewithal and seeks to reap the results. But tactics has also become an important conditioning factor of strategy, and as it changes, so does strategy. Battles and fronts are no longer necessarily restricted in space and time, and the distinction between battles and campaigns is no longer so clear-cut, as the tridimensional warfare of World Wars I and II demonstrated. Indeed, in World War II theatre commanders were as much con-

cerned with the actual fighting of armed forces in battle as they were with larger strategic decisions such as relations to allies, economic problems, and political questions on the ground. Although in theory strategy continues to occupy a middle ground between national policy and tactics, in practice the line dividing it from the other two fields has become difficult to draw.

Strategic leadership and war planning. Count Alfred von Schlieffen, the famous German military leader of the period just before World War I, once said: "A man is born, and not made, a strategist." But it is obvious that even a born strategist—if there be such a natural genius—has much to learn. In the past strategic leadership was a relatively simple affair. J.F.C. Fuller, the British student of warfare, pointed out in *The Foundations of the Science of War* (1926) that until relatively recent times the death, capture, or wounding of either of two opposing generals normally decided a conflict, "for the general *was* the plan." He could personally devise the plans and direct his troops. By the mid-20th century this was rarely possible. As warfare has become complicated, strategic leadership has become more difficult. The art has taken on many more facets, and systematic training is required to master them. The strategist has retired from the scene of battle, and large, specialized staffs have grown up to help him.

"Corporate" leadership in modern strategy

Although the responsibility for strategy remains the general's, many of his functions have been delegated to his planning staff. In modern states corporate leadership has become the rule in the management of military strategy, as in the direction of large business enterprises.

The example of an Alexander the Great completing his advance planning and leaping into battle at the head of his troops would in modern warfare be considered most unusual. Napoleon was wont to make his plans and then retire with his retinue of trusted advisers to survey the battlefield on horseback from the top of a hill. Generals in World War I were often pictured in their offices in large headquarters—usually in a château behind the lines, studying a map on the desk and dispatching orders via the telephone and motorcar at hand. In World War II the headquarters staffs of commanders in the theatres of war grew even larger and more elaborate. Tridimensional warfare—land, sea, and air—had enlarged the field of operations far beyond individual battlefields, and usually a high commander reached his decisions in a headquarters far removed from the field of battle and months before the battle itself took place. Far from striking the classic pose of the officer on a well-schooled charger, some of the greatest generals issued their orders from desks and fought their most important battles at conference tables. As strategic planning became a highly organized affair, planning committees and conferences in the capital cities of the warring powers made the blueprints for victory in the global, coalition struggle. In their capital command posts, military leaders kept in touch with the manifold phases of the national government's war effort and dealt with the worldwide problems transcending those of the individual theatres of war. With the aid of new devices for rapid communication, these leaders and their staffs sought to set the patterns of strategy and keep abreast of the movement of armies as the Caesars and Napoleons had done in earlier eras.

Strategic planning as a peacetime function

As war became more total, war planning became a significant peacetime function of governments. The manufacture of strategic plans has become a highly specialized industry in modern military establishments. At the same time, more and more governmental agencies have been drawn into the business of planning for national security. The plans they produce may vary from a simple design to shift a small task force to a danger spot to an elaborate plan for the conduct of war in its entirety. To be realistic, strategic plans and estimates must constantly be reexamined and brought into harmony.

Against this general background in the nature of the art, it is now possible to sketch the important contributions made in key periods to modern strategic theory and practice. It is important to remember that the art of strategy has changed from age to age, just as has war itself, and that each is the product of its own society and time.

HISTORICAL DEVELOPMENT THROUGH WORLD WAR II

Though the serious and systematic study of modern strategy may be dated from the 18th century, various authorities have identified strategic precedents going back to earliest times. Students of warfare of primitive ages have associated with primitive tribes and clans a stratagem of surprise from darkness or by ambush, and they have identified a strategy of hunt and pounce, like that of a lion or tiger. The Bible points to the care with which Moses prepared his operations—an early form of advance planning. The ancient world developed a strategy of mass attack by phalanx, legion, or cavalry. Alexander, Hannibal, and Caesar, who combined in their own persons political and military direction of the state, planned their famous campaigns far ahead. They have been singled out as forerunners of the modern art of grand strategy. Writers in modern times have used the campaigns of these great captains to illustrate practically every known "principle of war." But important as their attention to strategic considerations in war and especially to strategic approaches to the battlefield may have been, the foundations of the ancient art of warfare were tactics and battles. To a considerable extent battles—often short and furious—also held the centre of the military stage in the European Middle Ages. Strategy was notably absent in the excursions of the Huns, the Muslims, and the crusaders. Far more important from a strategic viewpoint were the campaigns of Genghis Khan and his general, Sabutai, in the 13th century. Their advance planning and bold strategic maneuvers in broad sweeps from Mongolia across Asia and Europe showed an appreciation of strategic problems most unusual for their age.

The absence of strategy in ancient warfare

In the transition to modern times two other figures who touched on the field of strategy are often mentioned—Niccolò Machiavelli in the realm of military thought and Gustav II Adolf in the field of generalship. Machiavelli's *Art of War* (1520) emphasized the larger aspects of war, particularly the close relationship between the civil and military spheres. A century later, Gustav, ruler-general of Sweden, intervened in the Thirty Years' War and, maneuvering skillfully, drove his enemy's armies out of northern Germany.

18th-century warfare. After the death of Gustav Adolf in 1632, warfare again settled down to a slower pace and a more stable mold. The 17th and 18th centuries experienced the growth of professional armies loyal to the king. But the great cost of building and maintaining such armies led to a concern for their safety, a hesitation to risk them in bloody encounters, and a preoccupation with defense and fortifications. Strategy during this period was essentially of limited aim and was greatly concerned with the art of siegecraft, for which elaborate rules were prescribed. In Prussia of the mid-18th century, however, circumstances compelled Frederick the Great to try a new and aggressive approach and to break through the accepted military pattern of the day.

The aggressive, flexible strategies of Frederick the Great

Confronted at the outset of the Seven Years' War (1756–63) by a coalition of Austria, France, Russia, Sweden, and Saxony, Frederick found himself virtually surrounded. His task was to devise a strategy to defend his territory and not to dissipate his outnumbered troops. The strategy he evolved did not follow set rules or recipes. Indeed, never was the definition of strategy as a "system of makeshifts"—offered in a later age by the Prussian general Count Helmuth von Moltke—better demonstrated. In his planning Frederick capitalized on two valuable assets—his army, a superior and highly disciplined instrument of war, and a central position. He sought always to keep the initiative, to attack first one enemy and then another, to assemble at decisive points a force superior to that of his foe, and to avoid long, drawn-out wars. Using his central position to concentrate against individual armies of the enemy before they could be reinforced by others, he developed the classical "strategy of interior lines." But even Frederick, the statesman-warrior, could not entirely escape the conditions imposed by the warfare of his times. Indeed, the statesman imposed caution on the warrior. He could not expose his costly armies to the risk of destruction and bloody decision by battle. His battles were not those of

annihilation. In the end his wars were decided by reasons of state, and those wars left his nation exhausted.

The strategy of maneuver and position

The age that immediately followed Frederick chose to imitate his caution rather than his aim. Military theory was characterized by ideas of victory without battle, maneuvering for position, a system of lines and angles of operation. Geometric concepts and cunning tricks and artifices replaced the aim to destroy enemies. Great emphasis was put on terrain and the occupation of key geographic points. The 18th century, it must be remembered, was the era of enlightenment, and warfare conformed to the spirit of the age. Strategy, like all warfare, became "mathematical" and "scientific." Theorists optimistically maintained that a general who knew mathematics and topography could direct campaigns with geometric precision and win wars without even fighting. But the new mode of warfare ushered in by the French Revolution and the Napoleonic era was soon to challenge these optimistic assumptions.

Napoleonic warfare. The French Revolutionary and the Napoleonic periods (1789–1815) witnessed great changes in the methods of war—the revolution in society accompanying and reinforcing the one in warfare. When Napoleon, the first great military strategist of the modern Western world, burst upon the European scene, the groundwork for a new age in warfare had already been laid. The French Revolution gave birth (1793) to the "nation in arms," and all Frenchmen became liable for military service. The patriotic citizen-soldier succeeded the mercenary professional. Skirmish tactics, or the loose formation, replaced the straight line; divisional organization came into use, along with lightweight artillery of great range and firing power. When Napoleon came to reap the benefits of these changes, he completely transformed strategy as well as tactics. He applied the same basic principle to the one as to the other—never to divide his forces but to concentrate all his might against the enemy forces at the critical point. His emphasis was on careful preparation, on uniting his forces before the action, on overpowering weight of striking power, on shock attack, on great daring, and on bloody decision by battle. His methods were simple, direct, overpowering—even brutal; his aim was nothing short of the destruction of the enemy forces. Against such power, neat geometric calculations stood little chance and ordinary stratagems were helpless. Again and again he showed his military genius for bringing a mass to bear against the flanks of his enemy, for selection of battlegrounds advantageous to his forces, and for deploying his forces for battle. He gave supreme expression to the idea of victory by battle.

The overwhelming concentration of force

Though as a military leader, operational strategist, and tactician of the battlefield Napoleon is regarded by many as unparalleled, in the larger field of national or grand strategy he had shortcomings. Embodying in his own person the leadership of the state and its military affairs, he recognized the value of incorporating political and economic measures, along with military moves, to increase the chances of victory in war. But he could not successfully grasp and cope with the challenge finally put to him by Great Britain and its European allies. British strategy sought to meet the Napoleonic threat to Europe by using naval power to blockade the Continent and by conducting a war of exhaustion on land through peripheral warfare, such as the Duke of Wellington's famous campaign on the Iberian Peninsula. Napoleon's reply to the British naval blockade was the continental system prohibiting British goods from entering. But this helped bring his downfall, since he was needed everywhere—to hold the coast and to fight in Spain, in Holland, and against Austria and Russia. His veteran French forces were dissipated, and he had to rely on impressed nationalities of Europe. Eventually the coalition of his enemies was to use the methods and means of warfare that the French Revolution had introduced and Napoleon had perfected to reinvigorate their own forces and overwhelm him.

The protracted war of exhaustion

Despite his mistakes, Napoleon's preeminent place in the history of strategy is secure. His tactics and strategy influenced military leaders for a century. His maxims were widely studied and were said to have been carried in the saddlebag of the famous Confederate general of the American Civil War, Thomas J. ("Stonewall") Jackson. Students of strategy have long pointed to Napoleon's battles and campaigns for classical illustrations of "principles" of war—of surprise, mobility, concentration of force, and economy of force. Possibly more than that of any other general, his competent practice oriented modern military theory toward the search for underlying principles. Indeed, the art of strategy as evolved by theorists since 1800 may be traced largely to his operations. For this development two great interpreters of Napoleonic strategy, Antoine-Henri Jomini and Carl von Clausewitz, were especially responsible.

Carl von Clausewitz. Clausewitz (1780–1831), a Prussian, was the first great student of strategy and the father of modern strategical study. He was trained in systematic study of philosophy in the school of Immanuel Kant, thus it was natural for him to range widely over the whole field of military knowledge and to reduce Napoleonic warfare to a unified philosophical conception. His famous work *On War,* written as an outgrowth of his studies of Napoleon's campaigns, remains the best general study of the art of war. He died with his work unfinished, but his writings published after his death became the standard textbooks on war in Prussia and elsewhere. Their influence was felt profoundly in the Franco-German War of 1870–71, and leading generals of World Wars I and II were brought up on them and on the works of his followers.

The contributions of Clausewitz to strategic thought are many and diverse. To some his work is the Bible of strategy, and, like that great book, susceptible to many conflicting interpretations. His work set forth fully and clearly for the first time the relationship between political and military leadership. He dwelt on decision by battle as the first rule of war, on seeking the destruction of the enemy's forces, and on achieving superiority at the decisive spot. Rejecting the optimism and rationalism of the 18th century, he held that war was not a scientific game but an act of violence. Mathematical and topographical factors, he held, were important in tactics but less so in strategy. "We . . . do not hesitate," he asserted, "to regard as an established truth that in strategy more depends on the number and the magnitude of the victorious combats than on the form of the great lines by which they are connected." The key to victory was battle, however bloody. He defined strategy as the employment of battles to gain the end of war.

Victory through destruction of the enemy's forces

Clausewitz devoted much of his work to showing that war is both a social development and a political act. He went further and said that "war is not merely a political act, but also a real political instrument, a continuation of policy carried out by other means." War was therefore not an independent phenomenon unto itself to be handed over to soldiers and sailors. Again and again he asserted that military and political strategy must go hand in hand. "War," he declared, "admittedly has its own grammar, but not its own logic."

Clausewitz' emphasis on the aim of strategy as the destruction of the enemy's forces on the battlefield has had a great influence on subsequent military thinking. His disciples, however, have generally overlooked the fact that he also recognized another strategical form—a strategy of limited aim for limited warfare, of wearing down an opponent. When Clausewitz wrote, warfare was conducted in two dimensions, and it was rarely possible for one nation to impose its will without first destroying the opposing army. But Clausewitz recognized clearly what many of his followers in subsequent generations forgot, that the destruction was only a means to enforce policy and not an end in itself.

Antoine Jomini. In the history of military thought, the French general Antoine-Henri Jomini (1779–1869), one of Napoleon's staff officers and a contemporary of Clausewitz, presents a striking contrast to the Prussian philosopher of war. Lacking the philosophical bent of Clausewitz, Jomini concentrated his thinking on what he regarded as practical issues in war rather than on war as a whole. He became the chief expounder of Napoleonic methods, and out of his studies evolved a theory of strategy. Although he opposed "systems of war" purporting to provide for

all contingencies, he nevertheless believed that in the field of strategy certain rules and general principles—eternally true—could and should be formulated "as a compass for the commander-in-chief of an army." To establish these principles, he believed, was the major problem of military science.

Victory through the occupation of territory

The heart of Jomini's theory lay in the theatre of war and the campaign. But he thought primarily of occupying all or part of the enemy's territory rather than of annihilating his army. This occupation was to be achieved by progressive domination of zones of territory. Jomini emphasized throughout his work the proper choice by the general of decisive maneuvering lines and their adaptation to geometric configurations of zones of operation. Campaigns must be carefully planned in advance. The task of strategy is to make preliminary plans—to establish lines of operation and to bring military means into conformity with geographic realities of the chosen zone of operations. He laid down two basic principles: massing troops against fractions of the enemy by rapid movement and striking in the most decisive direction.

Jomini's great contribution to military thought lay in his definition of the place of strategy in warfare. Probably more than any other work, his *Summary of the Art of War* (1838; *Précis de l'art de la guerre*) fixed the major fields of modern military art. Subsequent wars were to cast doubt on much of his work, particularly on his conception of geographic campaigns and of the superiority of interior lines of operation. But, like Clausewitz in German strategic thinking, Jomini had an enduring influence on French military thought. His emphasis on planning for operations and on intelligence took root in military staffs and schools throughout Europe, and his work became the textbook for the conduct of the American Civil War.

American Civil War to World War I. Often called the first of the really modern wars, the American Civil War (1861–65) marked a transition to a new era in strategy. It gathered up new phenomena that had begun to influence warfare in the middle of the 19th century and whose fuller consequences were to be felt in the half century that followed. It was a period marked by refinement of the old in strategic theory and practice and by the addition of new strands—by such famous figures as Robert E. Lee, Ulysses S. Grant, and William T. Sherman on the battlefield and Count von Moltke, Count von Schlieffen, Hans Delbrück, and Alfred T. Mahan in the literature of strategy.

Strategies of North and South. The Civil War is significant in the history of strategy in a number of ways. The basis of strategy—particularly factors of time and space—began to change. The use of steam power for land and water military transportation received its first major test. Railroads gave strategy a new speed of movement but tended to make strategy stick to straight lines and fixed routes. The Civil War also tested ironclad ships and heavy naval ordnance. The relation among the combat arms was completely upset by the introduction of the long-range infantry rifle. The accuracy of long-range weapons in the hands of defending infantry shattered the effectiveness of the rapidly concentrated attack in which Napoleonic strategy had culminated. But, as so often has been noted in the history of warfare, armaments and weapons are more readily changed than ideas, and Napoleon's principles continued to be maintained, sometimes with disastrous consequences on the battlefield.

The growing speed and scale of warfare

Aside from the effect of new inventions, the Civil War revealed the growing importance of the economy and manpower in war. Industry was called on more and more, and conscription was adopted to provide manpower. The war also revealed the impact of systematic West Point training received by the leading generals on both sides. Finally, the Civil War was long studied for classic examples of maneuver and of offensive and defensive strategy and for lessons in the relation among policy, strategy, and means of war. Essentially the Civil War demonstrated local or theatre strategy and tactics. Though elements of grand strategy were at hand—political, economic, military, and psychological—the art was still not well understood or consistently applied. Despite conscription and the partial mobilization of industry and the railroads, there was no well-worked-out grand design correlating the widely scattered forces and the war industries that supported them.

Effects of political objectives on strategy

The strategies of the North and South were rooted in different political objectives. The objective of the North was to prevent the Confederate states from seceding from the Union, that of the South was to attain independence. Because the South was greatly inferior to the North in population and resources, it could not hope to conquer the North. The dual purposes of its strategy were to convince the North that forcing the South to remain in the Union was not worth the cost and to bring about foreign intervention in favour of the South. General Robert E. Lee, the great Southern leader, believed the best way to realize these objectives was to carry the war into the North and to defeat the Northern armies in their own territory. For a time, therefore, his strategy was essentially offensive. But after his defeat at Gettysburg, Pa., he no longer had the wherewithal to continue the offensive, and at the same time it became obvious that foreign intervention would not be forthcoming. From that time to the end of the war his strategy was defensive, with the object of wearing down the patience, if not the power, of the North.

To achieve its political object, the North, on the other hand, developed another strategy. The Federal design had three main goals: (1) to blockade and isolate the Confederacy, (2) to cut it in two, and (3) to strike at Richmond, Va., its capital. The naval blockade, though not completely effective, brought virtual commercial isolation to the South. Partition was gained by capturing Vicksburg on the Mississippi in July 1863 and by severing the east–west railroad connections. The capture of Vicksburg cut the South off from its sources of supplies beyond the Mississippi. Only gradually did the North change its design from that of attacking Richmond to that of striking at the main army of the Confederacy and the remaining sources of supply. Grant's elevation to the supreme command of the armies in March 1864 enabled him to put this concept into effect. The famous march of General Sherman through Georgia to the sea in the fall of 1864 was an outstanding example of strategic maneuver and surprise. Leaving his supply line, Sherman feinted against one city and attacked another, finally cutting off Lee's army in Virginia from its war resources in the South. The cooperation of the Federal eastern and western armies in a grand converging movement resulted in the evacuation of Richmond and, finally, in the surrender of Lee's army to Grant at Appomattox Court House in Virginia in April 1865.

The period from the close of the Civil War to the outbreak of World War I saw the further growth of trends already apparent. Space and time factors began to appear in a new light. A nation with a well-developed railway net gained significant advantages in war. The speed of mobilizing and concentrating armies became a basic element in strategic calculations, and the timetable based on it became the heart of staff plans drawn up in anticipation of war. Increased firepower in the machine gun, universal liability of able-bodied males for military service, rapid mobilization of reserve military units, and increased potential of fortifications influenced military planning.

The Prussian-German strategists. Strategic thinking in the half century before World War I showed a remarkable diversity. To the Prussian-German school—Moltke and Schlieffen—the new trends in warfare seemed to reinforce Clausewitz' teachings about battles and the aim of defeating the enemy's armies.

Moltke's emphasis on offense and "ad hoc expedients"

To Field Marshal Count Helmuth von Moltke (1800–91) belongs the chief credit for molding the Prussian army into a formidable war machine, which defeated the Danes (1864), Austrians (1866), and French (1870–71). Moltke agreed with Clausewitz that battles were the primary means of breaking the will of the enemy. But Moltke did not believe a strategist could follow a rigid set of rules. To him strategy was a system of "ad hoc expedients." It was "the art of action under the pressure of the most difficult conditions." No plan of operations, he believed, could look with any assurance beyond the first encounter with the main enemy forces. The offensive, according to Moltke, is "the straight way to the goal," whereas the defensive is "the long way around." He became famous

for his skillful conduct of operations on the outer line leading to encirclement. In addition to exploiting the altered conditions of space and time created by the railroads and improved highways, he capitalized on the possibilities offered by the telegraph for handling armies of great size. Recognizing that the field of operations had become too vast to be surveyed by the eye of the commander, he introduced a new system of delegating power to subordinate commanders. Broad directives took the place of detailed orders. Moltke always fought with superior forces, and his wars, culminating in that against France in 1870–71, are regarded by some authorities as classical models of conception and execution in military strategy.

Schlieffen's "strategy of annihilation"

Field Marshal Count Alfred von Schlieffen (1833–1913), chief of the Prussian general staff before World War I and ablest of Moltke's successors, carried the strong line of strategic reasoning running from Napoleon through Clausewitz and Moltke to its logical conclusion in his conception of a "strategy of annihilation." Like Moltke, he stressed the military side of strategy, the concentration on decisive victory by battle. But, unlike Moltke, he could not count on superior forces and had to prepare for war on two fronts. The basis for German strategy before World War I as developed by him was embodied in the famous Schlieffen plan. The plan was extremely simple. The bulk of the German forces were to attack the nearest opponent, the one in the west (France), and to defeat him in a great battle; meanwhile, in the east (Russia) the Germans would stand on the defensive. Schlieffen proposed to gain the decision in the great battle by means of an enveloping attack—if possible, by a double envelopment. Once the enemy in the west was defeated, the Germans would attack the foe in the east. This was the essence of the plan with which the Germans entered World War I.

Schlieffen's theories were to have wide influence, largely through his book *Cannae.* Analyzing Hannibal's great victory over the Romans in 216 BC, he had developed his theory of the battle of annihilation by means of encirclement and double envelopment. The decisive German campaign against the Russians at Tannenberg, East Prussia, in August 1914 was fought in this mold, and Schlieffen's theories were studied exhaustively in the higher army schools of the United States and Europe after World War I. As General Walter Bedell Smith, chief of staff to General Dwight D. Eisenhower, supreme commander of the Allied Expeditionary Force in World War II, pointed out, General Eisenhower and many of his staff officers, products of these schools, "were imbued with the idea of this type of wide, bold maneuver for decisive results."

Political-military strategy: Delbrück and Mahan. Moltke and Schlieffen thought of war as military action—the speediest decisive defeat of the main opponent. But the closing years of the 19th and early years of the 20th centuries witnessed the emergence of new approaches and different emphases in strategy. Two thinkers looking to past history for light on the problems of their times made signal contributions to strategic theory—one, Alfred Thayer Mahan (1840–1914), an American, in the field of naval strategy, the other, Hans Delbrück (1848–1929), a German, in the area of military strategy. Each recognized an intimate relationship between war and politics in every age, that political and military (or naval) strategy must be in harmony. Each showed an awareness of the growing importance of the economic bases of strategy, of state policy, geographic position, and available means as determinants of the mode of strategy, and of accommodating strategic action to suit the particular times and needs.

Delbrück's "strategy of exhaustion"

Carrying forward a line of thinking already suggested by Clausewitz, Delbrück presented his theory of the "strategy of exhaustion"—of wearing down an opponent by a variety of means. Clausewitz had merely indicated the existence of two methods of conducting war—one aimed at annihilation of the enemy, the other limited warfare. Delbrück expounded on the differences. The sole aim of the strategy of annihilation he identified as the decisive battle. The second type he called variously the "strategy of exhaustion" and "two-pole strategy." The commander could move between battle and maneuver; the political object of war could be obtained by other means than battle—by occupying territory, blockade, destroying crops or commerce. In Delbrück's view, Alexander, Caesar, and Napoleon had been strategists of annihilation; Pericles, Gustav II Adolf, and Frederick the Great, equally great generals, exponents of the strategy of exhaustion. Holding that the strategy of exhaustion was just as valid as the strategy of annihilation—each depending on the political aims and means at hand—Delbrück's theories ran counter to the military thinking of his day and brought down a storm of criticism about his head. But he persisted in reminding his age, intent on victory by battle, of other important and forgotten aspects of Clausewitz' teachings.

While Delbrück was battling his military critics in Germany, a scholarly navy captain and teacher at the Naval War College in Newport, R.I., was quietly breaking ground in pursuing his brilliant researches in military and naval history and strategy. This pioneer was Alfred Thayer Mahan, indefatigable student of the strategy of Napoleon and Jomini. His masterly works, *The Influence of Sea Power upon History, 1660–1783* (published in 1890) and *The Influence of Sea Power upon the French Revolution and Empire, 1793–1812* (published in 1892), marked a revolution in naval thought. While advances in technology were affecting naval architecture and weapons, and steam, armour plate, and rifled guns were coming into vogue, Mahan aimed to bring naval strategic thinking up to date. The books and articles that poured from his pen down to World War I about warfare of the second dimension—the sea—had a profound influence on the theory of warfare and on naval policy and strategy in many countries.

An advocate of a big navy, of overseas bases, of national greatness through sea power, he was the American apostle of "looking outward." Mahan emphasized the significance of commerce in war and of economic warfare through the application of sea power. His researches convinced him that the nation or group of nations that commanded the seas could best draw on the trade, wealth, and economic resources of the world and was the more likely to win wars. Strongly influenced by Jomini's teachings, he looked for fundamental truths and formulated "principles" of naval strategy. Naval strategy and sea power, he recognized, were conditioned by a nation's insular or continental situation. To Mahan, a central position gave the same great advantages on the sea or on land—interior lines. Concentration of force he viewed as a fundamental principle of land and sea warfare. The backbone of fleet strength, in his opinion, was the battleship or capital ship. Under Mahan's tutelage, the twin theories of coastal defense and commerce raiding, which hitherto held sway in American naval strategy, gave way to the theory of command of the sea. Command of the sea can be defined as that condition under which friendly ships can use the sea freely but under which an enemy, though venturing out to sea, cannot do so with any security. Mahan's concepts of naval strategy and faith in preponderant naval power and the use of the navy as an instrument of national power were accepted by the U.S. Navy and President Theodore Roosevelt. His doctrines stimulated the trend toward overseas expansion and growth of the navies of the world between 1898 and 1914. The sudden acquisition of an overseas empire in the Spanish-American War of 1898 greatly changed the strategic position and problems of the United States. Its emergence as a world power, beginning in these years, was to have important bearings on the strategic balance of power among the nations of the world, an equilibrium that World War I altered profoundly.

Mahan's command of the sea

World War I. *The war.* World War I, the first of the great coalition wars of the 20th century, was an important landmark in the story of the evolution of modern strategy. Never was the phenomenon of cultural lag as applied to warfare more clearly demonstrated. Beginning in the accepted mold of strategic planning popular since 1870, it soon ran head on into countertrends that were altering the very bases of strategic action and that strategic thinking in the intervening years had not yet fully grasped. Despite the experiences in the South African and Russo-Japanese wars with the machine gun as a defensive weapon of tremendous firepower, French and German military leaders at the outbreak of the war continued to put their faith in the

The failure of mass offensives

offensive. In fact, they were convinced that new weapons and methods of control, the radio and telephone, actually improved the offensive capabilities of their mass armies. The universal underestimation of the effect of modern firearms on the defense had important repercussions on strategy both during and after the war.

The first moves in the war began in 1914 as French and German strategists had planned. In seven days the Germans concentrated more than three million men on the eastern and western fronts from mobilization points. In approximately the same time the French assembled 1.2 million men on the western front. Both sides made heavy use of railroad lines to speed assembly of great masses of troops. Both sides were determined to attack. Out of the movements of mass armies came the first battles on the frontiers. As Schlieffen had planned, the Germans catapulted into Belgium, but the enveloping wing was not as strong as Schlieffen, who had died the previous year, had wished. It was compressed into a smaller corridor by the political decision not to violate Dutch neutrality. The anticipated six-week campaign of annihilation against France envisaged by Schlieffen could not be executed. The French attack also soon hit a snag. Although the French army's right wing reached the Rhine, its centre was endangered by a German pincer movement. Only a hasty retreat and a counteroffensive at the Marne River saved Paris. "Pinwheel strategy"—each side attacking and driving the enemy back—had stalled badly.

Meanwhile, on the eastern front, the German prewar strategy of holding until France had been quickly defeated was compromised by the desire of the Austrian ally to push against the Russians, partners of the French. The German victory at Tannenberg counterbalanced the Austrian defeat at Lemberg (Lvov). The eastern front became stabilized.

By the close of 1914 the war had become a stalemate on both the eastern and western fronts. The conflict had resolved itself into trench warfare from Switzerland to the English Channel. Machine guns and artillery took over the battlefield. The conflict had settled down into a war of position, and strategic mobility was lost. World War I became a classic case of arrested strategy.

Breaking the defensive line

The first phase of the war was over by the end of 1914. Prewar plans had failed; the war of movement, of mass offensives, had ceased. The big question thenceforth was how to dig the war out of the trenches. In answering that question important elements of grand strategy came into play. The heavy demands upon industry for munitions of war multiplied, and technology was called upon for new means—the tank and poison gas—of breaking the stalemate. Britain's naval blockade to starve Germany took on added significance. The German countermeasures helped bring the United States into the war in 1917. The United States was not prepared for war, however, and the buildup of its forces across the Atlantic was slow. The Germans, seeking in 1918 to forestall the full impact of U.S. might, put their resources into a great offensive that came close to succeeding. When the Americans finally arrived in force, they played a valuable part in military strategy in reducing the salients within the Allied lines. Eventually the German allies were defeated; the German armies reached a point of exhaustion and the homeland a stage of semistarvation. Germany asked for an armistice.

The lessons. Although much has been written about World War I, the strategic lessons of that conflict for coalition warfare have not been fully comprehended. Never was the dependence of strategy on statecraft more clearly demonstrated. As political circumstances of the war changed, strategy changed. The Central Powers, led by Germany and Austria-Hungary, never had a common plan of campaign or effective unity of command. The Allied side achieved unity only under necessity. Along with the military factors, economic and psychological considerations proved important in conducting the war and gaining victory. Although the aim of annihilating the enemy was paramount with both sides—especially in the opening campaigns—the desire to exhaust him also influenced strategy, and fresh confirmation was given to Moltke's description of strategy as a "system of makeshifts."

The conduct of mass warfare

Military leaders in World War I had to master three basic factors in strategic calculations: masses of men, technological advances, and wide areas. The movement of huge masses became an art in itself, for armies had taken on unprecedented dimensions. Millions of men were in action. Railroads and motor transport became important not only for concentrations but also for establishing new strategic points on the fronts themselves. The arena of war embraced whole continents. Battles lasted for days and weeks, and the fighting continued even after the great battles were over.

New weapons came into play. Aerial reconnaissance enabled a commander to gain some insight into the enemy's intentions and movements. New means of communication—telephone, radio telegraphy, the automobile, and the airplane—promoted faster execution of orders and unified command over widely scattered forces. The overwhelming firepower of modern weapons checked the effectiveness of the attack, long considered the ideal path to victory. The tank, however, offered fresh possibilities in redressing the balance between the defensive and the offensive. Tactics became more than ever a prelude and conditioning factor of strategy, since without freedom of movement, strategy was only an academic exercise. Tactics came to mark the beginning rather than the conclusion of an operation.

There were also larger strategic influences at work. If World War I was a war of masses, it was also a war of matériel. War was becoming increasingly total and cut deeper into the life of the nation. Some of the foremost leaders and students of World War I—notably Winston Churchill and Georges Clemenceau—recognized that military strategy had become but a part of a greater national strategy. Symptomatic of this thinking was Clemenceau's widely quoted statement that war was too important a business to be left to soldiers. More than ever strategy and politics would have to be correlated. The increasing totality of modern war would have to be matched by a broader national strategy. The large impact of the war in the international sphere—the effects of the defeat of Germany, the weakening of England and France, the rise of the Soviet Union on the strategic balance of power in the world—could not yet be foreseen.

Between World Wars I and II. The period between 1918 and 1939 saw strategy once more in process of flux. As an outgrowth of the experience of World War I, strategy came largely to mean defense. In France, particularly, a mentality favouring fixed defenses began to take hold, eventually leading to the building of the concrete fortifications of the Maginot Line, bordering Germany. The belief was strong that field fortifications aided by the machine gun would contain any attack. The huge losses of World War I would thereby be avoided.

Douhet: air supremacy. Countertrends, however, were soon to dispute this prevalent emphasis in strategic thinking. One strong challenge came from the new school of exponents of air power. In World War I the air arm had had its beginnings. The period between the end of World War I and the beginning of World War II saw it come into its own; air forces and air organization expanded greatly. Theorists began to develop the strategy of warfare of the third dimension. Foremost among these was the Italian general Giulio Douhet (1869–1930). He first presented the doctrine that the air arm alone would decide wars of the future. In his view, land and sea forces would no longer be decisive. On the ground, armies could act henceforth only on the defensive, since attack, and with it the decision, could be gained only through the air. Air power could quickly conquer time and space. The air arm could circumvent every kind of ground resistance and nullify fortified positions and obstacles of terrain. It could strike at the enemy's sources of power before his armies could fire a shot. It could strike at his capital, industrial centres, and communications. In short, it could so reduce his ability and willingness to resist that he would surrender. Douhet proposed to expand the air arm as much as possible, keep land and sea forces only as support for war in the air, and gain control of the air by defeating enemy air forces in battles or destroying them in their airfields. He made strategic bombing and the industrial objective—

The neutralization of land and sea forces

strikes at the opponent's heart—the core of his doctrines.

Douhet's epoch-making ideas found many supporters in other countries. This school of thought generally argued that huge armies would no longer be necessary. The opponent's will could be overcome even if his armed forces remain undefeated. Some of Douhet's adherents went further and demanded the abolition of land and sea forces altogether. In any event, the rise of air power accentuated the need of thinking of strategy as dealing with something more than the movements of armies on land or of ships at sea.

Fuller: mechanized armour. Meanwhile, army leaders began to advocate another solution to break the strategic stalemate of World War I. To overcome the superiority of the defensive, they put their faith in developing a modern cavalry of tanks and armoured, motor-driven vehicles. The best known among the great champions of mechanization and motorization that arose in Great Britain was Major General J.F.C. Fuller (1878–1966). These advocates saw in the armoured vehicle, combining firepower, extreme mobility, and armoured protection, the best answer to overcoming defensive forces relying on machine guns. This system was particularly suited to needs of an insular country, protected by a strong air force and navy, and of a relatively small army intended primarily for expeditionary purposes in support of continental allies. But this solution on the ground found support in Germany, the Soviet Union, and the United States. In France, Charles de Gaulle bucked the strong tide of opinion and advocated tank warfare to restore a strategy of mobility and the offensive.

In the late 1930s the Germans combined air power and tanks into a new form of assault that also aimed to overthrow defensive superiority. Developing a highly mobile form of warfare for lightning strikes and mechanized attacks, they were to contribute the art of the blitzkrieg—the spearhead of a conquering, offensive strategy that Hitler unleashed in World War II.

Ludendorff: total war. In Germany, too, other influences supporting offensive strategy came to the fore. To overcome strategic stalemate of the World War I variety, the German general Erich Ludendorff contributed his theory of total war. He envisaged total mobilization of a nation's human power and resources for war. The nation at war would be led by a supreme military commander; strategy would dictate policy. The concept of total war moved geography and economics into prominent positions in Nazi thinking. Even before World War I the British geographer Halford J. Mackinder had posed the potential threat of a heartland power, in control of Eurasia, to sea power—a counter to Mahan's theory of control of the seas. The German geopoliticians after the war took over the "heartlands" concept, and through their teachings the concept of control of Eurasia became embedded in Nazi statecraft. Their doctrines gave support to the main strands in Hitler's offensive strategy—continental expansion, autarky (national economic self-sufficiency), and *Lebensraum* ("living space").

Complete mobilization of resources

Before war burst upon Europe in 1939, it was apparent that important changes also were brewing in naval strategy. All major sea powers were producing high-speed battleships, and the aircraft carrier was becoming a significant and integral member of the fleet. In the crucible of World War II, the emerging elements in ground, air, naval, and nonmilitary strategy were to take clearer shape. (Ma.M.)

World War II. *The first phase.* In spite of the ideas of military reformers such as Douhet and Fuller, the lesson derived from World War I by orthodox strategists throughout Europe was that war between industrial societies would involve the total mobilization of all national resources and would be a test as much of economic strength and civilian morale as of military skills.

Germany was no exception. Hitler's military advisers warned him that the Third Reich, which had begun to rearm only in 1934, would not be ready to confront France and Britain until 1941 at the earliest. Hitler in September 1939 ignored this advice. He recognized that his adversaries were even less ready for war than he was, and he had, in the combination of infantry, armour, and air support developed by elite units of the Wehrmacht, an ideal instrument for his immediate objective of overrunning Poland. It proved equally successful when turned against the West in May 1940. Hitler's panzer divisions broke through the French defenses at the Ardennes and cut the Allied armies in two. The forces to the north had to be evacuated over the beaches of Dunkirk with the loss of all of their heavy equipment. A week later the Germans attacked the demoralized remainder of the French armies to the south, and on June 16 the French government asked for terms. Germany accepted the surrender of nearly two million French prisoners and had lost only 50,000 men.

The French defeat had been as much moral as military. The people had faced the prospect of war with dread and had accepted with relief the assurance of their military leaders that World War I had proved the invincibility of the defensive. The "Maginot mentality" had little to do with the Maginot Line itself—that system of fortifications along France's eastern frontier that was designed quite properly to economize on personnel and that the Germans never attacked. Rather, the French military leadership simply could not believe in the possibility of open warfare and therefore had provided their forces with neither the equipment nor the training nor the communications to undertake it.

The "Maginot mentality"

The British had relied on France to provide the main land forces while they were to deploy their traditional strength at sea and contribute a long-range bomber force to attack German industrial strength at the source. But the Royal Air Force (RAF) in 1939 had neither the range nor the equipment nor the capacity to attack the German homeland in the necessary strength, and it was to take three years for its raids to have any serious effect. The RAF came into its own in a defensive role in the summer of 1940 when Hitler launched his Luftwaffe to obtain command of the air over Britain as a preliminary to a seaborne invasion. Britain had made a massive investment in radar early-warning systems and fast-climbing fighters, and these aircraft, fighting over their own territory, just turned the scale in what became known as the Battle of Britain. Hitler postponed the invasion and decided instead to neutralize Britain by submarine blockade and air bombardment while he prepared to attack the Soviet Union.

If Hitler was to achieve his long-term objective of a self-sufficient Third Reich controlling the grainfields of the Ukraine and the oil of the Caucasus, he had to destroy the Soviet Union sooner or later. His decision to do so sooner was probably precipitated by his need for those resources if he was to defeat a Britain sustained by all the resources of a United States—which, although at this stage determined to keep out of the war, was equally determined not to see Britain lose. Further, Hitler and his military advisers held Russian military capability in contempt. The incompetence of the Red Army had been shown when, in the Russo-Finnish War of 1939–40, it had taken three months to break the resistance of the small state of Finland. The directive for attack was issued on Dec. 18, 1940, and the attack launched on June 22, 1941.

At first all went as expected. German armoured divisions, driving deep in a series of encircling attacks, had by the beginning of December occupied 900,000 square miles (2,340,000 square kilometres) of Soviet territory, taken three million prisoners, and reached the suburbs of Moscow. But the Soviet leadership was prepared for such setbacks. Soviet prewar military doctrine had stressed total mobilization of the population, elasticity in operations, and defense in depth. Soviet war industries lay behind the Ural Mountains, beyond Hitler's reach. Reserve forces were switched from the Far East and counterattacked north of Moscow on Dec. 5, 1941. Two days later Japan attacked the U.S. Pacific Fleet at Pearl Harbor, and the war became global.

Global war. Japan had been gradually expanding its power on the Asian mainland since the beginning of the century and had been at war with China since 1937. In 1940 the defeat of France, Britain, and The Netherlands in Europe exposed their possessions in Southeast Asia to attack. Japan's military leaders resolved to seize them so as to establish an autarkic empire immune to the eco-

Japan's strategy of expansion

nomic sanctions by which the United States was attempting to check their expansion into China. The attack at Pearl Harbor was a preemptive strike to gain command of the western Pacific. Simultaneously, Japanese forces attacked the Philippines, Malaya, and the Dutch East Indies. The British fortress of Singapore capitulated with 130,000 troops to a Japanese force 15,000 strong on Feb. 15, 1942. An American garrison held out at Corregidor in the Philippines until May, but within three months of launching their attack the Japanese had gained all of their immediate objectives.

The United States thus found itself at war, alongside two hard-pressed allies, with two triumphant military empires that were still expanding their power. Germany, checked through the winter, opened a powerful offensive in the spring of 1942 toward the Caucasus. In June, Japan launched a further drive toward the U.S. bases in Hawaii. This was foiled on June 5 by the decisive Battle of Midway, which not only established U.S. command of the seas but confirmed the role of the aircraft carrier as the new "capital ship" of naval warfare. By the end of 1942 Germany and Japan had expanded to the limit of their capacity. At the long Battle of Stalingrad (August 1942–February 1943), German offensive power was broken by Soviet armies as decisively as the Japanese had been by the U.S. Navy at Midway. The Allies were now able to mobilize and deploy an overwhelming superiority of resources. But German and Japanese defensive strength was still intact, and German submarines prevented the full deployment of Allied strength until the spring of 1943.

British and American war leaders agreed that defeat of Germany should be given first priority—a decision accepted with reluctance by the U.S. Navy, whose prewar planning had been directed primarily against Japan. But whereas the U.S. strategy for a European war was simple—accumulation of a large force in Britain in 1942, a cross-Channel attack and decisive battle in France in 1943—the British urged an initial concentration in the Mediterranean, where their troops had been fighting in North Africa since 1940, in order to knock Hitler's ally Italy out of the war and divert German forces from the Russian front. Both agreed, however, on a combined bombing offensive against Germany, and by 1944, in spite of German countermeasures, Allied bomber forces were inflicting lethal damage on the German economy. A series of compromises resulted in Allied landings in French North Africa in November 1942, in Sicily and Italy in 1943, and finally a cross-Channel attack in 1944, just as the victorious Soviet armies were advancing into eastern Europe. Invaded from east and west, its cities destroyed from the air, Germany capitulated on May 8, 1945.

Although Britain fought a successful campaign to recapture Burma (Myanmar), and the Soviet Union participated in the last days of the war, the direction of the war against Japan was the exclusive concern of the United States. The U.S. Army favoured a land-based approach via New Guinea and the Philippines to mainland China, whence a major invasion could be launched against the Japanese homeland. The Navy pressed for a direct advance via the islands of the central Pacific, so as to blockade and starve out Japan. Resources were in fact available for both strategies. With complete command of the sea, U.S. forces had by March 1945 captured island bases from which their heavy bombers could destroy Japanese cities one by one. The atomic bombs dropped on Hiroshima and Nagasaki on Aug. 6 and 9, 1945, were only the coup de grace in this process of destruction.

Victory of superior resources

The orthodox prewar strategists were proved correct: World War II was indeed eventually won by the mobilization of superior resources, ruthlessly and often wastefully employed against militarily more skillful foes. The use of nuclear weapons signaled both the consummation and the transformation of total war. (M.E.Ho.)

STRATEGY IN THE NUCLEAR AGE

Strategic thinking after World War II, at least with regard to conflict between the great powers, was dominated by two related developments. The first was nuclear weapons, which raised the prospect of war as the ultimate catastrophe. This led to a shift of focus in strategic thinking toward the deterrence of war rather than the waging and winning of war.

The second development was a decades-long continuity in the East–West conflict, with two alliances each dominated by a superpower—the North Atlantic Treaty Organization (NATO) by the United States and the Warsaw Pact by the Soviet Union. Although attempts to reproduce these alliances in continents other than Europe met with scant success, their stability within Europe meant that they were virtually taken for granted. Therefore, there was less postwar interest in the issues of alliance formation and disintegration, which had preoccupied earlier generations of strategists.

The atomic bomb and American strategic thought. The first successful test of the atomic bomb took place in New Mexico in July 1945, as the leaders of Britain, the Soviet Union, and the United States met at the Potsdam Conference to discuss the shape of the postwar world. This context coloured the early American appreciation of the potential foreign-policy role of the new weapons, with the result that nuclear strategy thereafter became bound up with the twists and turns of the Cold War between East and West.

However, the decision actually to use the bomb against Japan reflected the more immediate urge to end the war as soon as possible and certainly before it became necessary to mount an invasion of the mainland. The atomic bombing of Hiroshima and Nagasaki in August 1945 was a means of shocking Japan into surrender. The choice of civilian rather than purely military targets, and the consequent immense loss of life, reflected the brutalizing experience of the massive air raids that had become commonplace during the war. Afterward it was assumed that any future atomic bombing would also be against cities. As weapons of terror, they appeared to have brought 20th-century trends in warfare to their logical conclusion.

Weapons of terror

The first nuclear weapons were in the range of other munitions (the bomb that destroyed Hiroshima was equivalent to the load of some 200 B-29 bombers); also, at least initially, the weapons were scarce. The key development introduced by atomic bombs was less in the scale of their destructive power than in their efficiency. By the start of the 1950s, though, this situation had been transformed by two related developments. The first was the breaking of the U.S. monopoly by the Soviet Union, which conducted its first atomic bomb test in August 1949. Once two could play the nuclear game, the rules had to be changed. Anyone who thought of initiating nuclear war would henceforth need to consider the possibility of retaliation.

The second development followed from the first. In an effort to extend its effective nuclear superiority, the United States produced thermonuclear bombs, based on the principles of nuclear fusion rather than fission, upon which the atomic bombs were based. This made possible weapons with no obvious limits to their destructive potential. Opposition to this development by influential nuclear scientists, such as Robert Oppenheimer, was disregarded by President Harry S. Truman on the grounds that the Soviet Union would not suffer from any comparable moral inhibitions.

This move was not matched by a pronounced nuclear bias in U.S. strategy. The weapons were still scarce, and it seemed only a matter of time before whatever advantages accruing to the United States through its lead would be neutralized as the Soviet Union caught up. The Truman administration assumed that the introduction of thermonuclear weapons would extend the time available to the United States and its allies (including NATO) to build up conventional forces to match those of the Soviet Union and its satellites. A series of events, from the Berlin blockade of 1948 to the Korean War of 1950–53, had convinced the United States that the communists were prepared to use military means to pursue their political ambitions and that this could be countered only by a major program of Western rearmament.

Endurance of conventional strategy

Massive retaliation. The administration of President Dwight D. Eisenhower, which came to power in January 1953, saw things differently. It reflected on the frustrating

experience of the inconclusive conventional war fought in Korea and wondered why the West had not made more of its nuclear superiority. Eisenhower was also extremely worried about the economic burden of conventional rearmament. Assigning a greater priority to nuclear weapons provided the opportunity to scale down expensive conventional forces. By this time the nuclear arsenal was becoming more plentiful and more powerful.

The strategy that emerged from these considerations became known as "massive retaliation," following a speech made by U.S. secretary of state John Foster Dulles in January 1954, when he declared that in the future a U.S. response to aggression would be "at places and with means of our own choosing." This doctrine was interpreted as threatening nuclear attack against targets in the Soviet Union and China in response to conventional aggression anywhere in the world.

Massive retaliation was widely criticized. In the United States the Democratic Party, whose policy under Truman was being reversed—and the army and navy, whose budgets were being cut at the expense of the air force's Strategic Air Command—charged that it placed undue reliance on nuclear threats, which would become less credible as Soviet nuclear strength grew. If a limited challenge developed anywhere around the Sino-Soviet periphery (the two communist giants were seen to constitute a virtual monolith), and the United States neglected its own conventional forces, then a choice would have to be faced between "suicide or surrender."

First and second strikes. Massive retaliation was also criticized for failing to appreciate possible areas of Soviet superiority. This criticism grew after the Soviet Union demonstrated its technological prowess by successfully launching the first artificial Earth satellite (Sputnik 1) in October 1957, not long after it had also made the first tests of an intercontinental ballistic missile (ICBM), the SS-6. Concern grew that the Soviet Union was outpacing the United States in missile production, so leading to a "missile gap." (It might have been argued that after a certain level of destructive capability had been reached by both sides, an effective stalemate would be reached and extra weapons would make little difference, promising only, as British prime minister Winston Churchill put it, to make "the rubble bounce.")

However, by this time nuclear strategy was becoming much more sophisticated. With the RAND Corporation, a think tank based in Santa Monica, Calif., taking the lead, new analytical techniques were being developed. These were often drawn from engineering and economics, rather than the more traditional strategic disciplines of history and politics. In a celebrated RAND study of the mid-1950s, a team led by Albert Wohlstetter demonstrated that the air bases of the Strategic Air Command could be vulnerable to a surprise attack, after which retaliation would be impossible, thereby exposing the United States and its allies to Soviet blackmail.

A devastating surprise attack was considered possible because, with improved guidance systems, nuclear weapons were becoming more precise. Therefore, it was not inevitable that they would be used solely in "counter-value" strikes against easily targeted political and economic centres; instead it was just as likely that they would be used in "counterforce" strikes against military targets. A successful counterforce attack that rendered retaliation impossible—known as a "first strike"—would be strategically decisive. If, however, the attacked nation possessed sufficient forces to survive an attempted first strike with retaliatory weapons intact, then it would have what became known as a second-strike capability.

Counter-value and counter-force strikes

Other strategists, such as Thomas Schelling, warned that if both sides sought a first-strike capability, this could lead to an extremely unstable situation, especially during a period of high political tension when both were nervous as to the other's intentions. If it were feared that an enemy first strike was imminent, then there would be powerful pressures to attack first, and if the enemy recognized these pressures, then that would encourage him to get in his strike. Schelling described this as the "reciprocal fear of surprise attack."

On the other hand, if both sides were confident of their second-strike capabilities, then there would be considerable stability, as there would be no premium attached to unleashing nuclear hostilities. The benefits of a mutual second-strike capability led to the concept of arms control, by which potential adversaries would put less priority on simply lowering their force levels (as advocated by proponents of disarmament) and more on removing incentives to take the military initiative in the event of a severe crisis.

Mutual assured destruction. In the event, technological developments supported the second strike. Initially, long-range bombers had to be kept on continual alert to prevent them from being eliminated in a surprise attack. When ICBMs moved into full production in the early 1960s with such systems as the U.S. Titan and Minuteman I and the Soviet SS-7 and SS-8, they were placed in hardened underground silos, so that an unlikely direct hit would be required to destroy them. Even less vulnerable were submarine-launched ballistic missiles (SLBMs) such as the U.S. Polaris and the Soviet SS-N-5 and SS-N-6, which could take full advantage of the ocean expanses to hide from enemy attack.

The futility of anti-nuclear defense

Meanwhile, attempts to develop effective defenses against nuclear attack proved futile. The standards for antiaircraft defense in the nuclear age had to be much higher than for conventional air raids, since any penetration of the defensive screen would threaten the defender with catastrophe. Progress was made using surface-to-air missiles (SAMs), such as the U.S. Nike series, in developing defenses against bombers, but the move to ICBMs, with their minimal warning time before impact, appeared to render the defensive task hopeless. Then, during the 1960s, advances in radars and long-range SAMs promised a breakthrough in antiballistic missile defense, but by the early 1970s these in turn were countered by improvements in offensive missiles—notably multiple independently targeted reentry vehicles (MIRVs), which could swamp any defenses. (The first MIRVed ICBMs were the U.S. Minuteman III and the Soviet SS-17.)

Measures of civil defense, which could offer little protection to the civilian populace against nuclear explosions and, at best, only some chance of avoiding exposure to nuclear fallout, also appeared hopeless in the face of the overwhelming destructive power being accumulated by both sides.

By the mid-1960s fears had eased of a technological arms race that might encourage either side to unleash a surprise attack. For the foreseeable future each side could eliminate the other as a modern industrial state. Robert McNamara, the U.S. secretary of defense for much of that decade, argued that so long as the two superpowers had confidence in their capacity for mutual assured destruction—an ability to impose "unacceptable damage" defined as 25 percent of population and 50 percent of industry—the relationship between the two would be stable.

The need to maintain strategic stability influenced the strategic arms limitation talks (SALT), which began in 1969 and became the centrepiece of President Richard M. Nixon's policy of détente with the Soviet Union. In 1972, with the ABM Treaty, the two sides agreed to ban nationwide antiballistic missile systems, thereby confirming the primacy of the offense. Attempts to consolidate the strategic standoff with a treaty limiting offensive weapons proved more difficult. (In 1972 only an interim freeze had been agreed.) The second round of talks was guided mainly by the concept of parity, by which a broad equality in destructive power would be confirmed. However, the difficulty in comparing the two nuclear arsenals, which differed in important respects, resulted in long and complex negotiations. A treaty called SALT II was agreed on in June 1979, but by this time détente was in decline, and it was dealt a final blow with the Soviet intervention in Afghanistan at the end of that year. In addition, the strategic underpinnings of arms control had been undermined by a growing dissatisfaction in the United States with the principles of mutual assured destruction.

Alternatives to assured destruction. Critics found the condition of mutual assured destruction—which had become known by its acronym MAD—alarming. If MAD

failed to deter, then any war would soon lead to genocide. In addition, if the threat of retaliating with nuclear weapons was used to deter only nuclear attack, then the value of nuclear threats in deterring conventional aggression would be lost. In principle, this could undermine the commitments made to allies to use nuclear weapons on their behalf if they faced such aggression.

Particularly alarming was evidence that the nuclear strategy of the Soviet Union envisaged using nuclear weapons in a traditional military manner much as if they were conventional weapons—that is, at most to obtain a decisive military advantage in a conflict and at the very least to reduce the damage that an enemy might do to Soviet territory (if necessary, by launching preemptive strikes). During the negotiations that led to SALT II, critics also argued that the momentum behind the Soviet ICBM program, in combination with improved guidance systems that gave unprecedented accuracy to MIRVed missiles, had opened a "window of vulnerability" in the U.S. deterrent force. They expressed concern that the Soviet Union, by deploying the SS-17, SS-18, and SS-19 ICBMs, was building a force of such size and accuracy that just a portion of it could attack and destroy the U.S. Minuteman and Titan ICBM force without killing huge numbers of civilians. Although this would not be a true first strike, since U.S. bombers and submarines could retaliate, these latter delivery systems were not accurate enough to produce an equivalent counterforce attack against Soviet missile silos. Instead, the United States would be forced to escalate the war by retaliating against cities. This repugnant act would be of no strategic value, however, because the rest of the untouched Soviet missile force would then be used to wipe out U.S. cities. The United States, therefore, would have placed itself in a position in which it would have to choose between surrender and slaughter.

The realism of this scenario may be doubted, given that no attack against U.S. ICBMs would be accurate enough to avoid massive civilian destruction; therefore, the Soviet Union could be certain that the United States would feel little repugnance at retaliating against Soviet cities. Nonetheless, it was used to criticize SALT II, a complicated treaty that offered few means of verification and did little to interfere with the Soviet ICBM program. It was also used to argue for the development of U.S. ICBMs comparable to the Soviet systems.

Countervailing with limited strikes

The first formal break with assured destruction came in 1974, when Secretary of Defense James Schlesinger announced that future U.S. nuclear targeting would be geared to selective strikes and not just the sort of massive attacks suggested by the philosophy of mutual assured destruction. Although President Jimmy Carter's secretary of defense, Harold Brown, was skeptical that either side would actually find such sophisticated nuclear strikes possible, he accepted the need to develop a range of targeting options to convince the Soviet Union that it could not gain the upper hand by such methods. This was the main theme of the "countervailing" strategy announced in 1980.

Ronald W. Reagan came to office the next year with a much more radical critique of MAD, and his presidency was devoted to attempts to escape from its constraints. Initially, this took the form of a search for offensive nuclear operations that would enable the United States to "prevail" in a protracted war with the Soviet Union, rather than just countervail. It involved upgrading the old civil-defense systems and deploying the MX, an experimental ICBM originally designed to survive a first strike through some form of mobile deployment. Neither of these ideas was politically popular. In the end, civil defense was rejected as impossible, and the MX (now named Peacekeeper) was deployed in Minuteman silos and in only a fraction of the originally proposed numbers.

In March 1983 Reagan announced the start of a second search for a means to escape from MAD. This time it was for a defensive system that could intercept ballistic missiles. Reagan spoke of his preference for protecting lives rather than avenging them, and of the possibility of rendering nuclear weapons "impotent and obsolete," but the vision could not be turned into reality. Although the Strategic Defense Initiative, or SDI (which critics dubbed Star Wars, after a science fiction movie), was given a high priority and billions of dollars for research, the idea of protecting society as a whole from nuclear attack soon appeared hopelessly impractical, given the diverse means of delivering nuclear weapons. The main question became whether SDI could protect key political and military assets from attack, but even here some of the more futuristic ideas—such as using space-based lasers to destroy ballistic missiles just as they were launched—proved technically demanding and expensive. Political support waned.

Arms reduction

Meanwhile, Reagan had replaced talks on arms limitation with the Strategic Arms Reduction Talks (START). At first the Soviet Union argued that no progress on strategic arms control was possible so long as SDI was being pursued. Mikhail Gorbachev, who became the Soviet leader in 1985, offered his own vision of how to escape from assured destruction in a speech of January 1986, in which he set out a radical disarmament agenda leading toward a nuclear-free world by the end of the century. In October 1986, at a summit in Reykjavík, Ice., Reagan came close to embracing this vision, although no agreement was reached because he refused Gorbachev's demand to abandon SDI. Nevertheless, the concept of arms reduction had taken hold, and START proceeded with a new emphasis on deep cuts in nuclear arsenals.

The switch to arms reduction suggested that Reagan's critique of MAD had concluded with the view that, given the difficulties of designing and deploying both discriminating offensive options and effective ballistic missile defenses, it was better to do away with nuclear weapons altogether. This constituted a formidable challenge to the orthodox view that nuclear weapons exercised a stabilizing deterrence on international misbehaviour and were a reassurance to America's allies, who faced preponderant Soviet conventional forces. Reagan was prevailed upon to moderate his critique, but not before doubts had been created as to the strength of the U.S. commitment to guarantee the security of its allies with nuclear weapons.

On the other hand, the alacrity with which Gorbachev embraced complete nuclear disarmament reflected the greater freedom of maneuver available to any Soviet leader as well as the subordinate role of the Warsaw Pact allies. Whereas NATO's European members were anxious to lock the United States into their security arrangements for fear that they would be unable to stand alone, the Soviet Union had drawn its allies into a pact that met its own security requirements—that is, extending its form of government into eastern Europe and creating a buffer between it and the hostile capitalist forces of the West. Members of the Warsaw Pact might be beneficiaries of a Soviet nuclear guarantee, but there was no question of shared decision-making on nuclear matters. In fact, during the 1970s Soviet doctrine had appeared to have the goal of extracting the maximum regional benefit from its nuclear arsenal—vis-à-vis both western Europe and China—while maintaining Soviet territory as a sanctuary from nuclear devastation. Its priority in any nuclear conflict would have been to confine nuclear exchanges to central Europe, while showing a certain respect for U.S. territory as a sanctuary in the hope of reciprocal treatment by the United States. If escalation had appeared inevitable, however, or if the United States had appeared to be preparing a first strike, then Soviet doctrine would have called for a preemptive blow against the United States' long-range arsenal in an effort to reduce damage to the Soviet Union.

Orthodox Soviet doctrine

This approach was undermined by evidence that U.S. nuclear doctrine and deployment showed no respect for geographic sanctuary and by the Soviets' own recognition of the sheer difficulty of managing a nuclear exchange in such a way as to reduce the vulnerability of Soviet territory. Even before Gorbachev, there had been a discernible trend in military thinking toward emphasizing the opening conventional stage of a war and toward achieving victory within that stage. Gorbachev accelerated this trend. Because he was not prepared to allow overambitious nuclear doctrines to interfere with his objective of improving relations with the West, he was much more prepared than his predecessors to compromise in arms control negotiations. In addition, he was influenced by the April 1986

disaster at the Chernobyl nuclear power plant, which demonstrated that radioactive fallout had little respect for national boundaries.

Flexible response. This gave a new twist to the long-standing debate within NATO over nuclear deterrence. The United States' allies had already learned to live with unavoidable doubts over the quality of the U.S. nuclear guarantee of European security. These began to surface in the 1950s, after the Eisenhower administration had embraced nuclear deterrence and the allies had agreed that it was natural to rely on the most advanced weapons available—especially those in which the United States then enjoyed a clear superiority. The alternative course—relying on conventional forces—would have caused severe economic strains, and there was deep pessimism as to the possibility of ever matching Soviet conventional strength.

The conventional buildup set in motion under the Truman administration had one important requirement: that the Federal Republic of Germany be rearmed. This set in motion a sharp debate in Europe that was coloured by memories of the recent war, but in 1955 a formula was found in which West Germany rearmed but was permitted no chemical or nuclear weapons and was part of NATO's military command. In return, the West German government sought a commitment by its new allies to the concept of forward defense, in which any aggression would be rebuffed at the border between East and West Germany. (With its lack of depth and its concentration of population and industry close to the East, the Federal Republic had no wish for its allies to trade German soil for more time in responding to a Soviet attack.)

Once it was decided that NATO would not attempt to match Soviet conventional forces, then forward defense meant, in effect, that nuclear deterrence was linked to the inter-German border. European members of NATO had no qualms with this arrangement, because it saved them the expense of sustaining large-scale conventional forces, and they did not believe that the Soviet Union had any interest in invading western Europe that would be worth the slightest risk of nuclear war.

In the early 1960s the administration of President John F. Kennedy, which confronted the Soviet Union over the Berlin Wall and the Cuban missile crisis, did not take such a relaxed view of Soviet intentions. Given what it saw as the Soviet capacity for retaliation, the United States thought it unlikely that any president would use nuclear weapons first, and it was hard to see how a credible deterrent could be fashioned out of an incredible nuclear threat. At the very least, the United States insisted, NATO should raise the nuclear threshold, that is, the point at which nuclear weapons would be necessary to stave off conventional defeat. This would be accomplished by extra conventional forces. New analyses suggested that it would be easier than hitherto assumed because previous assessments had exaggerated the strength of the Warsaw Pact. In addition, the Soviet leader, Nikita Khrushchev, who was convinced that nuclear weapons made it unnecessary to maintain vast armies, was imposing major reductions on his generals at that time.

The threshold between conventional and nuclear war

European governments argued in response that conventional forces simply could not provide a sufficient deterrent. Since Soviet territory would not be vulnerable in a conventional war, the Kremlin might judge that the risks of conventional war were acceptable. And even if the Warsaw Pact were defeated, central Europe would still be devastated. Therefore, all war had to be deterred, not just nuclear war.

In 1967 a compromise was found in the doctrine of "flexible response." Under this compromise, the Europeans recognized the U.S. requirement for an extended conventional stage, so that the first shots across the Iron Curtain would not lead automatically to nuclear holocaust, and the United States accepted the need for a clear link between a land war in Europe and its own strategic nuclear arsenal.

Limited nuclear war. Flexible response did not prescribe a particular course of action; rather, it retained for NATO the possibility that it would be the first to use nuclear weapons and suggested that this initially would involve short-range, tactical weapons.

When tactical nuclear weapons such as the Honest John rocket were introduced into the NATO inventory during the 1950s, the U.S. Army had supposed that these could be considered quite separately from intercontinental strategic missiles. If anything, tactical nuclear weapons were closer to conventional weapons and were to be integrated with general-purpose forces. A number of strategic thinkers in the United States, including Henry Kissinger and Robert Osgood, hoped that, if the West could reinforce its military strength in this way, it would be possible to take on communists in limited nuclear wars without resort to incredible threats of massive retaliation.

However, once the widespread use of battlefield nuclear weapons by NATO was simulated in war games in the 1950s, it became apparent that they would result in such death and destruction that they could in no way be considered conventional. Also, as Warsaw Pact forces obtained comparable capabilities with such weapons as the SS-1 missile, any Western advantage seemed neutralized. Unless a retreating defender used nuclear weapons immediately, any later use could well be over his own territory and against a dispersed enemy. And, if tactical nuclear weapons were used to impose great costs on the enemy, there would be a risk that the conflict could soon escalate to strategic nuclear use. Limited nuclear war, therefore, appeared a contradiction in terms.

Linking U.S. to European security

European governments were still loath to dispense with the weapons. Although they could not be considered ordinary weapons of war, their close integration with conventional forces meant that they were more likely than U.S. strategic nuclear forces to get entangled in a land war in Europe. The idea was to use the risk of escalating to total nuclear war with the United States as a powerful deterrent effect on the Soviet Union's actions in Europe. According to this strategy, deterrence did not require a certainty that nuclear weapons would be used, but only a risk. The consequences of miscalculation were so horrendous that a government would dare not gamble. However, the United States, whose own security was now being linked to peace in Europe, was still more concerned that miscalculation might nonetheless take place.

Certainly, NATO's procedures for "going nuclear" were designed to reduce the risk of unauthorized use. But this created a tension between theory, which suggested that deterrence was served by the risk that a conflict might get out of control, and practice, which exhibited a determination not to lose control. The tension was reflected in discussions over how to replace the first generation of tactical nuclear weapons as they became obsolete in the 1970s. If the next generation were made smaller and more precise, then this would imply a readiness to use them to fight a nuclear war rather than simply deter. An apparent readiness to wage nuclear war was at the heart of a controversy over the "neutron bomb" (actually a thermonuclear missile warhead or artillery shell of enhanced radiation and reduced blast), which was criticized for blurring the boundary between conventional and nuclear weapons and thereby making it much easier to go nuclear.

Even greater controversy was generated by NATO's decision in 1979 to replace the Pershing IA, a medium-range ballistic missile, with two weapons that would constitute a more powerful intermediate nuclear force (INF): the Pershing II intermediate-range ballistic missile (IRBM) and the Tomahawk cruise missile. The origins of the program to modernize the INF lay in two western European concerns over the U.S. nuclear guarantee. The first concern resulted from the tendency of the United States in the Strategic Arms Limitation Talks to concentrate on achieving symmetry between the nuclear forces of the two superpowers, while paying little attention to the superiority, within the European theatre, of the Warsaw Pact in both nuclear and conventional weapons. Particularly worrisome was the Soviet SS-20, an IRBM that was first tested in 1974 and deployed in 1977. Although the SS-20 did not signal any shift in Soviet policy (U.S. military bases in Europe and the British, French, and Chinese nuclear forces had long been targeted), it was the first new missile designed for this purpose to have appeared in some time. In 1977 Chancellor Helmut Schmidt of West Germany

argued that NATO should not tolerate Soviet superiority in such weapons. This suggested that the imbalance should be dealt with either through arms control or by an equivalent Western effort to upgrade its own INF.

The second concern placed far less stress on the SS-20 and more on the requirements, within NATO's strategy of flexible response, to be able to strike Soviet territory with systems based in western Europe in the event of full-scale war on the Continent. This requirement existed irrespective of the new Soviet missiles, and it was becoming problematic because of the age of NATO's medium bombers and the lack of any U.S. intermediate-range land-based missile in Europe. A modernized INF made more sense than systems designed for battlefield use, because they posed a direct threat to the Soviet homeland and thus challenged Soviet ideas of confining any nuclear exchanges to NATO and Warsaw Pact countries, with superpower territory accorded sanctuary status.

However, large-scale protests sprang up in Europe and North America after the decision to modernize. Voicing a concern that a new arms race was getting under way in Europe, they took on special urgency following the Soviet invasion of Afghanistan (two weeks after NATO's decision on the INF), with the decline of arms control, and with the election of Ronald Reagan, who had a hawkish reputation, to the U.S. presidency. The strength of the protests encouraged NATO to moderate its policy. The rationale for modernizing the INF was switched from the requirements of flexible response to the more politically marketable aim of matching the deployment of the SS-20, and in November 1981, at the start of negotiations on this issue, Reagan offered to eliminate NATO's INF if all SS-20s were removed. This "zero option" was rejected by Leonid Brezhnev, and, despite warnings from the Soviet Union that deployment of a modernized INF would mean the end of negotiations, the first Tomahawk and Pershing II missiles were delivered in late 1983. Yury Andropov promptly broke off the INF talks, hoping to force a breach in the unanimity of the NATO allies, but, when the expected crisis failed to arise, Konstantin Chernenko agreed to resume negotiations. Soon afterward Gorbachev was in charge, and he decided that the zero option was in the Soviet interest: eliminating the INF would remove a direct threat to Soviet territory in return for removing a larger number of Soviet missiles that could strike only the allies of the United States. In December 1987, Gorbachev and Reagan signed the INF Treaty.

The INF Treaty: withdrawing nuclear deterrence

Although America's allies saw that the treaty had political benefits in improving East–West relations, some strategists worried that it sounded the death knell for nuclear deterrence. One response by NATO was to see whether it would be possible to build up other nuclear systems by way of compensation, but the difficulty here was that the improved political climate undermined public support for such moves. In West Germany the question of modernizing the short-range Lance missile was coloured by the direct and almost unique threat this weapon posed to German territory. There had always been the strongest official support for the traditional concept of nuclear deterrence in that country, but, with the political climate improving, West German politicians such as Chancellor Helmut Kohl came to argue that yet another nuclear modernization program would send the wrong signals to the East. They were also unhappy at the apparent readiness of the United States and Britain to retain Germany as a battlefield for short-range nuclear exchanges while securing the removal of intermediate- and long-range systems that threatened their own territories. The Soviet Union possessed large numbers of short-range missiles and had been modernizing them for a decade with such systems as the SS-21, but Gorbachev indicated a readiness to negotiate their complete elimination. British prime minister Margaret Thatcher and U.S. president George Bush insisted that this would be imprudent, and, following their lead, NATO agreed in 1989 to postpone modernizing the Lance in the hope that negotiations on conventional force reductions would reach a satisfactory conclusion and thus reduce the importance of nuclear weapons as a means of compensating for the Warsaw Pact's conventional superiority.

The Bush administration was more orthodox on nuclear matters than its predecessor, but Reagan's interest in a nuclear-free world—highlighted by SDI, the Reykjavík summit, and the INF Treaty—had already encouraged discussion among some Europeans of the possibility of a European defense community that would be less dependent upon the United States. In practice this would require the substitution of a French and British strategic nuclear guarantee for an American. Britain had always, officially at least, committed its strategic nuclear forces (which since the late 1960s had been SLBMs) to NATO. Britain's rationale for maintaining a national nuclear force involved a combination of the political influence that could be brought to bear on its allies, especially the United States, and a claim to be contributing to the overall deterrent posture. France, by contrast, had always had a much more nationalistic rationale, but after the 1970s, following the introduction of the Pluton short-range missile, which could only land on German territory, it was obliged to consider the role that its *force de frappe* might have in the defense of its allies. In any event, neither Britain nor France was eager to take over from the United States the broader deterrent role; nor were those who had previously sheltered under the U.S. umbrella interested in a European alternative.

Conventional strategy. The main consequence of the developing uncertainties surrounding nuclear deterrence was an increased interest in conventional strategy. For the first two decades of the nuclear age, there had been little interest in this area; given the conviction that any war between the great powers would soon go nuclear, there seemed to be little point in preparing for nonnuclear engagements.

Meanwhile, France and Britain fought a number of colonial wars, with France's struggles in Indochina and Algeria particularly protracted and bitter. During the 1960s the United States became steadily involved in Vietnam, in which a weak pro-Western government in the South faced an insurgency backed by a communist government in the North. After 1965 there was a substantial commitment of all elements of U.S. military power, excluding nuclear weapons but including a bombing campaign against the North.

Renewal of conventional strategy

Partly as a result of these conflicts, interest began to revive in the likely character of a conventional war involving the major powers. In the West this was also a result of the adoption of flexible response, which demanded greater attention to conventional warfare. In the East as well, to some extent because of the shift in NATO doctrine, conventional warfare grew in importance. There was some irony in this. Flexible response reflected NATO's concern over Soviet conventional superiority, yet, under Khrushchev, Soviet forces had been cut back dramatically on the assumption that any future war would go nuclear from the start. After Khrushchev was ousted in 1964, the Soviet Union began a major buildup of conventional forces, and in 1967 military exercises were held that indicated the expectation of a substantial conventional stage at the start of a future war. Besides the need to remain strong in relation to NATO, by the early 1970s the Soviet buildup reflected concern over a possible threat from China, which had become extremely hostile and was rapidly improving relations with the West. Again, there was irony in this development. China and the Soviet Union had finally split in 1963 over Khrushchev's readiness to deal with the West, over his unwillingness to back the Chinese nuclear program, and also over a long-standing border dispute between the two countries. Later, the years of the Cultural Revolution (1966–76) convinced Brezhnev that the Chinese were dangerous and unstable; clashes on the border in 1969 led to hints from Moscow that it might take action against China's fledgling nuclear capability.

In the late 1970s this Warsaw Pact buildup, coupled with Soviet-supported operations in such Third World countries as Vietnam, Angola, and Ethiopia, stimulated NATO to improve its capacity to resist an offensive and mobilize quickly. This was based on the fear that, without sufficient warning to get mobilization under way, the strength of Soviet frontline and follow-on forces could overwhelm

NATO's thin peacetime lines of defense. Growing doubts over the credibility and durability of nuclear deterrence also increased the importance of improving conventional forces.

Even with increased allocations to defense, NATO governments remained pessimistic about their ability to match Warsaw Pact forces. Although the total military power of NATO was much greater, geography favoured the Warsaw Pact, since reserves from the East could reach the front much more quickly than reserves from the United States, which would have to make a hazardous journey across the Atlantic Ocean. Against this pessimism it was noted that greater numbers were normally assumed to be required by the attacker (a ratio of three to one was often cited, although the critical factor was not the overall ratio but the strength of the offense at the point of attack). Technological advances were said further to favour the defense, in that extremely precise and comparatively simple guided weapons could be used to take on tanks and high-performance aircraft, the central actors in any offensive.

This optimism was questioned by other strategic analysts. They noted that the natural advantage accruing to the defense would do so only if the attacker had to force a way through well-prepared defensive positions, rather than simply outflanking them. The ability to impose attrition on the enemy would be reduced if the enemy was able to fight a war of maneuver, in which an immobile defense might find itself caught off balance. Moreover, in a war of maneuver, the potential benefits of simple air-defense and antitank systems would soon be qualified by the need to make them mobile, which would put them in need of protection as well.

The proponents of maneuver warfare warned that this was the type favoured by the Warsaw Pact. The Soviet Union preferred the offensive because it would make it possible to defeat the enemy quickly, before the full weight of its power could be brought to bear. Soviet doctrine during the 1970s suggested that a key aspect of this offensive would be the neutralization of NATO's nuclear assets by overrunning key installations, with a possible shift to a regional nuclear offensive when the right moment arrived. By the early 1980s doubts over whether a war would last long enough for the right moment ever to arrive, and whether nuclear exchanges could be limited geographically, encouraged a greater stress by the Soviet military on obtaining a victory in the conventional stage.

Follow-on forces attack

The maneuver school eventually encouraged a shift in NATO thinking toward more mobile operations, as well as a greater willingness to contemplate attacks into Warsaw Pact territory in an effort to reduce the momentum of a Pact offensive. The strategy of follow-on forces attack (FOFA), for example, envisaged the holding of a Pact offensive on the ground while attacking the Pact's follow-on forces in the rear with air strikes. Such aggressive defense was criticized by peace movements as being too provocative. Instead, they proposed non-provocative strategies based on "defensive defense," which would lack any capability to go on the offensive. These ideas proved difficult to turn into practice, as any sort of mobile force could move forward, and few armies would tolerate being deprived of their capacity to counterattack.

Meanwhile, in the Soviet Union concern over the burden of high defense expenditures, combined with an awareness that the arms buildup of the 1970s had triggered a counterresponse from NATO, encouraged "new thinking" that actually picked up on the ideas of defensive defense. These ideas were received by Soviet military commanders with as little enthusiasm as they were received in the West. Nevertheless, they influenced cuts in Soviet forces, announced by Gorbachev in December 1988, that eliminated some military units of a clearly offensive nature without depriving the Warsaw Pact of its offensive options. However modest in themselves, the cuts raised the prospect of an end to the role of these forces in sustaining the Soviet Union's dominance over eastern Europe. When coupled with domestic political reforms in the Soviet Union, Poland, and Hungary, they also signaled the eventual demise of the postwar alliance system. The prospect of a declining Soviet hegemony, however, was clouded by growing disorder within eastern Europe (and even in the Soviet Union itself) and also by the considerable strength that Soviet military forces would retain in spite of the planned cuts. Moreover, there was no reason to presume that a security system in which the two superpowers played a less overbearing role would be free of conflict. This provided an argument for a continuing role for nuclear deterrence: to warn against the dangers of allowing any conflict to get out of hand, if not to deal with a specific military threat. Nevertheless, political changes within the Soviet bloc presented a historic opportunity to heal East–West divisions and, with that, to reduce the need for substantial armed forces of any kind. (L.D.Fr.)

Tactics of land warfare

FUNDAMENTALS

Tactics, the art and science of fighting battles, is concerned with the approach to combat, the disposition of troops, the use made of various arms, and the execution of movements for attack or defense. Carl von Clausewitz defined tactics as "the formation and conduct of single combats in themselves" and strategy as "the combination of single combats with each other. . . . In other words, strategy forms the plan of the war, maps out the proposed course of the various campaigns which compose the war, and regulates the battles to be fought in each."

Relations between tactics and strategy

No matter how precisely the two terms may be defined, however, there is seldom any fixed line of demarcation. Strategy is frequently limited by the tactical potentialities of the forces involved, and tactics in turn depend on such strategic considerations as the planning and conduct of campaigns. Actually, the two overlap more often than not. In recognition of this fact, the term "grand tactics" gained currency during the Napoleonic era, and 20th-century military commentators have used the phrase "operational strategy" to describe concentrations of such major units as corps and divisions in preparation for battle.

The elements of tactics are dual in character. Attack and defense, cohesion and dispersion, protection and mobility, missile and shock effect, moral and material values—these elements have been changeless throughout the ages, regardless of variations in arms and communications. The barbarian's bow was as effective a missile weapon in its day as the machine gun in its, and elephants were esteemed for shock effect many centuries before the invention of the tank.

Factors determining tactics

Seldom do armies decide arbitrarily on the tactics they will adopt. The swing from one pole to another is nearly always influenced by geographic, climatic, and economic factors. Nomads of the plains become cavalry warriors from necessity rather than choice, just as mountaineers fight on foot. Primitive tribes and nations, unsure of their military skill, were drawn toward the cohesion of the human wall, or phalanx. As such peoples grew in strength and confidence they tended to adopt formations of dispersion, which are better suited to aggressions. Thus the Romans of the early city-state period fought in an eight-rank phalanx, but the Romans of the conquering republic developed that marvelously flexible tactical instrument known as the legion.

One of the most cherished delusions of the ancient world was the belief that a formation had a shock effect in proportion to its depth. But adding ranks to the phalanx did not add to the material might of the attack; it merely contributed to the moral impetus of the attackers. Shock effect is rarely the product of an actual physical impact. The shock is communicated to the soldier's mind rather than his body, and during the cavalry age he was morally defeated by his dread of being trampled, though horses seldom ran down foot soldiers who stood their ground.

Basic tactical principles are few and simple. The complexity and difficulty appear when such principles as surprise, secrecy, deception, security, and the gathering of intelligence are put into actual practice.

So obvious are the tactical advantages to be gained by deception and surprise that they are sought intuitively by primitives. Both principles are combined in that most primitive of tactics, the ambush. Even wild animals know

enough to hide while lying in wait to attack. And the ambush has remained one of the most effective tactics of warfare; it was used by both British and German armoured columns in the complicated North African desert campaigns of World War II and by the Viet Cong during the Vietnam War. Surprise depends simply on the degree of secrecy with which a general masks his intentions and capabilities.

So essential is intelligence to the tactician that a good spy is sometimes worth more to an army than a good general. Reconnaissance patrols, prisoner interrogations, and observation or listening posts also play an important part in arriving at an estimate of an enemy's intentions and capabilities.

Security, like secrecy, has been maintained in all military ages by unremitting vigilance. Although the firepower of missile weapons has often been given the sole credit for the prolonged stalemate in World War I, violation of age-old principles of tactics cannot be overlooked. Several promising Allied offensives met failure because surprise and deception were sacrificed to make possible the tremendous artillery "preparations" that had no decisive effect on an enemy burrowed deep into the earth. Such large quantities of shells were required for these bombardments that the area and extent of the impending offensive were revealed to the enemy by aerial photographs of ammunition dumps.

Tactical formation. Far back in man's past his primitive ancestors perceived that superior numbers did not always make for superior force. A great deal depended on the tactical direction and speed with which this force was applied, and the solution was to group fighters so that they could give one another the most support while doing the most harm to the enemy. This was the beginning of tactical wisdom.

The line and column

The two basic formations are the line and the column, which differ only by 90° of the compass: a column of men in four files, facing north, has only to execute a "right face" to become a line of four ranks, facing east.

Neither of these formations is useful for all tactical purposes. Although the column is best adapted to marching and maneuver, its narrow front and vulnerable flanks are liabilities in combat. The line lacks depth and mobility, but it is more suitable for deployment on the battlefield. Mention should also be made of two other formations that appear less frequently in the chronicles of war—the hollow square and its counterpart, the 360° perimeter. Some stout stands have been made by men facing outward in all directions, but the defect of such formations is that ability to maneuver is sacrificed.

Even in the remote past it was evident that forming warriors into a line or column was not enough. Some attention also had to be paid to problems of deploying them so as to make the best of both shock and missile weapons. When men learned to fight on horseback, a cavalry arm came into being. The "eyes" of the primitive army, when used for scouting and reconnaissance, horsemen were also effective for pursuing a broken foe. A crude artillery arm appeared as early as the 8th century BC, when the Old Testament dates engines of war. Not even the tactical ancestor of the tank was lacking. Armoured chariots manned by archers and spearmen led the infantry in battles on the plains of Mesopotamia many centuries before the Christian era.

The tactics of attack. Descriptions of battles often give the impression that victories are won entirely by physical pressure, as measured by the casualties inflicted. It has been the experience of history, on the contrary, that moral or psychological forces play a much larger part in the outcome than material means. An army's combative spirit is not often broken because the ranks have been so thinned that the survivors are crushed by sheer human tonnage. They are beaten because of their conviction that further resistance is hopeless, and it is the business of the tactician to convince rather than crush his opponents.

The test may be found in the fact that casualties of 30 percent are usually the most an army can endure without losing combative spirit. Yet seven men out of 10 are still able to fight, and it may be that they outnumber the entire force of adversaries who accepted battle with a numerical disadvantage. History records instances of losers being routed while still possessing a 2-to-1 numerical superiority. They were not crushed; they were simply convinced.

The methods of creating a conviction of defeat may be reduced to a few plans of attack that repeat themselves throughout history. Most simple of all is the plain frontal assault all along the line. Justifiable and even desirable under some circumstances, it is as a rule the resort of unimaginative generalship. As for the direct rear attack, it is so rare as to be associated only with ambush or pursuit.

Effectiveness of flank and rear attacks

In an early age it was discovered that armies, like the individuals composing them, dread an attack from the flank or rear. An assault from one side or another is not enough, since an adversary can face in the new direction, whereupon the effort becomes a mere frontal attack. In order to deliver a true flank attack, part of the offensive force must fix the enemy's front, perhaps even defensively, while the assault force swings around to strike simultaneously from the side. If this force can make a wide enough swing to threaten an opponent's rear, the attack becomes a single envelopment. And in instances when the rear is assailed from both sides with concentric thrusts, a double envelopment is the result. Neither a flank attack nor envelopment may be possible when a general secures his flanks with natural obstacles. In that case it is sometimes possible to create two new flanks by rupturing an adversary's front and driving through the gap to strike at his rear. That is the battle of penetration, and it has some distinguished victories to its credit.

In any event, the effectiveness of the attack is in direct proportion to the conviction of defeat it can instill. As evidence, history is filled with occasions when the casualties of a beaten army were the result rather than the cause of defeat.

Offensive and defensive tactics. The relative merits of the offensive and defensive have been subjects of argument in all military ages. From one century to another each has dominated in turn, the new cycle being brought about by the introduction of new weapons or tactics.

The attacker is usually able to choose the time of the assault and sometimes the place as well. If there are several points where the blow might fall, he has an opportunity to utilize the principles of deception and surprise against an opponent who must prepare as best he can to meet every alternative.

The defender often has the advantage of choosing and fortifying his own ground, but at least temporarily he must concede the initiative to the enemy. In most instances, of course, he expects to regain it by means of a decisive counterattack after exhausting his adversary or placing him in an unfavourable position. Passive defenses are seldom the resort of a general who retains freedom of action, though there are exceptional occasions when the object is to inflict crippling casualties on a rash enemy.

Cordon defenses are put into effect along a frontier on which troops can be thinly dispersed and quickly reinforced by reserves in the event of trouble. The mobile, or elastic, defense is based on the theory that if the enemy makes a penetration his flanks will be open to counterattack. The so-called defensive-offensive action merely implies that the defender has initially planned to launch a counterstroke when he has an opportunity.

The delaying action

Breaking off contact with the enemy is a delicate operation calling for that specialized form of the defense known as a delaying action. The rearguard force holds just long enough to compel the pursuer to deploy, then pulls back to the next prepared position before becoming seriously engaged. In a two-line delaying action the first and second lines retire through each other in turn until the main body of troops is out of danger.

The ideal instrument would be an army equally formidable in offensive and defensive tactics, but there have been few such in military chronicles. Probably the most effective combination throughout history has been the strategic offensive and tactical defensive. Invading armies are often able to choose their ground and await an attack, since the enemy is obliged for the sake of national morale to defend the homeland. No better example of

this compulsion could be found than the situation faced by Roman generals after Hannibal's crossing of the Alps into Italy. One after another, they met him in headlong attack, only to go down to crushing defeat. Finally the proud republic was reduced to avoiding battle in order to recuperate.

Until the 19th century, offensive and defensive cycles were of long duration. Improvements in weapons have come so rapidly since the Napoleonic era that the swing from one to another is a matter of only a few years. Generals of all armies were confident at the outbreak of World War I that the offensive power of the arms of 1914 would prevail in a few months. The machine gun had the leading part in turning the conflict into a stalemate, however, and not until the last few months of the war did the tank provide the victors with an offensive weapon capable of breaking the deadlock. Even so, a school of military thought held after the war that the defensive remained powerful enough to curb an aggressor in the next conflict. That illusion was soon dispelled when the German armies of 1940 drove on to a succession of early victories.

HISTORICAL DEVELOPMENT

In fighting the unending war for survival, prehistoric man employed tactics that from the first involved the three elements of firepower, movement, and protection. He employed missile weapons (stones, wooden spears, darts) and hand or side weapons (stone and flint knives and hand axes); he used terrain and flame as protection against both animals and the elements, and he also used fire as a weapon to force beasts from the protection of caves. His first military tactics were decidedly primitive—single encounters won by the strongest opponent. But, as walled towns sprang up and armaments improved, as the wheel was invented and the horse domesticated, battle grew more sophisticated. Eventually tribes organized specific armies, trained and disciplined them, and developed metal weapons and armour.

The chariot. About 1800 BC the Hyksos, migrating south from Asia, introduced the horse-drawn chariot as a tactical weapon in a series of battles that wrested most of present Syria and Israel from the Egyptians. These vehicles were small, usually holding one or two men armed with the bow. The front was protected. In battle, chariots preceded units of foot soldiers, a tactic that thoroughly disrupted the Egyptian formations.

Egypt. The Egyptians quickly adopted the new tactics. Ahmose I and his successors developed a wider, lighter chariot and also introduced iron wheels with spokes, a notable improvement over the solid wooden wheels. In building a new empire, Thutmose I employed a navy to establish forward supply bases that supported armies spearheaded by chariots and using two types of foot troops: heavy infantrymen, who carried spears and long shields, and light infantrymen, armed with spears, bows, battle-axes, and slings. The light, broad chariots generally carried specially trained bowmen. Thutmose's armies were probably formidable enough, but his continued prosperity seems to have hinged more upon a wise overall strategy, particularly in the political sphere, than upon any special ground tactics which in any event were doomed by the rise of the Assyrians.

Assyria. The Assyrians, who later conquered Egypt and by 671 BC dominated the eastern Mediterranean, made war their chief business. Employing weapons made of iron (taken from the Hittites), the Assyrians in open battle relied on both cavalry and heavy chariots backed by archers, spearmen, and shield bearers. They developed a chariot heavier than the Egyptians' and used protective felt coverings for the horses. They deployed both light archers, who often fought as skirmishers without protection, and heavy archers, who alternated with spearmen in the use of shield bearers. They began to use mounted bowmen and swordsmen independently, the first cavalry force. They also relied on terror for tactical purposes and were famed and feared for hard, fast actions.

To take cities, the Assyrians employed catapults and battering rams. The latter grew very sophisticated, at times reaching 20 stories in height and manned by hundreds of soldiers; Josephus speaks of one that was pulled by 300 oxen and manhandled into position by 1,500 soldiers.

Persia. Assyrian tactics were adopted in part by the Persians who, commencing in about 553 BC, rose to power under Cyrus II. According to Herodotus' chronicle of these early wars, in the decisive Battle of Thymbra in 546 BC, Cyrus' chariots were backed by infantry followed by javelin-equipped soldiers and archers with cavalry on the flanks and in the rear. Cyrus enjoyed a reputation for cunning. Frustrated in his later siege of Babylon, he posted armies "at the point where the Euphrates river enters the city, and another body at the back of the place where it issues forth." He then took a body of troops and diverted the Euphrates by a canal into a marsh; his troops entered the city through the riverbeds, taking the Babylonians by surprise and winning the city.

Expansion and consolidation continued under Cyrus' successors, most notably Darius I (reigned 522–486 BC) and Xerxes (486–465 BC). Intent on protecting the empire in the east and west, Darius placed more emphasis on archers than on infantry and supported them with heavy chariots. To increase the shock value Darius attached scythes to the blades and wheels. A bit later, three curved blades were fastened to the pole, one straight forward, the others fanning out in front of the horses. These added to the terror of the attack.

Perhaps Darius' major contribution to military science was the reorganization of his ground force into divisions—each, of 10,000 men, organized into 10 battalions, each battalion into 10 companies, and each company into 10 sections. James H. Breasted described this organization as "one of the most remarkable achievements in the history of the ancient Orient, if not of the world." (*The Conquest of Civilization,* 1926.)

The infantry phalanx. *Greece.* The Greek city-states militarily emphasized the infantry phalanx, considered by some theorists to be the germ of all future European military development. Citizen-soldiers fought in an eight-rank phalanx of heavy infantry supported by contingents of light foot. Although Homer's *Iliad* mentions horse and chariots a millennium before the Christian era, cavalry was not suited to a mountainous peninsula, and the incessant strife of the city-states was usually decided by the clash of opposing hedges of infantry spears.

Greek warfare was distinguished more for brawn and fortitude than for the intellectual brilliance that made for greatness in other aspects of Greek life. Even the double envelopment that won the Battle of Marathon for the Greeks in 490 BC seems to have been the product of a fortuitous event rather than generalship. Attacking at a fast pace, before the Persian superiority in cavalry and missile weapons could become effective, the Greek hoplites (heavily armed infantrymen) of both wings hurled back the lightly armed invaders in their path. Meanwhile, their own hard-pressed centre gave ground before the enemy advance until it seemed that defeat was imminent. As it turned out, this withdrawal created the opportunity for the heavy infantry of the Greek wings to converge on the flanks and rear of the Persian host. Thus the battle ended with the demoralization and rout of an army at least twice that of the Greeks. (See Figure 1.)

The Battle of Marathon

Defending their homeland against the Persians, the Greeks learned to place increased reliance on cavalry and missile weapons—mercenary horsemen from the plains of Thessaly and bowmen from Crete. The first Greek to distinguish himself as a tactician was Epaminondas in 371 BC. Facing a Spartan army of 10,000 men with a force of 6,000 that included undependable allies, the Theban commander drew up a strange oblique order of battle by advancing his left wing and reinforcing it with his best troops to a depth of 48 ranks. On the other wing he placed his doubtful allies. The Battle of Leuctra opened with a cavalry clash won by the better-trained Theban horse. Then Epaminondas bore down with the infantry of his heavy left wing on the right of the Spartan phalanx, and the factor of surprise probably had more to do with the outcome than physical shock. At any rate, the Spartan right recoiled in disorder on its own centre, and a phalanx thrown into confusion was little better than a mob. The

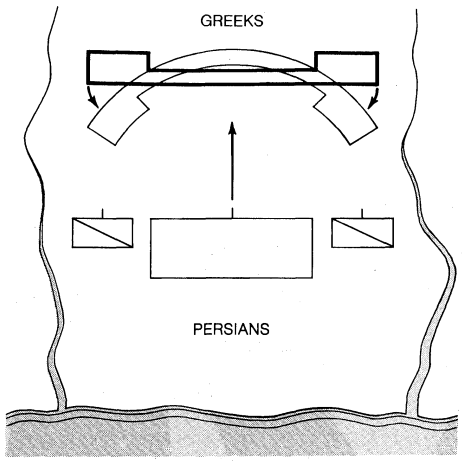


Figure 1: The Battle of Marathon (490 BC), showing tactics of concentric attack.

muscular tactics of Sparta were of no avail to demoralized men too crowded together to use hand weapons, and the Thebans gained a one-sided victory.

Macedonia. Epaminondas was mortally wounded while winning a second battle with his oblique order, and it remained for a barbarian prince to assemble the tactics of the disputatious Greeks into a system capable of conquest. Young Philip II of Macedon (382–336 BC) studied Epaminondas' methods while spending three years in Thebes as a hostage. He returned to his northern kingdom to organize the first army of ancient history to combine effectively such arms as infantry, cavalry, and artillery.

Philip's use of the phalanx

The new tactical system was built about the phalanx. Philip strengthened the mobile human fortress by doubling the eight-rank formation of the Greeks and arming his warriors with a longer and heavier spear. He perceived that the two chief weaknesses of the phalanx were the vulnerability of its flanks and the inability to maneuver or even retain its solidarity on rough ground. As a tactical remedy for both defects, he organized and trained a light infantry such as the ancient world had never known before—a disciplined force of archers, slingers, and javelin men. Their function was to act as tactical hinges on both sides of the phalanx, lending mobility as well as protection. Since the young nobles of Macedonia preferred to fight on horseback, Philip had no difficulty in organizing a heavy cavalry arm far superior to the establishments of the Greek city-states. Squadrons of light horse were also employed for scouting and skirmishing.

That the army had an efficient baggage train, medical corps, and engineering section is evident from its accomplishments. In his development of "artillery," moreover, the Macedonian king went far beyond his predecessors, who had found war engines useful only for siegecraft. He originated the idea of carrying only the essential parts on the march and depending on his engineers to cut timbers for the framework whenever needed.

On the battlefield the full phalanx of about 8,000 heavily armoured spearmen was doubtless a fearsome spectacle as it advanced with menacing steel points. Several such bodies of troops might be employed as the central bulwark of the battle line, the pivot either of offense or defense. On the wings were the heavy cavalry squadrons trained for shock effect. Between the phalangial centre and cavalry wings, the tactical hinges of light foot supplied the flexibility for maneuver.

Alexander at the Battle of Arbela

This was the army that Philip led to the conquest of Greece and his son Alexander (356–323 BC) made the instrument of world conquest. The Greco-Macedonian tactical system met its greatest test on the field of Arbela, or Gaugamela, in 331 BC. Darius III, the Persian monarch, had a numerical advantage of perhaps 5 to 1 over Alexander's 7,000 cavalry and 40,000 foot. The action took place near the ruins of ancient Nineveh on a plain suited to the Persian cavalry and chariots. Despite the odds against him, Alexander rejected the plan of a night attack urged by his nobles. He opened the battle by seizing the initiative with a rapid advance of his right wing, followed in echelon by the rest of the army. The local superiority in numbers enabled him to drive so deeply between the enemy's left and centre as to reach the rear (see Figure 2). A more perfect example of the battle of penetration can scarcely be found in the pages of history, for the Persian host was routed while still possessing a great numerical superiority over Alexander's army.

Alexander's campaigns took him as far as modern India and Afghanistan during the six years after his conquest of Persia. Among his most troublesome foes were the Scythian horsemen who preyed upon his flanks and rear, seeking to cut off detachments. That the nomadic riders of the Asian plains could be formidable opponents was convincingly demonstrated in later centuries when no fewer than four Roman armies were lured to their destruction.

The Scythians and Parthians were made natural warriors by a harsh environment. Shunning action at close quarters, fleeing when pressed, then returning to the attack in swirling envelopments, they could discharge arrows while riding at top speed.

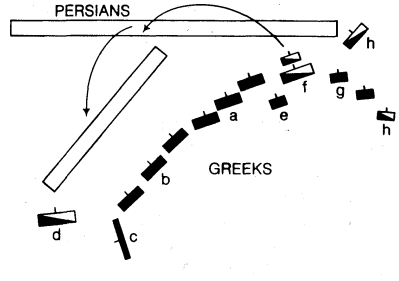
Alexander refused to be drawn into extended pursuits by feigned flights. Noting that a lack of discipline among the Scythians made them reckless at times, he trained his light infantry to launch surprise counterattacks and beat the nomads at their own game by cutting off detachments. Once the lightly armed Scythians were at close quarters, they were no match for Alexander's cavalry.

The Roman legion. The conqueror's death was followed by the rapid crumbling of the great empire he had founded during his lifetime. But while Alexander's former generals disputed for mastery in the east, a young and vigorous war power was rising in the west. Fighting in a simple phalanx of heavy infantry at first, the Romans learned by defeat that the human wall is not suited to battle on rough ground.

Maniple, cohort, and legion

The early republic. The new tactical system that gradually evolved was the legion. Under the republic it consisted of small and supple infantry units known as maniples, each numbering 120 men in 12 files and 10 ranks. In a typical battle order the first two lines, drawn up in chessboard array, were composed of 10 maniples each. An interval equal to a maniple's front of 60 feet separated the units, enabling the second line to merge with the first or the first to fall back into the second. The advantages of mobility could thus be exchanged for those of solidarity, since the two lines combined had a depth of 10 ranks and a front of 1,200 feet.

In the third line 10 units of light foot were alternated with a like number of reserve units, the former numbering 120 men each and the latter 60. The three lines were 250 feet apart, and from front to rear one maniple of each classification formed a cohort. This was the Roman equivalent of a battalion, numbering 420 men. Ten cohorts made up the infantry strength of a legion, but 20 were customarily combined with a small cavalry force and other supporting units into a self-sustaining little army of about 10,000 men.



a RIGHT OF PHALANX
b LEFT OF PHALANX
c LIGHT INFANTRY
d CAVALRY
e SHIELD BEARERS
f HEAVY CAVALRY
g LIGHT INFANTRY
h LIGHT CAVALRY

Figure 2: The Battle of Arbela, or Gaugamela (331 BC), showing tactics of rear attack.

The two famous Roman infantry weapons were the *pilum* and *gladius*—the first a seven-foot javelin used for both throwing and thrusting and the latter a 20-inch cut-and-thrust sword. For protection the legionary had a metal helmet, cuirass, and convex shield, but aggressive action was deemed a Roman's best defense. The first line attacked on the double, hurling javelins and diving in with swords before the enemy had time to recover. Then came the maniples of the second line, and only a resolute foe could rally from the two successive shocks.

When Hannibal descended from the Alps in 218 BC with an army of Iberian, Numidian, and Gallic mercenaries, the Roman Republic paid the penalty of placing too much reliance on heavy infantry and failing to develop a cavalry arm and missile weapons. Two Roman armies, destroyed in the early encounters, were outmaneuvered and cut to pieces by the Carthaginian horse. Finally, the Romans incurred a disaster that has made the name of Cannae a symbol ever since of the battle of annihilation and double envelopment (see Figure 3). Hannibal provoked the action, though much outnumbered by a Roman army of 85,000 men. Adopting a convex formation of his foot in the centre, he gradually withdrew it in a feigned retreat until the Roman infantry rushed in recklessly, raising the victory shout. When the converging legionaries were too crowded together to use their weapons, the Carthaginian closed the trap by assailing the flanks and rear of the Roman infantry with his cavalry after driving the inferior enemy horse from the field. The result was a slaughter of helpless men that lasted until only about 15,000 scattered fugitives survived Rome's greatest disaster.

Tactic of double envelopment

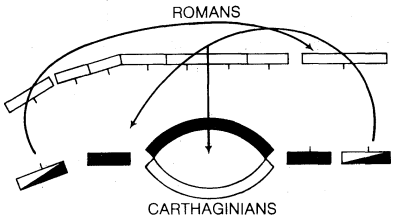


Figure 3: The Battle of Cannae (216 BC), showing tactics of double envelopment.

The late republic and empire. The patriotic citizen-soldier of the republic's early years was gradually replaced by a professional soldiery serving for pay and donatives. The cohort of an average field strength of 360 men gradually replaced the maniple as the tactical unit of increasingly larger Roman armies. Ten cohorts made up a legion in the operations of Sulla and Caesar. The 3,600 heavy infantry were supported by enough horse and light infantry to bring the numbers up to 6,000 in all.

The maniple survived only as a unit of a cohort of eight or 10 ranks. Four cohorts in the first line and three each in the second and third lines were drawn up in chessboard array, with intervals between units so that one line could merge with another (see Figure 4). Seven legions in three lines, comprising about 25,000 heavy infantry, occupied a mile and a half of front. This was an average of about 11 men to a metre of front, as compared with 28 in Alexander's army.

The practical good sense that distinguished Roman tactics was at its best in field fortification. No greater demonstration was ever seen in the ancient world than the 50 miles of works with which Caesar gained the submission of Vercingetorix and 80,000 Gauls at Alesia. The Romans began by surrounding the strong natural defenses with a system of walls and trenches. When Gallic allies marched to the rescue of his opponent, Caesar surrounded his own army with works of circumvallation and beat off all attacks. Finally the relieving army disintegrated and Vercingetorix was starved into surrender.

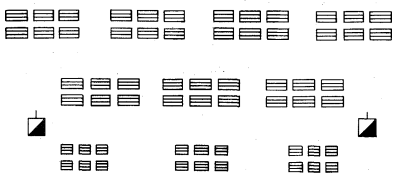


Figure 4: *Legion in battle order.*
Four cohorts in the first line and three each in the second and third lines with cavalry.

Byzantine cavalry. After the fall of Rome in AD 476, the Eastern or Byzantine Empire managed to survive by means of bribing the most formidable barbarians and inducing them to move on to the occupation of Italy. The respite enabled Constantinople to develop during the reign of Justinian (AD 527–565) a cavalry tactical system that won victories on three continents.

This system was built about the horse archer. The invention of the stirrup had given an impetus to cavalry warfare, and the historian Procopius has described the tactics of the mounted mercenaries whom Belisarius led to conquests in Persia, North Africa, and Italy.

The mounted archer

> But the horsemen of the present time go into battle wearing corselets and fitted out with greaves which extend up to the knee. From the right side hang their arrows, from the other the sword. And there are some who have a spear also attached to them and, at the shoulders, a sort of small shield without a grip, such as to cover the region of the face and neck. They are expert horsemen, and are able without difficulty to direct their bows to either side while riding at full speed, and to shoot an opponent whether in pursuit or flight. They draw the bowstring along by the forehead about opposite the right ear, thereby charging an arrow with such impetus as to kill whoever stands in the way, shield and corselet alike having no power to check its force. Still, there are those who take into consideration none of these things, who reverence and worship the ancient times, and give no credit to modern improvements. (Procopius, *Works,* trans. by H.B. Dewing [Loeb Classical Library], vol. i, p. 7; Harvard University Press, Cambridge, Mass., 1914.)

Byzantine tactics were eminently practical, the spirit of chivalry being unknown in an empire constantly fighting for existence. Greek in speech as well as tradition, Constantinople reduced war to a science with a series of treatises and manuals. The *Tactica* of Emperor Leo the Wise, written about 900, urged that operations should be varied to take advantage of each foe's peculiar weaknesses. Fraud and feigned flights were suggested without apology. If a foe could be bribed more cheaply than defeated, Eastern Roman emperors deemed it no disgrace to pay tribute.

Since most of Constantinople's foes were mounted, the tactical system placed the emphasis on heavy cavalry, both archers and lancers. But even though the infantry ranked as a secondary army, it was well trained and combined with mounted troops in proportions suited to tactical demands. Squadrons of light horse were employed for scouting and reconnaissance.

Byzantine armies stood between the West and Asian invaders at a time when the impoverished Europe of the Dark Ages would probably have been overrun. Charles Martel and his Franks had difficulty in repulsing a single large-scale Moorish plunder raid on the field of Tours in 732. Yet Constantinople held its own in an 83-year war with the Saracens when Islām was at the height of its conquering momentum.

At the outset the Muslims were cavalry warriors using much the same tactics as the Parthians and Scythians. They differed from other plains horsemen in their ability to learn from their foes. After experience taught the need, the Saracens organized a heavy cavalry of their own and adopted many of the East Roman tactics. Most astonishing of all, these former desert dwellers took to the sea and built ships large enough to launch two great land and water invasions that were repulsed outside the walls of Constantinople. Only the secret Byzantine weapon known as Greek fire (an inflammable composition, the ingredients of which are unknown) saved the imperial city in 673.

The Franks and Vikings. It was fortunate that Europe was spared a severe test, for as late as the 8th century the Franks could raise only untrained and largely unarmoured men using the most primitive tactics. The low esteem in which the Franks were held by skilled warriors is evident from Leo's *Tactica.* Admitting that the headlong charge of mounted Franks with lance and broadsword could be formidable, he nevertheless advised that one could

take advantage of their indiscipline and disorder; whether fighting on foot or on horseback, they charge in dense, unwieldy masses, which cannot manoeuvre, because they have neither organisation nor drill. . . . Hence they readily fall into confusion if suddenly attacked in flank and rear—a thing easy to accomplish, as they are utterly careless and neglect the use of pickets and vedettes and the proper surveying of the countryside. (C.W.C. Oman, *A History of the Art of War in the Middle Ages,* vol. ii, p. 204.)

Charlemagne made an unceasing effort to create a disciplined army of mailed cavalry out of the Frankish host of unarmoured foot, but despite his reforms Europe was long helpless against the plunder raids of the Vikings.

Amphibious warfare of the Vikings

Their single-decked ships with curved prows and shield-hung bulwarks were fast and seaworthy, and experience had taught them effective tactics of seizing a beachhead after pushing up such rivers as the Loire and Meuse. The Vikings were formidable fighters with ax and sword, and their forays grew so bold during the last half of the 9th century that they sacked London and laid siege to Paris.

There was not time to summon large armed forces from a distance against these fleet raiders. Each district had to fend for itself, and the feudal system burgeoned as Europeans paid military duty to the local lord upon whom they relied for protection. Cavalry tactics were also given another impetus, for only mailed horsemen could cope with the Vikings. Seldom has the social progress of an age been so largely shaped by tactical factors.

The medieval knight. In 1066, at the Battle of Hastings, the last infantry host of western Europe went down to defeat in England before the descendants of Norsemen who had settled in France and learned to fight on horseback. The mounted and armoured men-at-arms had become virtually the tactical unit of artless battles that had as their first object the seizing of opponents to hold for ransom. Serfs who fought on foot were held in such low esteem that no count was taken of their numbers or losses.

In contrast to the supremacy of the offensive in Asia, the defensive was all-powerful in medieval Europe. Concentric fortification added to the difficulties of besiegers by making it necessary to attack walls within walls as well as outer wards and foreworks, while the crossbow provided the garrison with a means of enfilade fire from rounded keeps. Meanwhile, the gradual transition from chain to plate armour made the mailed knight and his barded charger more invulnerable to the weapons of the day.

The wretched communications of medieval Europe were a deterrent to maneuver, even if the unruly armed forces of the 13th century had been disciplined enough to attempt it. It was seldom possible to force a battle on an unwilling opponent, and methods of supply were so primitive that an army often had to abandon a half-finished campaign after eating a province bare.

That the tactical superiority of the knight owed a great deal to his monopoly of the prerogative of bearing arms was demonstrated by despised Flemish burghers at the Battle of Courtrai in 1302. French men-at-arms, putting down a revolt, launched their usual thundering charge, only to bog down in disorder upon reaching marshy ground they had not troubled to reconnoitre. The Flemish foot clubbed the knights out of the saddle and dispatched them as they lay helpless in the mud, too burdened with armour to defend themselves.

The cavalry of Genghis Khan

The Mongols. Except for a few crossbowmen, Europe's armies had no infantry worthy of the name. A similar situation prevailed in Asia, where the Mongols led by Genghis Khan built up a formidable cavalry system. The bow, lance, and scimitar were the weapons of fierce horsemen who used two varieties of bows and three "calibres" of arrows for various tactical purposes. The Great Wall of China, with its 1,500 miles of earth-filled brick or granite, proved to be no serious obstacle to invaders who swept like a pestilence from the Yangtze to the Vistula during the 13th century without incurring a single major defeat. Whether they could have continued to win victories in the mountains and forests of western Europe is open to speculation, for they returned to Asia after defeating the feudal horsemen of Hungary and Poland.

The late Middle Ages in Europe. *Longbow and halberd.* The use of gunpowder in the 14th century made it inevitable that the supremacy of the man of iron in his house of stone could not be long maintained. But the first handguns were not as effective as either the crossbow or English longbow, and it was actually the rise of disciplined forces of yeomen using infantry tactics that restored the balance between the offensive and defensive in medieval Europe.

Although a few weapons using gunpowder were present on the field of Crécy in 1346, the outcome owed more to the discipline and tactics of an English army in which every man received a fixed and regular wage. For it was truly a national army that Edward III led to France, and the French knights would probably have been defeated even if the invaders had been armed with a less effective weapon than the six-foot longbow. Its superiority over the crossbow of the Genoese mercenaries was proved in a preliminary exchange of missiles. Edward had placed his dismounted men-at-arms in the centre as spearmen, and on both projecting wings the lightly armoured archers stood behind ditches and pointed stakes. They poured deadly flights of arrows into the front and flanks of the advancing French knights until 16 successive charges had been repulsed with heavy losses.

The Swiss pike phalanx

Meanwhile, another infantry tactical system was winning victories in central Europe at the expense of the Austrian knights. Swiss burghers and peasants, fighting for their liberties, depended on a medieval phalanx composed of spearmen and warriors armed with the halberd—an eight-foot shaft that had the blade of an ax, the point of a spear, and a hook for pulling a rider out of the saddle.

The Swiss phalanx of the 15th century consisted of three-fifths halberds, one-fifth spears, and one-fifth missile weapons, chiefly crossbows, with handguns slowly gaining in favour. In the interests of mobility the men of the Swiss cantons discarded all armour except the helmet and cuirass. Marching 30 miles a day to meet the enemy, they attacked with a celerity and discipline that were disconcerting to medieval adversaries. If hard-pressed, the Swiss "formed the hedgehog" by facing outward in all directions as men from the rear ranks filled any gap in front.

Early victories of gunpowder. After beating mailed horsemen who had resisted tactical change, the English and Swiss in their turn were defeated because they lagged behind the military progress of the age. It had taken weapons of gunpowder a century and a half to have much influence on tactics, but at the siege of Orléans in 1428–29 both sides used cannon firing stone balls. The French guns were better served than the mortars and bombards brought by the English, who met defeat at the Battle of Patay after abandoning the siege.

Charles VII of France (1403–61) raised the first standing army of the Middle Ages—9,000 permanent troops, paid and armed by the king. This force gradually won back the French territory conquered by the English in the Hundred Years' War, and at the Battle of Formigny in 1450 the invaders were decisively beaten after making their customary defensive stand with the longbowmen on both wings. Instead of launching their usual charge, the outnumbered French brought up culverins, or medieval fieldpieces, and cut the English archers to pieces with enfilade fire. The Swiss phalanx was granted a longer period of grace, but there was no hope and eventually it too was to pay with defeat for clinging obstinately to outworn tactics.

Use of artillery by the Hussites

Of all the early exponents of gunpowder weapons, perhaps the most remarkable was the army of Bohemian peasants commanded by Jan Žižka (*c.* 1376–1424). Several crusades were proclaimed by Pope Martin V against these dissenting followers of Jan Hus, but the German and Hungarian knights encountered the most advanced tactical system of the new age of gunpowder. Combining the strategic offensive with the tactical defensive, the Hussite military genius drew up a line of "wagonforts"—armoured wagons affording cover to infantry armed with both handguns and crossbows. Between the vehicles were cannon mounted on four-wheeled carts and protected by infantry with pikes. Behind both wings the cavalry waited in readiness to deliver either a counterstroke or lead a pursuit.

The combination of the three arms continued to pre-

vail without a single reverse after the death of the blind leader. More than 50 battles and combats were won by the Hussites in 14 years, but civil war led to the destruction of the armies that had been invincible against the invading knights.

The age of transition. In 1453 it was evident that a new military age had dawned when the Turks breached the walls of Constantinople with a siege train of 70 cannon. After 40 days of bombardment enough towers had been leveled and holes opened in the masonry for a general land and sea assault to succeed.

If weapons of gunpowder could batter down the strongest walls of Europe, it was plain that a defensive based on iron men in stone castles was no longer dominant. By the middle of the 15th century the plate armour panoply had become so heavy that special breeds of horses, the ancestors of Percherons and Belgians, were raised as chargers. Only armour of prohibitive weight could stop a ball from a handgun at close range, and men were beginning to suspect that mobility offered more protection than iron.

The end of armour

Fortification and siegecraft had changed but little in 2,000 years. Down to the 14th century an Assyrian of 700 BC would not have found much to astonish him in the attack and defense of walled strongholds in Europe. But after the fall of Constantinople it was realized that the wall was obsolescent. Works of the future must be low and screened, burrowing into the earth for security. Besiegers must dig their way forward in a maze of parallels and approaches affording shelter from the fire of the besieged.

About this time the ponderous handgun was being superseded by the harquebus, a comparatively portable matchlock that could be fired from the shoulder. Improvements in the manufacture of the wheeled cannon had resulted in slender guns of cast iron or bronze, thus increasing mobility, and the invention of the trunnion permitted the muzzle to be elevated or depressed for more accurate aim.

New firearms such as the harquebus gun and lighter cannon were utilized by brilliant captains in the struggle between France and Spain for control of Italy early in the 16th century. Tactical values changed from one battle to another with bewildering rapidity. On the field of Cerignola in 1503, the French infantry and Swiss mercenary pikemen were mowed down by the fire of Spanish harquebusiers who counterattacked and captured the enemy's entire artillery train. Nine years later, at the Battle of Ravenna, the Spaniards again used entrenched harquebusiers. But this time the French moved 24 cannon to the enemy's flank and surprised the Spanish infantry with an enfilade fire that decided the battle.

Pike and harquebus. The Spaniards founded a new school of infantry tactics that was to rule the battlefields of Europe for a century and make possible incredibly daring conquests in the New World. Solid columns of infantry were drawn up in groups known as battles. Various combinations of "shot and pike" were used, but usually the harquebusiers were drawn up on the corners of battles 25 ranks deep. After firing at the word of command, each rank withdrew to the rear to reload and gradually moved forward by successive volleys until its turn came again. It was the function of the pikes to protect the harquebusiers while reloading. When the enemy's ranks were broken by firepower, the pikemen evolved from square into line and advanced, shoulder to shoulder, in a massive charge calculated to sweep the field.

"Shot and pike" tactics

The tercio of about 3,000 foot was the basic Spanish tactical unit. Consisting of 12 companies of about 250 men each, the fighting strength in 1534 was divided equally between pikemen and harquebusiers. Cavalry was represented in one-twelfth of infantry strength, and artillery by an average of one gun to each 1,000 foot.

The value of tactics and discipline was illustrated during the reign (1519–56) of the emperor Charles V when small bands of Spaniards won amazing victories over hordes in Mexico, Peru, and the islands of the Caribbean. The Indians were not wanting in sacrificial courage, and their advantage in numbers compensated for inferiority in weapons. One Aztec might have been a match for a Spaniard in single combat, but a thousand Aztecs were routed by a hundred Spaniards because they lacked the tenacity instilled by discipline as well as tactical system and precision.

Gustav Adolf and the Thirty Years' War. It fell to the Swedish king Gustav II Adolf (reigned 1611–32) to institute tactical innovations that earned him the title "father of modern tactics." The Swedish king organized his infantry into brigades, each consisting of two regiments that in turn included two battalions. Each battalion of 1,000 men was divided into eight companies of an average strength of 75 muskets and 60 pikes.

The beginning of modern tactics

Swedish armourers had developed a lighter musket, and Gustav is credited with being first to introduce paper cartridges to speed reloading. His six ranks advanced in open order and closed up into three, the first rank kneeling and the other two standing while firing a volley.

The Swedish artillery also benefited from innovations. Advances in metallurgy having made possible a lighter fieldpiece, Gustav recalled most of his old ordnance. These pieces were recast into fourpounders of three feet 10 inches (117 centimetres) in length, 400 pounds (181 kilograms) in weight, and 2.6-inch (6.6-centimetre) calibre. Four men or a single horse could handle the new gun on the field, though the fourpounders of other armies weighed half a ton (450 kilograms) or more.

The first artillery cartridge was developed as ammunition—a thin wooden case holding a prepared powder charge wired to the ball. Improvements were also made in the "bomb," or shell—a hollow iron sphere filled with gunpowder and fired with a lighted fuse. Reloading was speeded to such an extent that the Swedish fourpounder could be discharged eight times while the best musketeers fired six volleys. This was the first regimental piece in history, and each Swedish regiment of 1,000 men brought two fourpounders into the field.

Early shells

It was characteristic of the forthright Swedish king to train his cavalry entirely for shock effect with pistol and sabre. He organized Europe's first regular dragoon corps—a tactical hybrid made up of men armed with carbine and sabre who could fight either in the saddle or on foot.

The decline in body armour was accelerated when Gustav himself discarded all protection save a leather buffcoat. Field fortification was also reduced to its simplest elements by an army priding itself on the boast, "The Swedes do not defend their men with walls; the Swedes defend their walls with men."

In the Battle of Breitenfeld (1631), the first large-scale test, Gustav opened the action by committing 100 cannon to a long-distance artillery "preparation." Then his self-sustained battalions, each a mobile combat group of pikemen and musketeers with their own fourpounder, were able, by virtue of rapid shifts, to concentrate fire where it was most effective (see Figure 5). Of the 33,000 opposing troops, 7,000 were captured and 12,000 killed or wounded. Only about 2,000 survivors could be rallied after the pursuit.

From Captain Liddell Hart, *Great Captains Unveiled*, Blackwood

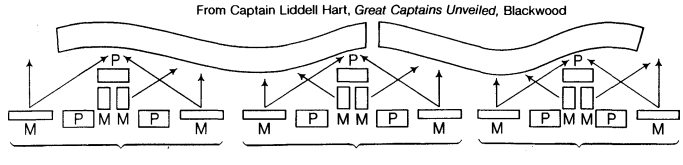


Figure 5: Companies of musketeers and pikemen drawn up in combat formation by Gustav II Adolf.

Although Gustav won another victory in the Battle of Lützen the following year, he paid for it with his life. In two years of campaigning he had overrun Germany, but after his death the army soon degenerated into another of the plundering hosts that made the war one of the most destructive of history.

The age of linear tactics. After the bloody battles of the Thirty Years' War (1618–48) resulted in a prolonged stalemate, the Viscount de Turenne, commanding mercenaries paid by France, cleared the enemy from Bavaria in a campaign of maneuver, compelling the elector Maximilian I to sue for an armistice. Avoiding battle unless he had an

overwhelming advantage, the French tactician relied upon caution, precision, and surprise to drive the enemy into ravaged areas while retaining his own sources of supply.

Thus the savage excesses of the war were made to appear futile by tactics that foreshadowed the most moderate era of strife in modern history. The founding father of the new military age was the Marquis de Louvois (1639–91), minister of war at the court of Louis XIV of France. Recognizing that tactics are dependent on supply, he organized a quartermaster general's department and a trained commissariat. As fixed points of maneuver, he established magazines at strategic centres for storage of reserves of food, ammunition, and equipment. The French minister also foresaw that two recent improvements in weapons would lead to tactical change. The first was the socket bayonet, which superseded the pike by providing every musketeer with his own means of protection. Equally important was the invention of the flintlock musket, making it possible to dispense with the clumsy lengths of smoldering "match."

Tactical measures of Louvois

Even Gustav Adolf had not been able to reduce the depth of his formations to fewer than six ranks. But now that it was no longer necessary to combine pikes with muskets, Louvois perceived that the day of linear warfare had dawned. His famous drillmaster, General Jean Martinet, whose name has lived as a byword, trained the troops of the new French armies to deploy from column into line and advance in three ranks, pausing to fire platoon volleys at the word of command. The effect depended on precision of approach, and the men were trained to move forward in unison, keeping step at a stately cadence of 80 paces a minute.

The French army had no equal in numbers as well as quality when Louis XIV began the aggressions that were to involve him in wars throughout his long reign. The Battle of Seneffe in 1674 made it plain that grievous losses were to be expected by the best troops when opposing lines traded volleys at 50 paces as a prelude to charging with the bayonet. Only iron discipline could fit men to stand firm and fill the gaps in the ranks, for the death casualties alone in this clash amounted to 17 percent for both armies.

The American Revolution. Many of the tactics of the Napoleonic era had their roots in the American Revolution. It was evident to observers that the American frontier rifle ranked as history's first infantry firearm of precision. Although it took twice as long as a smoothbore musket to reload, it had from two to three times the effective range in addition to an accuracy that was beyond comparison.

Aimed fire versus volley fire

Although a majority of the Americans carried smoothbore muskets, the farmers and villagers of the New World were accustomed to hunting small game at a time when laws against poaching and possessing firearms prevailed in Europe. The greenest American recruits took aim instinctively, while European regulars were trained to point the piece in the general direction of the foe and fire at the word of command. Thus, at the Battle of Bunker Hill, most of the British shots passed harmlessly over the heads of Americans who inflicted casualties of 45 percent with their limited ammunition by taking "aim at the handsome waistcoats." Because of the broken country, light-infantry tactics flourished in the American Revolution. And since all the armies were small, a two-rank line gained approval as a means of adding to infantry firepower.

French Revolutionary and Napoleonic tactics. The "horde tactics" of the French Revolutionary armies evolved from military poverty. Lacking magazines, the sansculottes gained in mobility by becoming less dependent on bases and lines of communication. Lacking tents, they learned to bivouac in the field, where all units could be more swiftly concentrated. Lacking supply wagons, they learned to march at a more rapid cadence than their opponents. Lacking training for the precision of linear tactics, they gradually developed a more effective attack of their own.

At least two lessons of the American conflict were put into effect—firing at will, as distinguished from volley fire, and taking natural cover on the skirmish line. Upon coming in contact with the enemy, the bolder spirits of a revolutionary army swarmed out to fire from ditches or behind hedges. The platoon volleys of linear tactics were not effective against these elusive French skirmishers, who took to their heels when threatened with a bayonet charge. Fighting of this sort was so interesting and comparatively safe that more and more men participated. Finally, at the Battle of Hondschoote in 1793, French forces concentrated swiftly and almost literally mobbed an Allied rearguard detachment with a wild bayonet attack.

Bonaparte's early victories. It is taking no credit away from Bonaparte's genius for war to point out that he adapted rather than originated his tactics. He was the heir not only of French Revolutionary tactics but also of an army of battle-hardened veterans who had won some notable victories over the Allies. From the old royal army he inherited the system of dispersion and concentration of divisions and brigades.

Bonaparte was not a great battlefield tactician. He was at his best in that hybrid between tactics and strategy known as "grand tactics"—the assembling of units prior to battle for a last-minute concentration by forced marches on the battlefield itself. No longer could an adversary refuse combat, as in the days of linear tactics. Bonaparte forced such an opponent to fight by sending a column around to threaten his line of communications. When the enemy was brought to bay, Bonaparte could summon his own outlying columns as mobile reserves in less time than could the armies trained for linear tactics.

Bonaparte and the offensive column

The preliminary swarms of skirmishers, the cannonading by regimental guns firing grapeshot at close ranges, the battlefield concentration, and the attack by bayonet columns of infantry—these were the four stages of Bonaparte's early victories. Some of them were gained by envelopment, some by penetration, and others by plain frontal assault, according to the dictates of his intuition.

Wellington's tactics. At a time when other generals were imitating Napoleon's methods, the Duke of Wellington defeated French columns with linear tactics and depended on seaborne supplies while outmaneuvering an enemy living on a country eaten bare. He, too, was an adapter rather than an originator, for the British tactical system was largely the creation of Sir John Moore. In 1803, in training troops to meet an expected Napoleonic invasion, Wellington reduced the standard three ranks to the two-rank line of the American Revolution. He applied another American lesson by organizing three regiments of riflemen for his light division, soon to win fame as the most celebrated unit of British battles in Spain.

Moore was killed in his first peninsular operation, but Wellington applied his tactics so successfully as to become one of history's few unbeaten generals. The French, as Europe's masters, felt obliged to attack. Their opponent profited by being able in every major battle to choose his own ground and draw up his "thin red line" behind a low ridge. This gave him the double advantage of protecting his men from ricocheting cannonballs and compelling the French bayonet columns to advance uphill. A screen of skirmishers from the light division was deployed along the forward slope. Since their rifles outranged French muskets, they were able to harass the enemy columns with aimed fire while the British artillery thinned them with shrapnel. The two-rank line gave the redcoats a wider front for musketry, and Wellington met the main French shock with steady platoon volleys. If the enemy wavered, he ordered a British bayonet counterattack.

This was the tactical formula, repeated with few variations, that won victories over Napoleon's marshals for five years in Spain. Wellington proved to be a match for the French in swift concentrations, and on one occasion his light division completed a march of 43 miles in 22 hours. But it was with infantry firepower that he beat opponents who trusted too much in the bayonet, for Napoleon had refused to adopt the rifle and neglected musketry training in spite of his assertion that "the power of infantry lies in its fire."

Half a century of transition. At Waterloo, Wellington needed only the tactics that had won for him in Spain. But an era came to an end on that June day in 1815, when infantry, cavalry, and artillery were in close contact and

The rifled gun

could be controlled by the general as easily as a regiment a century later. All this was changed by the widespread adoption of the percussion cap and long bullet, which multiplied the range as well as accuracy of the rifle and made it a universal military weapon. No longer was it possible to place the guns in front of the infantry, as was done sometimes during the Napoleonic era, with the cavalry close behind. No longer was it possible to mass cannon only 500 yards from the enemy, as at Borodino, and blast a bloody path to victory. Thus the old battle order, which had changed but little in essentials since the day of Gustav Adolf, was relegated to the past. To command an army as Wellington did at Waterloo was out of the question; to render its separate parts cooperative demanded staff work and not merely a powerful voice.

During the half century preceding the American Civil War, inventions came so rapidly that soldiers could not keep pace. The revolver and repeating carbine opened up new vistas for the cavalry and rendered obsolete the charge of lancers. Rifled and breech-loading fieldpieces worked a revolution in artillery tactics, just as the telegraph, the railroad, and the steamship made it necessary to revise previous ideas of military transport and communication.

Few precedents were available when green armies of citizen-soldiers met in the opening engagements of the U.S. conflict. But if they had more to learn, they also had less to forget than the tradition-ridden soldiers of Europe. And before the four-year struggle ended, the Americans had put into effect many of the tactics that would shape the outcome of World War I.

The American Civil War. Four million men bore arms in the U.S. conflict. More than 2,000 combats took place, of which 149 were of enough importance to be called battles. Yet these far-flung operations were dominated throughout by the defensive power of the rifle, supported by the ax and the spade.

Return to the defensive

The railway gun, the land mine, the wire entanglement, the observation balloon, the hand grenade, the field telegraph—these were some of the innovations or revivals of the Civil War. But they were second to the demonstrated ability of the man in the rifle pits to hold his own against three attackers. Since most of the operations occurred in wooded country, soldiers of both sides learned to cut "head logs" with firing slits to give added protection along the parapets of trenches. Generals were taught by experience that marksmen in log-faced earthworks, supported by rifled artillery, could seldom be dislodged except at a ruinous cost in casualties. General Robert E. Lee at Gettysburg and General Ulysses S. Grant at Cold Harbor—the best tacticians of the war—met bloody repulses when they underestimated the firepower of opposing riflemen.

On the other hand, paradoxical as it may seem, the defensive power of the entrenched rifleman enabled Lee to launch successful attacks on opponents with a material advantage. Relying on the proved ability of one rifleman to stand off three adversaries, the Confederate general dared repeatedly to divide his army in the face of a more numerous enemy.

European soldiers professed to see in the Civil War only the incoherent clashes of armed mobs, and in 1870 the French and Prussians made it evident that they had learned nothing from the American precepts. They sent cavalry in suicidal charges against infantry men with breechloading rifles with an effective range of 1,000 yards; they took needless losses from rifled artillery as the penalty of close formations; and they persisted in infantry frontal attacks despite the fact that neither side succeeded in taking a single position by such tactics.

World War I. *The Battle of the Marne.* It was the tragedy of World War I that the century of comparative European peace since Waterloo seemed to have made romanticists out of the professional soldiers who held in their hands the lives of millions of conscripted citizens.

The machine gun

Despite the lessons of the U.S. conflict, confirmed by the machine guns of the Russo-Japanese War, France's generals of 1914 had a mystical faith in the *offensive à l'outrance*—the headlong offensive, compounded of mass, velocity, and morale. They believed that the prewar scale of two machine guns to a battalion was adequate; and as late as 1915 the British commander Sir Douglas (later Earl) Haig asserted that it was "a much overrated weapon" and agreed that the French scale was "more than sufficient." The German generals were more realistic, but they too had complete faith in the offensive and regarded any reference to defensive advantages as bordering on defeatism. Both the French and German tacticians considered a defensive action excusable only when "each man will die in place rather than give ground."

The Germans, who were better prepared at the outset, seemed to be proving their case as they swept through Belgium and northern France. The French soldiers, uniformed in red breeches for the sake of morale, endured frightful losses. But since the German generals held that "the abandonment of close-rank formations is an evil which should be avoided whenever possible," their troops in turn provided the French artillery with lucrative targets.

In the Battle of the Marne the French stopped the invaders by taking advantage of blundering last-minute revisions in the German envelopment plan. Then began the "race to the sea" that was to result in a continuous front of trenches from the Channel to the Alps. The German losses of the first four months were 747,465 killed and wounded; the French had 854,000 casualties during the same period.

Deadlock on the Western Front. In the spring of 1915 the French and British believed that artillery was the key to the deadlock. But even after a prolonged bombardment by 1,140 pieces of heavy artillery, 53 French infantry divisions were able to advance only 3,000 yards on a 12-mile front. The enemy simply weathered the storm of steel in deep dugouts until the barrage lifted, then manned the machine guns to cut down attackers plodding across no-man's-land under heavy burdens of equipment.

Use of poison gas

The Germans risked the censure of neutral nations in an attempt to win a decision with poison gas in the Neuve Chapelle sector. But the Canadians held out until they could be reinforced, and millions of soldiers were soon carrying gas masks as all armies resorted to chemical warfare. Steel helmets also were generally adopted as protection against shell fragments.

The offensives of 1917. The Western Front was not the only theatre of war, but it was recognized that France would be the area of ultimate decision. In the spring of 1917 it was still thought that by increasing the density of artillery the infantry could make headway behind a rolling barrage timed to the advance. But even a ratio of a gun to six yards of front was not enough, and the German machine guns continued to reap their deadly harvest.

It had been necessary for a civilian, David Lloyd George, to provide the British army with a realistic scale of machine guns and trench mortars in spite of the reluctance of Haig and Lord Kitchener. Another civilian spoke for the dead of France when Premier Georges Clemenceau remarked that "modern war is too serious a business to entrust to soldiers."

Although Belgian engineers had warned against destroying the drainage system of reclaimed lands, 103,000 tons of shells were fired in the preliminary British bombardments of the third Battle of Ypres, known to the soldiers as Passchendaele. Mustard gas was added to the horrors of the 100-day battle as both armies wallowed in the swamp created by 4.3 million British shells. Haig's gains were negligible despite 300,000 casualties, and the Germans considered Passchendaele "the supreme martyrdom."

Early use of tanks

One of the few successes of a gloomy year was the British breakthrough at Cambrai on a six-mile front with 381 tanks supported by infantry. Attacking by complete surprise, after a last-minute hurricane barrage, the clanking cavalcade ripped through defenses in depth and took 7,500 prisoners and 120 guns at a low cost in casualties. Adequate infantry forces had not been provided by high-level British generals who envisioned a cavalry exploitation, and the territorial gains of Cambrai were lost to a German counterattack. Even so, the operation had demonstrated that armour was the key to a decision on the Western Front.

The Germans, meanwhile, had developed a key of their own in a new concept of offensive tactics known as infil-

tration. It called for the rapid advance of dispersed "storm groups," armed with automatic rifles and light mortars, preceding the main body of infantry. These specialists were to probe for weakness and break through wherever they could find an opening. Their mission was to surprise enemy machine guns and artillery from the flanks and rear while leaving the strong points to be enveloped frontally by following waves of infantry, closely supported by artillery and tactical aircraft. These new tactics of infiltration, carried out by secretly trained troops, promised to create enough demoralization for a penetration as far back as corps command posts and communication centres.

Leaders of both sides realized early in 1918 that it was destined to be the year of decision. All the nations that had participated from the beginning were war-rotted except Great Britain and Germany, and it was a question whether the newly trained U.S. divisions could cross the Atlantic in time to compensate for the defection of Russia, convulsed by revolution.

The mobile warfare of 1918. Infiltration won the first round. The German spring offensive in Picardy resulted in gains for a few days such as the Western Front had not known since 1914. Within a week a salient had been extended to a maximum depth of 40 miles in the British sector by German columns that threatened to drive to the sea. Then, suddenly, the attack ground to a halt and the front was stabilized once more.

This loss of momentum was no coincidence. What had happened was that the Germans had outrun their supplies, for the rail transportation and horse-drawn artillery of 1918 were not adequate to nourish a rapid advance. General Erich Ludendorff made four bids for a decision, and each time the early gains merely resulted in creating another salient and two more German flanks.

The last effort, on July 15, 1918, was a complete failure. Attacking on the Champagne front, the Germans were trapped in the French elastic defenses planned by General Philippe Pétain. Here, for the first time in the war, land mines were used on a large scale, and the attackers never reached the main French positions.

The riposte came three days later when Americans and French Moroccans counterattacked on July 18 along the right face of the salient created by the German drive to the Marne. Complete surprise was achieved by rushing the assault troops up in trucks under cover of darkness. Without any preliminary bombardment, the infantry advanced at dawn with 352 tanks and penetrated so deeply as to threaten the whole salient when other French and U.S. troops attacked the opposite face. Before the Germans could pull out, they left 30,000 prisoners behind and suffered about the same number of killed and wounded.

Mechanized armour and troop transport

The precept of motor transport set at Verdun in 1916 bore further fruit as Entente generals learned that they could gain surprise unfailingly by these tactics. Dispensing with bombardment and bringing the assault forces up in trucks meant that there would be neither ammunition dumps nor troop concentrations to warn the enemy by means of his aerial observation.

The Amiens salient was next to be reduced, as 13 British divisions attacked on August 8 with 450 tanks, while armoured cars manned by automatic riflemen spread confusion in German rear areas. On September 2 another British attack struck the enemy in Flanders, and on the 12th an army composed of new U.S. divisions wiped out the Saint-Mihiel salient.

The end was in sight during October as the British on the left and French in the centre rained blow after blow, striking by surprise when and where such strikes were least expected. These victories stemmed in large measure from the pressure exerted on the enemy as the U.S. troops on the right slugged their way forward in the 46-day Battle of the Meuse-Argonne against an enemy who was using a fourth of his entire strength on a 20-mile front in difficult terrain.

World War II. *The post-World War I reaction.* During the postwar years the reaction of the victorious powers was in the direction of defensive tactics. The casualties inflicted by the machine gun were remembered when the victories of the tanks were forgotten. This tendency went so far as to resemble in some instances the wishful tactical thinking of 1914 in reverse.

The Maginot Line

The French, with traumatic memories of two German invasions in half a century, translated their theories of the all-powerful defensive into the concrete and steel of the Maginot Line—a bristling belt of underground forts and tank traps extending from the Vosges westward to the Ardennes, where it was believed that the hills and woods would turn back armoured columns. The British also had a defensive military philosophy, the main tenet of which was that postwar improvements in weapons added to the risks and perils awaiting a future aggressor.

When the Spanish Civil War served as a rehearsal for Soviet as well as German and Italian tactics, it was belatedly realized that improvements in weapons favoured the offensive to such an extent as to endanger laggards in preparations. Tanks were many times as fast and powerful as those of 1918; motorized artillery, on rubber tires, showed an enormous gain in mobility over horse-drawn, iron-wheeled vehicles; and motor trucks poured from the assembly lines by the thousands to solve the transport and supply problems of a new lightning warfare.

Early German victories. Any doubts as to the effectiveness of German preparations were dispelled during the first two years of World War II as Hitler's armoured columns swept on to victory after victory. The Maginot Line was outflanked by German tanks that broke through in the Ardennes and drove from the Meuse to the Channel in a week, cutting off 335,000 men of the British Expeditionary Force and attached French units. It seemed an authentic miracle when these beleaguered troops were evacuated by sea from Dunkerque (Dunkirk), leaving their heavy arms behind, as the German columns raced on to Paris.

Blitzkrieg tactics

From the late summer of 1939 to the autumn of 1941 the Germans never met with a serious reverse in land warfare. Their triumphs were gained with such swiftness, such precision and economy of means that the blitzkrieg, or lightning war, was popularly accepted as a brilliant new tactical system. Actually it was the legitimate offspring of the infiltration tactics of 1918 mated with tank tactics. There was nothing new that could not be explained by the gains of two decades in the offensive powers of such weapons as the tank, motorized artillery, and tactical aircraft. With a few exceptions, such as antiaircraft guns, defensive weapons had lagged in the armament race while armour and artillery were leaping ahead from the three miles per hour of 1918 to the 30 miles per hour of 1938.

The French army of 1940 was essentially a World War I instrument, patched here and there with tactical changes, just as the Maginot Line was in effect a modernized western front. Despite warnings that the enemy might strike through the gap in the Ardennes between the French and Belgian fortified lines, the French in the Sedan area were unprepared for German "tactics of the space and gap." Only three penetrations on a 12-mile front, each no wider than 2,000 yards, were enough as the starting point of a major breakthrough. Like a flood enlarging holes in a dam, tanks and combat groups in armoured cars poured into the gaps, depending on surprise and speed as well as Stuka dive-bombers for flank protection. Their direction of attack was lateral as well as forward, the object being to widen the gaps for following German forces. Within four hours the invaders were overrunning command posts and communications centres in French rear areas too demoralized for effective resistance. The Maginot Line had simply been bypassed and left behind by the torrent of armoured invasion that inundated northern France in a week.

Predictions that Hitler would meet his downfall in the U.S.S.R. were not upheld by military results of the first few months. The Soviet army seemed doomed to extinction by its very bulk as German armoured envelopments swallowed up prisoners by the division and corps. Political factors influenced tactics when Hitler's doctrines of racial superiority led to cruelties that drove Russian peasants into guerrilla warfare. Effective resistance of this sort, combined with Soviet scorched-earth tactics and a seemingly endless reserve of manpower, enabled the Soviet army to survive until the German transport bogged down in autumnal mud.

Even so, the tactics of Hitler's armies were so successful during the first two years that a serious defect might easily be overlooked: the apparent lack of preparation for amphibious tactics in a global war. Although the British were desperately short of arms after the Dunkerque evacuation, the Germans had neither the training nor the equipment to land troops across a few miles of water. In Crete they proved their mastery of airborne attack techniques, but a clumsy attempt at an amphibious combat landing ended in disaster.

The war in the Pacific. After the United States and Japan entered the lists as enemies, the war in the Pacific necessarily consisted in large part of amphibious operations. Events showed that the U.S. forces had developed an effective system of amphibious combat tactics. Three continents were opened up to invasion by this system, which won a score of major victories without suffering a single major defeat. In the 1920s the "Orange" plans, adopted by the United States for offensive operations in the event of war with Japan, called for attacks on that nation's shield of fortified islands in the Pacific. By a directive (1927) of the joint board of the U.S. Army and Navy, the Marine Corps was given the mission of preparing for the conduct of landing operations.

U.S. amphibious tactics

The outer barrier of Japan's island defenses was breached by U.S. ship-to-shore assaults. Months of detailed planning and rehearsal preceded each operation. Supply ships were "combat-loaded" so that tanks, trucks, artillery, fuel, rations, and water would be made available to the landing force in the order needed. Intricate firing plans were drawn up for artillery and rocket bombardments to scorch the target area while armoured landing craft took the assault troops to the beaches under cover of air strikes. The transition from one element to another was the critical moment. Maneuver was seldom possible during the controlled chaos of the landings, and repulse meant a disaster that could hardly be retrieved. Once the beachhead was won, special units unloaded cargo ships, established supply dumps, and regulated traffic while the assault troops drove inland with the support of tanks and artillery as soon as they could be landed.

The amphibious offensive had become by 1943 the most decisive tactical system of World War II, not only in the Pacific theatre but also in North Africa and Europe. Neither the Japanese nor the Germans ever found the clue to a successful defense. Landings in Morocco, Algeria, Sicily, and Italy were followed by the climactic operation of June 6, 1944, when 4,100 ships put ashore five divisions complete with tanks and artillery on the beaches of Normandy. Nine days later in the Pacific an equally successful amphibious combat landing on Saipan brought U.S. bombing planes within striking distance of Tokyo and led up to the assaults on Iwo Jima and Okinawa that made Japanese defeat inevitable.

Significance of amphibious warfare in World War II

Wide variety of World War II tactics. So varied were the geographic and climatic conditions that it is difficult to generalize about the tactics of a war in which a score of major campaigns were being waged simultaneously in different parts of the world.

Tank battles on the plains of the U.S.S.R. began with tremendous artillery "preparations" and ended with armoured envelopments in which as many as 20 divisions were surrounded. In such instances the beleaguered forces were sometimes able to form a "hedgehog," or circular formation, and fight their way out of the encirclement with the aid of airborne supplies. In the long run it was a grinding war of attrition in which the Soviet army gained the upper hand largely as the result of arms and equipment delivered by Britain and the United States in spite of heavy losses suffered by sea convoys.

Armoured columns in the desert warfare of North Africa achieved a mobility that makes their operations instructive to the student of tank tactics. But General Erwin Rommel's defeat was foreordained by the inability of the Germans to mount an amphibious assault on Malta. The campaign in the desert was actually a battle of supply, and while the British held Malta they could provide air escorts for convoys of cargo ships in the Mediterranean.

On all fronts, special units equipped for special tasks were employed to an extent never known before. Infantry was the arm of decision in most operations, but it became standard practice to organize task forces ranging in size from platoons to divisions, reinforced by tanks, artillery, rocket launchers, or such other support weapons as might be indicated for the specific mission.

So potent were the missile weapons of World War II that survival for the individual soldier depended on cover and concealment. Small infantry units took on increasing importance as experience taught that latitude of decision must be left to troops in direct contact with the enemy, since victory often depended on the initiative and intelligence displayed by platoon commanders.

Reliance on small units

Some of the infantry combat teams became standardized. The British commando units, made up of soldiers, sailors, and marines organized on a permanent basis, conducted raids throughout the war. In Burma (Myanmar) the British developed special penetration groups of infantry to operate behind the enemy's lines, attacking headquarters and supply dumps while being supplied by air.

All armies made an unceasing effort to gain a decision by the introduction of new weapons and tactics. But the German V-1 jet-propelled bomb came too late to change the outcome, as did the more formidable V-2 rocket. On the Allied side, the 2.36-inch rocket launcher and the 57-millimetre and 75-millimetre recoilless rifles provided the infantry with weapons that were tank killers under favourable circumstances. The variable-time artillery fuse enabled a shell to be exploded automatically in electronic reaction to waves reflected from the target itself. These tactical innovations seemed trivial, however, after the atomic bomb was dropped on Hiroshima on Aug. 6, 1945.

The atomic bomb

Thus, at the conclusion of World War II the participants were confronted by the fact that "the bomb" might render obsolete all the tactics of victory. Even the discovery of gunpowder does not represent a comparable turning point in tactical history, for the first handguns were not as deadly as the crossbow. Nor did gunpowder have a rapid evolution, considering that the trinitrotoluene (TNT) of the 20th century was only about twice as powerful as the explosives of the Middle Ages. Yet the world of 1945 had to adjust itself overnight to the necessity of revising all military standards of the past.

Tactics in the nuclear age. *War in Korea.* It is one of history's ironies that the first war of the strident new nuclear age should have been waged with tactics reminiscent of the Western Front in 1917. Combat in Korea (1950–53) consisted for the most part of infantry fights with small arms on the platoon and company levels.

Called a "police action" at first, the intervention of 1950 proved during its first year to be the fourth most costly military effort of U.S. history. After a preliminary period of mobility, the United Nations and Chinese communist forces faced each other in lines of trenches and defenses in depth extending across the peninsula. Neither side attempted to hold its front in force. Bunkers and outposts on the main line of resistance sheltered combat groups whose duty it was to maintain contact with the enemy and warn of a large-scale attack. Patrols ranging in size from an infantry squad to a platoon crept through no-man's-land in the darkness as both sides found the ambush an effective tactic.

The Korean War showed that even in the nuclear age an Asian peasant army using semi-guerrilla tactics can be formidable in mountain terrain suited to its operations. Atomic weapons would not have been effective in such a war, and the Korean experience made it appear unlikely that sound infantry tactics would soon become outdated.

The two great tactical innovations of the war were the combat helicopter and the revival of infantry body armour. Rotary-wing aircraft had been developed too late for an extensive test in World War II, but in Korea the helicopter proved its value immediately in tactical as well as logistical respects. A U.S. brigade commander found it possible to maintain combat control over units 25 miles apart with the use of rotary-wing transportation. When a U.S. squadron of transport helicopters reached the front, it was demonstrated that a battalion of infantry could be lifted in a few hours to the main line of resistance in moun-

Introduction of helicopters and body armour

tainous country where an all-day march otherwise would have been necessary. On an occasion when an infantry company was surrounded, six helicopters flew in 10 tons of ammunition and evacuated casualties, thus enabling the encircled troops to fight their way out of the trap.

The revival of body armour, after a lapse of several centuries, was made possible by 20th-century plastics. Nylon pads and thin slabs of Fiberglas were combined in a sleeveless vest weighing about eight and a half pounds (3.9 kilograms) and allowing freedom of movement. The armour was capable of stopping .45-calibre automatic pistol bullets, submachine gun slugs, and most shell, mortar, or grenade fragments of a velocity less than 1,100 feet (335 metres) per second.

All U.S. combat troops in Korea were equipped with the garment before the war ended. On the basis of U.S. Army and Marine Corps surveys, it was established that body armour prevented from 60 to 70 percent of chest and abdominal wounds, while reducing the severity of 25 to 30 percent of the wounds resulting from penetrations. Aside from humanitarian aspects, this meant that a large proportion of the enemy's most effective antipersonnel weapons were rendered useless. The tactical value of body armour was increased by its contribution to morale. Troops on the defensive had their foxholes and trenches, but the vest-protected men advanced in the open during an attack, when protection was most needed. The result, according to U.S. infantry officers in Korea, was a marked increase in combative spirit.

Later tactical organization: Vietnam. In 1961 a new divisional structure consisting of an "all-purpose" divisional base with three brigade headquarters to which could be attached varying numbers and types of combat maneuver battalions was introduced by the U.S. Army. This reorganization provided a division whose size (10,500 to 19,950 men) and equipment could be tailored to each specific mission and to the type of terrain encountered in its accomplishment. Other results of the change were improved balance between nuclear and nonnuclear firepower, greater internal flexibility, and an improved command and control structure.

Still another change in tactical thinking occurred during the war in Vietnam in the 1960s and '70s with the establishment of special forces for carrying on small-unit guerrilla warfare behind enemy lines or cooperating with indigenous populations in antiguerrilla operations. The helicopter again proved its value in Vietnam, where troops and supplies, including heavy equipment, could be speedily transported over impassable jungle areas. A significant new development was the introduction of armed helicopters that played an active role in attacking the enemy. Helicopter flight crews wore a lightweight, flexible "flak jacket" that covered the upper torso. U.S. Marine Corps ground troops were not required to wear such protective body armour, because of the heat and humidity, but it was available for those who chose to wear it. (L.Mo.)

Naval tactics

FUNDAMENTALS

Being the activities of battle itself, tactics are conceived and executed at the literal and metaphoric centre of war's violence. Tactical science is an orderly description of these activities, and tactical art is the skill required to carry them out in combat.

The search for constants. It should be said that, in order to achieve victory, willpower and courage must always accompany tactical art and science and often dominate the outcome of battle. These qualities are not tactics, but they are related to tactics in the way a sound decision is related to the resolution with which it is implemented. There is no finer example than Horatio Nelson. In the Battle of the Nile (Aug. 1–2, 1798), not only were Admiral Nelson's tactical decisions brilliant, but he had so imbued his captains with his thinking that, when they saw a chance for surprise by attacking the disengaged side of the French fleet, they were quick to seize it and gain a decisive advantage (see Figure 6). Still, their decisions only established the basis of that great victory, for the French fought with desperation, and it took hard fighting by British tars, inspired by Nelson's charismatic leadership, to fulfill the promise of victory.

In a similar manner, new technology is not tactics, but it may have a decisive effect in both altering the face of battle and affecting its outcome. Navies put special emphasis on warships and aircraft. It is well said that on the ground men are served by their weapons while at sea weapons are served by men. Lest his readers be too enamoured of élan and fighting spirit, Rear Admiral Bradley A. Fiske used a telling example in *The Navy as a Fighting Machine* (1916). He pointed out that in the American Civil War the Confederate ironclad *Virginia,* with 10 guns, handily defeated the Union sloops-of-war *Congress* and *Cumberland,* which carried a total of 74 guns. One day later the Union's *Monitor,* carrying two guns in a turret, fought the *Virginia* to a standstill. Courage and resolve were powerless against progress and armour.

The American naval strategist Alfred Thayer Mahan made perhaps too much of the influence on tactics of technological progress. In his seminal *The Influence of Sea Power upon History, 1660–1783* (1890), he wrote that, due to new fighting systems, "from time to time the structure of tactics has to be wholly torn down but the foundations of strategy so far remain, as though laid upon a rock." Mahan appreciated the utility of naval history for the discovery of strategic constants—that is, principles of strategy that have remained valid throughout technological change. Tacticians, on the other hand, are conscious of tactical constants as well, especially the following: the power of concentrated force (rarely in history has a naval tactician withheld a reserve); the special value of surprise; the abiding need for cohesion brought about by sound command and combat doctrine; the consummate goal of attacking effectively first; and the unique role played by timing and timeliness.

Five constants of naval warfare

The study of trends. Naval officers also study history for its trends, because trends are the only clue as to how tactics are changing and are the best check against the fatal sin of preparing to fight the last war. The trend that has influenced all else in the conduct of naval battle is the increasing range and lethality of naval weapons. Paradoxically, greater lethality has not led a trend toward greater loss of life. The first reason is that, unlike ground combat, the principal aim at sea is to put the fighting machine, not the fighting man, out of action, and modern machines are (thus far) sensitive to damage. Second, it is a longstanding constant that naval battles, once joined, are fast-moving and decisive.

To sketch how the range of weapons has affected naval tactics, a simple structure that describes the processes of combat must be established. First is firepower delivery itself. Second is the scouting process, which gathers information by reconnaissance, surveillance, cryptanalysis, and other means and delivers it to the tactical commander. Third is command itself—or command and control (C^2) in modern parlance—which assimilates the information, decides which actions are called for, and directs forces to act accordingly.

Combat being the activities of force against force, there is a natural antithesis to all three processes described above. First, the effect of enemy firepower is reduced by shooting down the incoming aircraft or missile, by maneuvering to avoid a torpedo, and by ship survivability or "staying power"—that is, the ability to continue fighting after suffering damage. Second, when scouting was accomplished by ships or aircraft flung out ahead of a formation, information denial was accomplished by screening—that is, by flinging out an opposing line of ships and aircraft. Modern ways to confound the enemy's scouting effort are keeping radio silence and jamming his radars, both of which deny him information. Third, enemy C^2 can be confused by deceptive signals or decoy forces. It can also be crippled or delayed by electronically jamming enemy communications.

The essential elements of naval combat

The seven processes described above—namely, firepower delivery, scouting, C^2, the three countermeasures against them, and maneuver—are the raw materials of naval tactics. To achieve success, they are synthesized into a

harmonious blend of action and counteraction. For example, a modern naval screen of ships or aircraft defends a formation both by destroying enemy aircraft or missiles and by denying tactical information. The screen itself may even be so central in importance that it becomes the focus of enemy attack, with destruction of the screen being tantamount to destruction of the force. Thus, the study of naval tactics has become more than the study of formations, firepower, and maneuvers. The increase in weapon range has been paralleled by scouting and the control of forces at longer and longer ranges; these in turn have opened up more avenues to gain information and confuse the enemy's picture by electronic means.

Tactics in the modern era. The study of tactics has always emphasized actions between fleets that gain or challenge control of the sea lines of communications. That traditional emphasis is retained here, but in the 20th century three other types of combat at sea have demanded greater attention.

War from land to sea and from sea to land. While navies have always had as their ultimate objective an influence over events on land, aircraft and missiles have extended the range and amplified the influence. Likewise, land-based systems have made their growing influence felt on warships and sea-lanes alike. Putting ground forces ashore from the sea by amphibious landing is an operation that has neither gained nor lost importance since the earliest galley warfare, but modern combined-arms tactics are quite different and require separate attention. A by-product of the extended range of modern weapons is the greater complexity of joint operations. The reorganization of many armed forces has been in large measure a response to the demand for well-coordinated operations.

Nuclear weapons. The United States, the Soviet Union, the United Kingdom, France, and China keep a considerable strategic deterrent force at sea in the form of submarine-launched ballistic missiles. The safeguarding or threatening of nuclear submarines has inspired a set of tactics unique in history. These tactics are among each nation's most closely guarded secrets and have never been used in anger, so that all knowledge of them is theoretical, based on mathematical computations, computer simulations, and constrained exercises at sea. This does not mean that such tactics, if ever used, will prove unsound. During the great transition from sail to steam and big gun, keen study by naval tacticians throughout the world developed the sound tactics that were finally practiced in World War I.

Theoretical tactics in the modern era

Raiding. War against trade is the war of an inferior navy that cannot compete for command of the sea but that, instead, dispatches raiders to deny the enemy its free use. These tactics of sea denial are those of predator and prey, of hunter and evader, and are as unique from the force-on-force tactics of major sea battles as are guerrilla war tactics from those of decisive land battle. With the maturity of submarines, these tactics have become so important that a separate section is devoted to them (see below *Guerrilla war at sea: the submarine*).

HISTORICAL DEVELOPMENT

In the following examination of the history of naval tactics, a shift in importance between elements of combat will be apparent. In galley warfare, sheer power dominated the outcome, and maneuver of numerous small ships, much as on land, contributed to its concentration. In warfare under sail, great firepower could be concentrated in individual ships, and doctrine, formations, and signal flags were means of controlling the slow-moving, wind-constrained formations. With battleships, steam power gave freedom to maneuver in any direction, and the range of the big guns allowed a concentration of fire from the whole formation. The defensive element came to prominence, symbolized by armour. Tactical decisions had to be made before the enemy was in sight, so that scouting became more evident as a tactical ingredient. Control had to be exercised over much greater distances, which expanded the possibilities of exploiting the means of control, notably through radio direction finding and code breaking.

The growth in the tactical influence of scouting, anti-scouting, command and control, and countermeasures against command and control continued through the era of the aircraft carrier and into the era of the guided missile. These elements have become as important to tactical success at sea as firepower, consuming comparable thought and resources.

The age of galley warfare. Galleys being relatively unseaworthy, war at sea among the ancients was always near land. Pictures of billowing sails notwithstanding, masts and canvas were stowed for battle, and oars were the means of propulsion. The most destructive weapon was a ram in the bow, which dictated a line abreast as the tactical formation. In the line abreast, two lines of opposing galleys approached each other head on, with the ram of each vessel unobstructed by the ships on either side. Momentum was the key to the ram's destructiveness, so that sprint speed—as much as seven or eight knots (nautical miles per hour, that is, eight to nine statute miles per hour or 13 to 15 kilometres per hour)—was as important as maneuver. Multiple banks of oars afforded speed, and the geometry of their arrangement fascinated naval architect-historians of later eras.

The ram and the line abreast

Major battles comprised hundreds of ships on a side. Battles occurred because of the threat of invasion, so that many armed men were present. These participated as archers or boarders. Rome developed grappling hooks and the *corvus* (a long boarding plank spiked at the end) to secure the victim ship while disciplined legionnaires fought their way on board.

Scouting the enemy formation was a subordinate issue, although contemporary descriptions indicated that formations and maneuvers, showing ingenuity and cunning, played a large part in the outcomes. Since battles were nearly impossible in foul weather, good visibility permitted the deployment of the lines abreast, often in two echelons, much as the commander intended. Descriptions of the battles (and the period was rich in them) were usually couched in the terms of land warfare—such as the routing of a flank, or an attempt to crush by encirclement. Galley tactics were so similar to land tactics that a reserve was actually held back—a practice that ever afterward was regarded in navies as a mistake. An inference cannot be drawn that a commander had tight control of his ships in action, however, and the correct image of a galley battle would be that of a wild melee, with oars smashed, hulls crushed, armour-clad soldiers drowned, losses enormous, and battles decisive with lasting consequences.

The ultimate battle under oar was at Lepanto on Oct. 7, 1571. Fought between a combined fleet of the Mediterranean's Christian nations and the vaunted Turkish navy, this battle was reported in great tactical detail. More than 200 Christian oared vessels met 270 of the Turks' in the Gulf of Patras on the Ionian Sea. By this time, three to five guns were fitted in the bows of galleys, and harquebuses were fired by Spanish soldiers. But, as was usual in galley warfare, the outcome was decided by boarding and hand-to-hand fighting with sword and pike. The Christian fleet under Don John of Austria prevailed in a bitterly contested donnybrook; losses to the Turks were placed at 30,000 killed, against 8,000 among the victors.

The age of fighting sail. By the middle of the 17th century guns arrayed along the sides of fighting ships were the decisive weapon. Heavy guns required a gun deck and a short, sturdy hull, which were at odds with the galley's requirements of lightness and length. Thus, the shift to sail was a victory of fighting strength over maneuver. Tactically, sailing navies became victims of the wind's whim, but strategically they benefited from nearly limitless range and, compared to the frail galleys, greatly improved seakeeping.

The column, or line ahead, became the logical tactical formation for bringing the most guns to bear. With all the ships of a battle line following one another, their guns could face the enemy line without obstruction. The three Anglo-Dutch Wars of 1652–74 saw the first closely studied battles of sail and gun. In them the column was as much a means for command and control as it was for concentration of fighting force, for as long as a fleet maintained station in line ahead, each ship separated by a

The naval gun and the line ahead

scant 200 or 300 yards, cohesion was assured, maneuvers were coordinated, and any malingering by reluctant captains was obvious.

The line was not a formation that permitted the concentration of fire, however, for naval guns in a rolling platform were effectively accurate at only about one-quarter of a mile, and the range for penetrability of shot was even less. In effect, engagements were decided within pistol shot, a battle line being a thin ribbon of death, miles long but scarcely 300 yards wide. When the English fought the Dutch in the 17th century, this was not considered a problem, because the tactics of both sides called for closing with the other aggressively. But in the 18th century their French opponents felt that their strategic interests lay in avoiding battle at close quarters, and in the Anglo-French wars the Royal Navy endured a long period of indecisive actions handicapped by a tactical doctrine so rigidly interpreted by courts-martial as to have become tactical dogma. These *Fighting Instructions,* though soundly conceived when first issued in 1653, were unsuited to this new opponent, for the implementing system of signals was unimaginative and constraining. Indeed, the two most admired tactical writers of the day, Paul Hoste and Sébastien François Bigot de Morogues, were French.

Toward the end of the 18th century, the British admiral Richard Kempenfelt began to unshackle the Royal Navy with a better system of signaling. The new freedom of maneuver came finally and forever to be embodied in the tactical genius and personal inspiration of Horatio Nelson, whose matchless victories at the battles of the Nile, Copenhagen (April 2, 1801), and Trafalgar (Oct. 21, 1805) drew the enduring admiration of naval tacticians.

Tactical study during this era concentrated on maneuver. "Breaking the line" of the enemy fleet was one aim, because this broke the enemy's tactical cohesion and made it possible to overwhelm individual ships by bringing greatly superior force to bear on each of them in turn. Popular aims were raking (firing a broadside the length of an enemy ship from across the bow or stern) or doubling (concentrating force by putting ships on both sides of the enemy line; see Figure 6). The most reliable way to concentrate gunfire was to build it into ships vertically by stacking gun decks one over the other. Later tacticians demonstrated analytically what every fighting seaman of the seafaring era knew instinctively: with equal competency exhibited on both sides, not only would a two-decked frigate beat a lighter, one-decked corvette, and a three-decker beat a frigate, but the loser would come away having done very little damage to his bigger opponent. Whence came the big "ship of the line," a three-decked ship that could stand in the line of battle and beat down smaller opponents while surviving to fight again.

The age of steam and big gun. Tactics and technology complement each other, and there is no better period in history for studying their interrelationship than the shift from sail to steam in the 19th century. The shell gun (raised to naval attention during the Crimean War by the Battle of Sinope, Nov. 30, 1853) compelled navies to adopt the iron sheathing of hulls. This pointed the way to all-metal hulls (iron, then steel), which in turn both permitted and demanded as a response the installation of rifled, breech-loaded guns of major calibre. Concurrently, iron boilers and screw propellers made steam propulsion practical and gave great new freedom of maneuver. Navies were unfettered tactically from the wind, but only at the strategic price of having to remain within steaming range of coaling stations.

Shell guns and armoured hulls

The sweeping consequences of these and other technological innovations lacked the crucible of war in which to test them, for it was an era of Pax Britannica, with the maritime peace kept by the Royal Navy. The *Monitor* and the *Virginia* (at the battle of Hampton Roads, March 8-9, 1862) marked the short-lived ascendancy of armour and the defense. This led to a brief revival of the ram and to some very speculative tactical concepts that looked outrageous in later days.

But the superiority of defense at sea did not last long. The tactical–technical turning point came from the observation of a few battles in East Asia around the turn of the century and from an often overlooked bit of military technology.

From *A History of War at Sea* by Helmut Pemsel, copyright © 1977 Lionel Leventhal Ltd.; permission granted by U.S. Naval Institute, Annapolis, Maryland

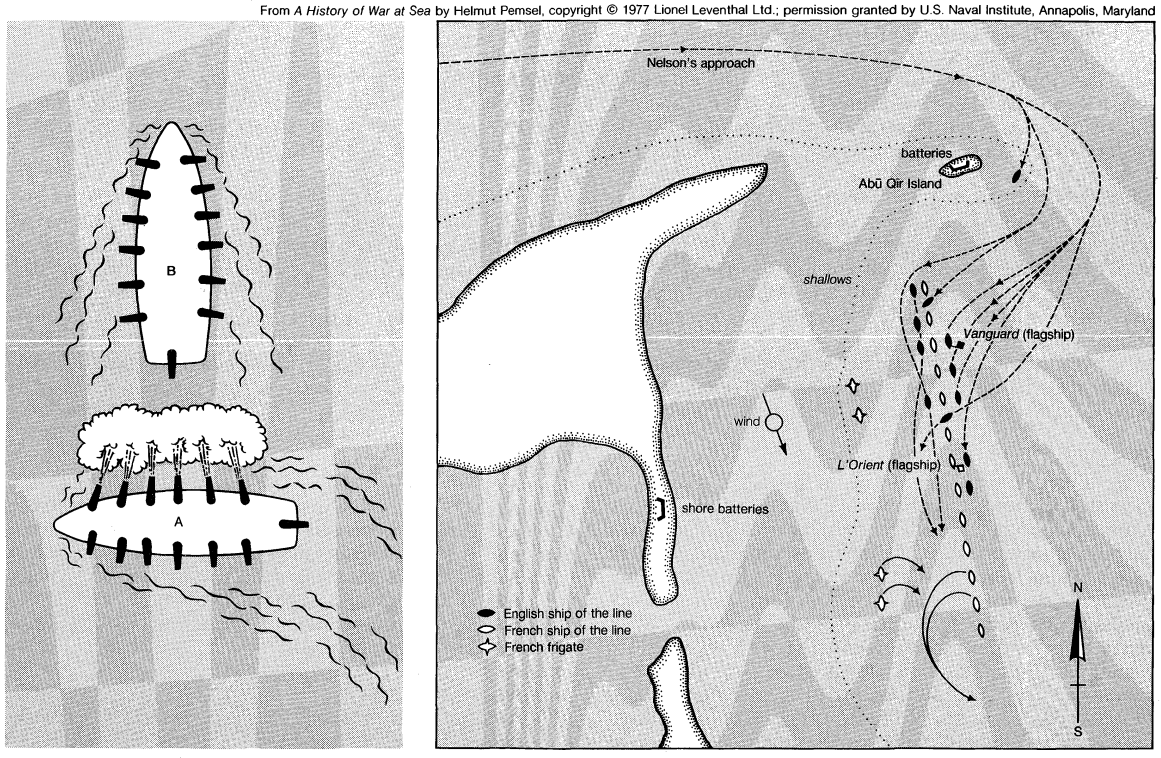


Figure 6: *The tactics of raking and doubling.*
(Left) In the raking position astern, ship A maneuvers across the stern of ship B and fires a broadside the length of the defenseless ship. (Right) The most brilliant feat of doubling was executed at the Battle of the Nile (Aug. 1–2, 1798). Nelson's fleet, rounding Abū Qīr Island, came upon a French fleet anchored in a line near the shore. Half of the English fleet passed inshore of the French, and the other half remained to seaward. With a light following wind, the English slowly worked down the French line, attacking each ship from two sides. Only four French ships escaped; two were sunk, and the rest were badly damaged.

The battles were those of the Yalu (Sept. 17, 1894), the Yellow Sea (Aug. 10 and 14, 1904), and Tsushima (May 27–29, 1905), in which the gun regained primacy to such an extent that the Russian vice admiral Stepan Osipovich Makarov could confidently write, "A good gun causes victory, armour only postpones defeat." The new technology was fire control, which enabled major-calibre rounds to be placed on target at five, then 10, and ultimately 15 miles.

By World War I the tactical issues were settled. First, big guns would dominate, a burly battleship firing a battery of them in broadside. Second, although armour could "postpone defeat," it was all but powerless against torpedoes; therefore, a destroyer screen was essential to protect the fleet. Third, a fleet cruised in compact formation but quickly deployed in column to fight. Since the opposing fleets would close with each other at a relative 40 miles per hour, a scouting line had to be thrown out well in front to report enemy movements by wireless. Despite the elephantine appearance of a line of battleships, there would not be a moment to lose when the enemy was sighted and no margin for error.

A column of battleships was not like a column of sailing ships. With greater range and improved fire control, gunfire from most of the ships on each side could reach most of the ships on the opposite side, making concentration of firepower by a whole fleet feasible and expected. The advantage was worked out mathematically in what were called the "N-square law" and the "square law of attrition": success would build on itself, so that any small advantage at the outset of an engagement would compound in favour of the superior force. With long-range gunnery, the advantage accrued fleet-wide, not merely ship by ship as in the days of fighting sail.

Crossing the T

A positional advantage could be added to this firepower advantage if the fleet "crossed the T" of the enemy, that is, if its own column crossed in front of the enemy column at a right angle and with the ships at the head of the enemy column within range of its guns. From this position at the top of the T, all the guns of the fleet could fire upon the head of the enemy column, while only the first enemy ships could return fire. This was the raking position sought by an individual sailing ship writ large, for at battleship gunnery ranges the whole force could concentrate successively on each enemy ship as it approached. It was reckoned that in fair weather and good visibility a fleet could destroy an enemy in the capped position in 20 minutes; or, if an enemy of equal strength could be surprised with unanswered fire within effective range for as little as five minutes, he would be demolished with little harm to the victor. There would be no leisurely approach, no chance to recover from a missed maneuver or a wrong turn. In fact, in practice the very swiftness of decision worked against maneuvering to cross the T. Much was made of the successful use of this tactic by the Japanese admiral Tōgō Heihachirō against the much slower Russians at Tsushima, but commanders at sea understood that the fast pace of battle worked against a T-crossing except by accident or surprise.

With one salient exception, there were no unforeseen tactics in World War I. The exception was the ease with which a fleet could be surprised at sea. The Battle of Jutland (May 31, 1916) was fought in great confusion, owing to a fog of smoke from the stacks and guns of 250 ships as well as the sloppy work of the commanders of the two scouting forces. The German commander, Reinhard Scheer, twice had his T capped for lack of visibility; for the same reason, the British commander, Sir John Jellicoe, twice was unable to exploit this, the ideal tactical position.

The lesser engagements of the war were also marked by surprise, but from another source. At one time or another, both the Germans and British broke each other's codes. Special intelligence and attempts to entrap a weaker enemy were rife throughout the war, leading to surprise in each of the battles in the North Sea: Helgoland Bight (Aug. 28, 1914), Dogger Bank (Jan. 24, 1915), and Jutland itself.

The extended range of naval aircraft

The age of the aircraft carrier. Early in World War II the primary instrument for delivering naval combat power became the aircraft carrier. The reason was range: aircraft could deliver a concerted attack at 200 miles or more, whereas battleships could do so only at 20 miles or less. The foremost tactical question during the transition in the 1920s and '30s was whether aircraft could lift enough destruction to supersede the battleship. Into the 1930s skeptics were correct that aircraft could not. But by the end of that decade, engines were carrying adequate payloads, dive-bomber and torpedo-plane designs had matured, carrier arresting gear and associated flight-deck handling facilities were up to their tasks, and proficient strike tactics had been well practiced. U.S. and Japanese naval aviators were pacesetters in these developments.

There was a subordinate tactical question as well: could the enemy be found at the outer limits of aircraft range? The ability to attack fixed targets such as the Panama Canal or Pearl Harbor, and to achieve surprise in doing so, had been amply demonstrated in naval exercises as well as in battle, but finding, reporting, and closing on ships at sea was a greater challenge. Without detracting from the courage and skill of aviators, it may be said that effective scouting was the dominant tactical problem of carrier warfare and had utmost influence on the outcomes of the crucial carrier battles of the Pacific Theatre in 1942: the Coral Sea (May 4–8), Midway (June 3–6), the Eastern Solomons (August 23–25), and the Santa Cruz Islands (October 26). In those closely matched battles the quality of U.S. and Japanese aviators and their planes was virtually on a par. When the United States won, it did so by superior scouting and screening, owing in large measure to air-search radar and to the advantage of having broken the Japanese code.

The command and control structure polished by the U.S. Navy during the war was the third vital component, after scouting and the delivery of firepower. The tangible manifestation of modern C^2 was the Combat Information Center, which centralized radar information and voice radio communications. By 1944 the tactical doctrine of coordinating fighter air defenses, along with the now much strengthened antiaircraft firepower on ships of the fleet, was so effective that in the Battle of the Philippine Sea (June 19–21, 1944) more than 90 percent of 450 Japanese aircraft were wiped out in a fruitless attack on Admiral Raymond Spruance's 5th Fleet.

Circular formation of carrier groups

The new tactical formation was circular, with carriers in the centre defended by an antiaircraft and antisubmarine screen composed of their own aircraft plus battleships, cruisers, and destroyers (see Figure 7). For offensive purposes, a circle allowed a rapid simultaneous turn by all ships in a task group in order to launch and recover aircraft. For antiaircraft defense, the circle was shrunk in diameter as tightly as possible so that each screening ship, by defending itself, helped defend its neighbour.

The new battle paradigm called for a pulse of combat power to be delivered in a shock attack by one or more air wings. Despite every intention, though, air strikes against alerted defenses were rarely delivered as compactly as practiced, nor were they as decisive tactically as naval aviators had expected. In the five big carrier battles, one attacking air wing took out an average of only one enemy carrier. (Viewed strategically, this average, along with losses of aircraft of around 50 percent per battle, was enough to govern the pattern of the Pacific war.) Since it took more than two hours to launch, marshal, and deliver an air strike, it was difficult to attack before an enemy counterstrike was in the air. Successful command at sea depended as never before on effective scouting and communication, because in order to win a decisive battle, in World War II as in all of naval history, it was necessary to attack effectively first.

Dominant though it was, carrier-based air power did not control the seas at night. With a modicum of success, the high-quality ships of Germany exploited the hours of darkness, especially during the winter months and in northern waters. In the bitterly contested campaign for Guadalcanal in the fall of 1942, guns ruled supreme at night and very nearly tipped the balance in favour of Japan. Expecting to be outnumbered as a result of the Five-Power Naval Limitation Treaty of 1922, the Imperial Japanese Navy had practiced night tactics assiduously in order, as they supposed, to whittle down the U.S. battle line during its

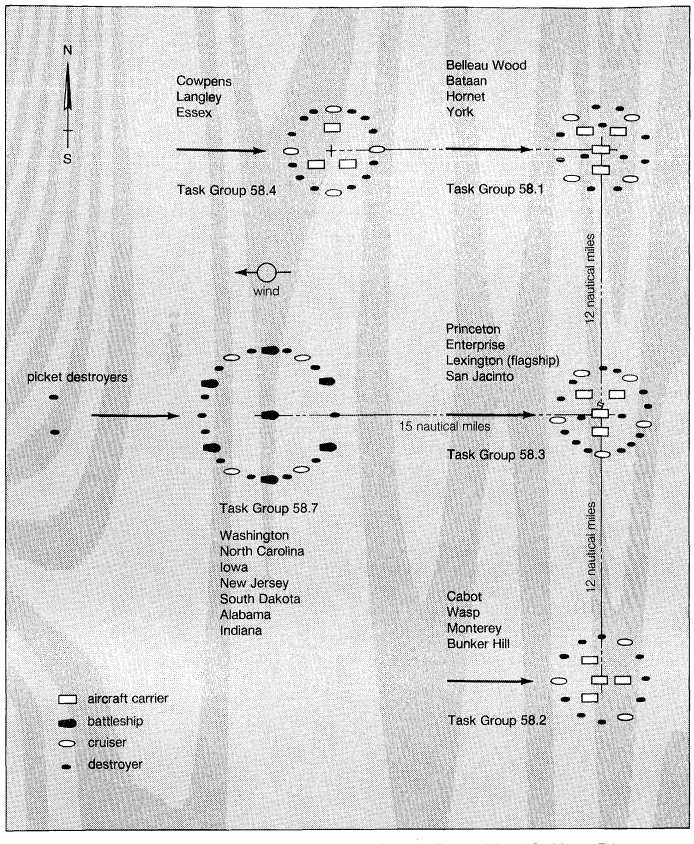


Figure 7: *Circular formations of Task Force 58, U.S. Navy 5th Fleet, Battle of the Philippine Sea, June 1944.*
Each task group is arranged in a circle approximately four miles in diameter. This preserves the protective screen when, breaking off a westward pursuit of the Japanese fleet, the ships turn east, into the wind, so that aircraft carriers can launch their aircraft.
After S.E. Morison, *The Two-Ocean War* (1963); Little, Brown and Company

slow march west across the Pacific. Having developed the matchless Long Lance torpedo, they installed it liberally in light cruisers and destroyers and developed tactics that would hurl a barrage of the long-range weapons in the direction of the enemy line—at the same time taking care not to expose the beams of their own ships to a counterstroke. Standard U.S. doctrine, on the other hand, called for fighting in column, employing guns as the primary weapon; the advantages that should have accrued to the Americans at night from superior radar were largely squandered. Between August 1942 and July 1943, in the cruiser–destroyer battles of Savo Island, Cape Esperance, Tassafaronga, Kula Gulf, and Kolombangara, Japanese night tactics prevailed. Not until mid-1943, with tactics attributed to Captain (later Admiral) Arleigh Burke that exploited the radar advantage in full, did the U.S. Navy redress the balance.

Still, naval aircraft were the weapons of decision. Although the duels of the great carrier fleets received more attention, air strikes from sea to shore were as crucial in securing control of the seas. Strikes by the British at Taranto, Italy (Nov. 11, 1940), by the Japanese at Pearl Harbor (Dec. 7, 1941), and by the Americans in the South Pacific at Rabaul (Nov. 5 and 11, 1943) and Truk (Feb. 17–18, 1944) were as important to that end as were the more sensational fleet engagements.

Also, in 1944 and 1945 the U.S. 3rd and 5th fleets, 27 fast carriers strong, took the war successfully against entire complexes of airfields in Formosa (now Taiwan), the Philippines, and Japan itself. A traditional tactical maxim, "Ships do not fight forts," was suspended for the duration of the war.

In the closing days of the war in the Pacific, the Battle of Okinawa served to indicate the nature of future combat at sea. By that time the U.S. Navy had reduced the Japanese Navy to impotence, and manned aircraft could not penetrate the sure American defenses. Nevertheless, during the three-month campaign for Okinawa (April–June 1945) the U.S. Navy lost 26 ships and suffered damage to 164 more—this time to Japanese kamikazes (suicide pilots) flying out of airfields in Japan. The pilots who flew these one-way missions were delivering, in effect, human guided missiles. Kamikazes showed that missiles could, on sufficient occasion, get through otherwise impenetrable defenses. The missile-guidance technology exhibited in the late stages of the war in Europe indicated that missiles would be the kamikazes of the future. And the atomic bomb offered the ugly threat of "one hit, one kill" at sea.

The age of the guided missile. At the end of World War II the supremacy of the U.S. Navy was as pronounced as that of the Royal Navy in the 19th century. With no enemy battle fleet to fight, it staked out the classic role of dominant navies throughout history—projecting its influence over land. Carrier-based aircraft, nuclear missile-carrying submarines, and amphibious-assault units extended that influence greatly. While the U.S. Navy served to link the North Atlantic Treaty Organization (NATO) across the Atlantic, its carrier-centred battle fleet stood ready to deliver sea power over the land. The principal opposing navy, that of the Soviet Union, was configured to challenge NATO's sea link and to confront U.S. aircraft carriers. The result was a new, asymmetrical tactical environment: a surface fleet facing a "fleet" composed mainly of submarines and land-based aircraft.

Modern naval warfare at the Falkland Islands

On a smaller scale than the U.S.-Soviet naval competition, the Falkland Islands War between the United Kingdom and Argentina in 1982 exhibited the tactical environment of sea-based forces fighting land-based forces in the guided-missile era. In this, the only extended naval campaign after World War II, were observed several modern influences on naval combat. First, submarines were formidable weapons, not only in the sinking of an obsolescent pre-World War II cruiser (the *General Belgrano,* formerly the USS *Phoenix*) by a nuclear-powered attack submarine (HMS *Conqueror*) on May 2 but, less obviously, in the harrying of the whole British fleet by one Argentine diesel-electric submarine. Second, the nature, if not the full extent, of the threat of modern air-launched antiship missiles was seen in two Argentine attacks, first against the destroyer HMS *Sheffield* (May 4) and then, after penetrating fleet defenses, the supply ship *Atlantic Conveyor* (May 25). Also, a land-to-sea missile struck and damaged the destroyer HMS *Glamorgan* (June 12), presaging more strikes from land in future maritime wars. Third, the British relearned lessons of damage control and ship survivability, while the Argentines found that aircraft armed only with unguided bombs were outclassed by ships with surface-to-air missiles. Fourth, and perhaps most fundamental, both sides saw the crippling effect of inadequate scouting, for both were without first-line sea-based air surveillance. Both had to manage with makeshift sources, such as picket submarines and commercial aircraft for conducting reconnaissance.

Despite the opposition, the British put forces ashore, maintained sufficient control of the airspace, and kept open the very long lines of supply, and this enabled them to retake the islands. A hasty prediction, made by some modern tacticians, that surface warships would be driven from the seas by modern missiles, did not prove true. Indeed, in the Falklands conflict, the recorded history of sea battles was reaffirmed. By their very nature, sea battles, once joined, still tended to be fast, deadly, and decisive. The commander of a fleet, always the most expensive component of an armed force, might, as Winston Churchill said of Jellicoe at Jutland, lose his ships and the war in an afternoon.

Electronic scouting

In response to growing weapon range, the collection and delivery of tactical information continued to grow in importance and consumed more manpower and facilities. Radar and electronic intelligence satellites, over-the-horizon radars, large surveillance aircraft, and electronic signal collectors of utmost sophistication were all manifestations of this trend. These scouting devices had their antitheses in electronic jammers and countermeasures—in effect, antiscouting systems.

In theory, modern communications have permitted the coordinated delivery of missiles or air strikes at great ranges from vessels in dispersed formations, and the three components of naval combat power—firepower, scouting, and C^2—can be highly dispersed. The major navies of the world, however, have continued to build aircraft carriers and cruisers, indicating a reliance on concentrated battle fleets and on strong defenses rather than dispersal to avoid destruction. The tactical value of concentrated as opposed to distributed power will ultimately depend on whether the historical trend observed above—that is, the growing range and lethality of naval weapons—continues. Battleships delivered salvos of gunfire in a continuous stream of destructive power, and the tactical effect was the N-square law of accumulating advantage. An aircraft carrier delivered a pulse of striking power that, if successful, destroyed about its own weight of the enemy. The classic naval tactic of attacking effectively first was vital. The question of the guided-missile age is whether one ship armed with missiles can sink more than one of the enemy, in spite of the enemy's defenses and ability to absorb punishment. If that is now the case, attacking first will be everything. Missiles that outrange the enemy's will be valuable, but even more valuable will be a compatible scouting system that detects and tracks the missiles close enough to their moving targets for the missiles' terminal guidance systems to lock onto them.

The swift naval engagements of the Arab-Israeli War of October 1973 are enlightening. In that war Syrian and Egyptian Osa- and Komar-class gunboats were armed with Russian-made SS-N-2 missiles, which outranged the Gabriel missiles carried by the Israeli Saar boats. Both fleets were small in numbers and size, but speedy. Based on relative missile range and the obvious sufficiency of firepower on both sides, the Arab boats should have struck first and won handily. The Israeli Navy, however, had recognized its disadvantage and had developed tactics that emphasized better scouting and C^2 as well as the use of chaff to deflect and neutralize the homing mechanisms of incoming missiles. This superior combination won decisively for them against both opponents. Therefore it could not be concluded that the advantage on paper of highly destructive and potentially decisive long-range missiles would win unless it was coupled with good sensors, modern C^2 systems, expert command, and sound training.

Guerrilla war at sea: the submarine. When submarines first went to sea early in the 20th century, they were immediately recognized as an extraordinary threat to surface ships. By World War II they were so effective against warships that they sank nearly as much aircraft carrier tonnage as was sunk by aircraft. Postwar attack submarines, nuclear-powered and armed with missiles and more advanced torpedoes, now pose an even greater threat to surface warships.

In both world wars, submarines were also a serious threat to merchant shipping. In World War II, German U-boats nearly severed the lifelines to Great Britain, U.S. submarines successfully isolated Japan by nearly wiping out its merchant fleet, and, in the Mediterranean, British and Axis submarines vied in attempts to cut their opponents' communication with North Africa.

Solo operation of the submarine

Barring a brief period in 1942–43, when U-boats operated successfully in so-called wolf packs, submarines have always been solo performers, relying for successful attack on concealment and surprise rather than concentration of force. These tactics, quite different from fleet actions, are akin to guerrilla war at sea. The submarine stalks its prey while the target—a warship, merchant ship, or convoy—seeks clues as to its presence in order to take evasive action. Simultaneously, antisubmarine forces—destroyers, maritime patrol aircraft, or helicopters—are predators and submarines the prey. The tactical competition between the two is all search and screening, deadly hide-and-seek, for when the submarine closes, its target can do naught but try to escape the blow, and when antisubmarine forces localize a submarine, no help will come and it will either have to fight like a cornered beast or go silent and try to slip away.

If submarines were able to protect sea lanes from attack, including air attack, then they would be serious candidates to succeed the aircraft carrier as the capital ship of the missile era. But as soloists their role has remained that of sea denial, not sea control, that of spoilers rather than champions of sea power. They are the latest in a long line of raiders of the deep, carrying on a tradition of isolation and stealth that began in the 16th century, when the English privateers Francis Drake and John Hawkins seized treasure ships at sea and raided Caribbean possessions in the teeth of the Spanish navy. Later, in the American Civil War, Confederate raiders such as Raphael Semmes in the *Alabama* harried Northern shipping despite the overwhelming superiority of the Union navy.

To the modern mind, a convoy has become a group of merchant ships protected against submarines. But, beginning in the age of fighting sail, there was a long tradition of protecting convoys against surface raiders, called "cruisers." In *Some Principles of Maritime Strategy* (1911), Sir Julian S. Corbett sorted out the separate roles of the battle fleet and the cruisers: the former established control of the seas by its concentrated presence or in a climactic battle; the latter either struck at lines of communication or attempted to fend off other raiders by operating alone or in small detachments. Corbett also traced the influence of long-range radio communication and predicted that this development would allow navies to bring such a swift concentration of superior power that the utility of surface raiders would come to an end. That he was right was proved by the fate of such surface raiders of World War II as the German battleship *Bismarck,* which was sunk by an overwhelming combination of bombs, naval guns, and torpedoes.

Corbett fully appreciated the major role submarines would play against capital ships, but he did not grasp the extent to which submarines would become the cruisers of the future. Indeed, submarines have become the biggest threat to commerce, ahead of mines and aircraft. In addition, they are at or near the top of the list of effective killers of submarines. Guerrilla war is therefore the apt term for submarine warfare. The battle tactics are dispersion, surprise, strikes where the enemy is weak and unprepared, disappearance into the vastness of the ocean, and a continuing erosion of enemy morale and dilution of his resources. (W.P.H.)

Air tactics

THROUGH WORLD WAR I

Powered aircraft were first used in war in 1911, by the Italians against the Turks near Tripoli, but it was not until the Great War of 1914–18 that their use became widespread. At first, aircraft were unarmed and employed for reconnaissance, serving basically as extensions of the eyes of the ground commander. Soon, however, the need to deny such reconnaissance to the enemy led to air-to-air combat in which each side tried to gain superiority in the air. Fighter planes were armed with fixed, forward-firing machine guns that allowed the pilot to aim his entire aircraft at the enemy, and the effective range of these weapons (no more than about 200 yards) meant that the first aerial combat took place at very short range.

Origin of fighter tactics

By the second year of the war fighter tactics emerged on all sides emphasizing basic concepts that, with modification, remained applicable through the jet age. First was the surprise attack; from the very beginning of aerial warfare in World War I, "jumping" or "bouncing" unsuspecting victims accounted for more kills than did the spectacular aerobatics of dogfighting. Because a pilot's only warning system was the naked eye, attacking fighters, whenever possible, approached from the rear or dove out of the sun, where they could not be seen. The German ace Max Immelmann, in exploiting the superior abilities of his Fokker Eindeker to climb and dive quickly, helped expand aerial combat from the horizontal into the vertical dimension. Immelmann developed what became known as the Immelmann turn, in which an attacking fighter dove past the enemy craft, pulled sharply up into a vertical climb until it was above the target again, then turned hard to the side and down so that it could dive a second time.

Fighters operated at least in pairs, flying 50 to 60 yards apart, so that the wingman could protect the leader's rear. Flying speed averaged 100 miles per hour, and communication was by hand signaling, rocking the wings, and firing coloured flares.

The next role to emerge for military aircraft was ground attack, in which planes, by strafing with machine guns and dropping rudimentary bombs, aided an advance on the ground, helped cover a retreat, or simply harrassed the enemy. By the late stages of the war, ground-attack aircraft had forced almost all large-scale troop movements to be carried out at night or in bad weather.

By war's end a fourth vision of air power arose—that of an independent air force attacking the enemy far from the front lines, the purpose being to destroy essential elements of the enemy's war capability by bombing factories, transportation and supply networks, and even centres of government. This role, never effectively implemented in World War I, was spurred largely by the German air attacks on London. Carried out at first by zeppelin airships, the bombing was later done by aircraft such as the Gotha bomber, which, by flying at night and often as high as 20,000 feet (forcing the crew to breathe bottled oxygen through a tube in the mouth), operated beyond the ceiling of many defensive fighters.

Thus, the basic roles that aircraft would play in modern war were presaged in World War I: reconnaissance, air superiority, tactical ground support, and strategic bombing.

THROUGH WORLD WAR II

The all-metal monoplane represented a huge increase in performance and firepower over the aircraft of World War I, and the effects were first seen in fighter tactics.

Air superiority. Airspeeds of the new fighters jumped to more than 400 miles per hour, and some planes could operate at altitudes of 30,000 feet. Wing-mounted machine guns and aerial cannon were lethal at 600 yards, and pilots communicated with one another and the ground via the radio telephone. These developments—especially the greater speeds—led Germans participating in the Spanish Civil War (1936–39) to fly their Me-109 fighters in loose, line-abreast *Rotten,* or pairs, about 200 yards apart. Two of these *Rotten* formed a *Schwarm,* and this flexible formation—called "finger-four" by English-speaking airmen—was eventually adopted by all the major air forces in World War II.

The fighting pair and the finger-four

Attacking out of the sun was still favoured, both because it preserved the element of surprise and because diving added speed. An alert defending fighter pilot, however, might use his attacker's speed to his own advantage by executing a scissors maneuver, in which he would turn sharply one way and then the other, reducing his forward motion so that the speeding attacker would overshoot and find the intended victim on his tail. Tight maneuvers such as the scissors were most effective when attempted with such agile fighters as the British Spitfire and the Japanese "Zero." Fighters such as the Me-109 and the U.S. P-47 Thunderbolt, which were noted for their speed, best escaped by diving hard and pulling back up when the attacker had been shaken.

A diving maneuver called the split-S, half-roll, or *Abschwung* was frequently executed against bombers. Heavily armed fighters such as the British Hurricane or the German Fw-190, instead of approaching from the side or from below and to the rear, would attack head-on, firing until the last moment and then rolling just under the big planes and breaking hard toward the ground. The object was to break up the bomber formations so that individual ships could be set upon and destroyed.

Defensive fighter squadrons were directed by radar control stations on the ground to the vicinity of the bombers, at which point the pilots would rely once more upon the naked eye. This was adequate for day fighting, when enemy bombers could be seen miles away, but at night the pilots had to get within a few hundred yards before spotting a bomber's silhouette against the sky or against the conflagration on the ground. For this reason, night fighting was ineffective until radar was installed in the planes themselves. This beginning of the age of electronic warfare required a novel teamwork between pilot and navigator, and it was best carried out in two-seat aircraft such as the British Beaufighter and Mosquito and the German Ju-88 and Me-110. Some of these long-range, twin-engined night fighters also served as "intruders," slipping into enemy bomber formations, following them home, and shooting them down over their own airfields.

Ground attack. The German Air Force, or Luftwaffe, was configured primarily to fly in support of ground forces, and, in the Spanish Civil War and the first years of World War II, the Ju-87 Stuka dive-bomber was its principal ground-attack craft. In a typical Stuka attack, several planes would circle above the target, then one plane after another would peel off to dive almost vertically before releasing its bombs, pulling up, and returning to the circle to dive again. In the Pacific Theatre, carrier-based dive-bombers such as the U.S. Dauntless and Helldiver and the Japanese Type 99 "Val" applied this maneuver to naval warfare. Dropping straight down from a cruising altitude of about 15,000 feet and releasing their bombs from below 2,000 feet, these planes destroyed or damaged many battleships and aircraft carriers. During the assault phase of amphibious landings, U.S. dive-bombers helped compensate for the flat trajectories of naval guns in disabling Japanese shore defenses. Because dive-bombers generally had top speeds in level flight of less than 300 miles per hour, they were most effective where air superiority had been secured by fighters such as the Zero or the U.S. F6F Hellcat. Spitfire pilots of the RAF made such short work of unescorted Stukas that they referred to these one-sided dogfights as "Stuka parties."

Dive-bombing

Ground attack was most devastating when conducted by fighter-bombers, which were often converted air-superiority fighters. Taking advantage of their speed, British Spitfires and Mosquitos and U.S. P-51 Mustangs and P-38 Lightnings, flying very low to avoid radar detection, bombed and strafed countless airfields and infantry columns. Pilots of the P-51, after escorting bombers into Germany (see below), often freely attacked ground targets while racing back to England at treetop level. In North Africa in 1942–43, the Royal Air Force (RAF) perfected close-air support by concentrating its air power under a centralized control that was exercised jointly by the senior ground and air commanders in the theatre of operations. This system, by concentrating maximum force at decisive points as the desert campaigns unfolded, achieved a flexibility of employment that later emerged as the central tenet of air power.

Strategic bombing. World War II saw massive bombing of military targets and major cities. The big, slow-moving bombers operated in formations (sometimes numbering 1,000 or more) that were intended not to evade enemy defenses but to beat them back or simply swamp them with numbers.

The key to bombing during the day was to provide an escort of fighters adequate to turn back defending fighters. (Antiaircraft artillery was of little hazard to bombers flying above 20,000 feet.) During the Battle of Britain (1940–41), a typical formation of German He-111 and Do-217 bombers would cross the English Channel at about 15,000 feet. Close escort would be provided by Fw-190s weaving in and out of the formation, while high and top cover would be provided by Me-109s stacked behind the bombers up to about 25,000 feet. The added height would enable the Me-109s to jump the RAF's Spitfires and Hurricanes while they were still climbing. Even more effective were fighter sweeps, in which Me-109s would leave the bombers and attack distant airfields before the defending fighters could get off the ground. But the Luftwaffe, in one of the major miscalculations of the aerial war, usually confined its fast, deadly fighters to a closer escort of the bomber formations.

The U.S. Army Air Force learned the value of fighter sweeps in its long-range daylight bombing of Germany, but not before placing an unfounded faith in the capacity of its B-17 Flying Fortress and B-24 Liberator bombers to defend themselves with their own heavy armament. In late 1942 and early 1943 these bombers began to fly in what became known as the "combat box" formation, de-

The defensive box formation

vised by Colonel (later General) Curtis E. LeMay. In such a formation, a single combat wing of about 48 bombers would be divided into three groups, with the lead group flying at 20,000 feet and the others trailing in echelon at intervals of 500 to 1,000 yards and at slightly higher altitudes. Within each group would be three squadrons, composed of two elements of three aircraft each, and the bombers would be staggered in such a way as to give their guns as free a field of fire as possible to cover themselves and their fellows.

The defensive formation was sorely tested in 1943, when, flying beyond the radius of the fighter escorts then available (less than 200 miles), U.S. bombers suffered losses too severe to be borne regularly. Activity over Germany was curtailed until the arrival in force the next year of P-51s equipped with droppable external fuel tanks that enabled the fighters to fly escort the 1,000 miles to Berlin. With enough fighters to allow one escort for every bomber, some P-51s were cut loose to sweep the airspace hundreds of miles away. In this way, the Luftwaffe was finally overwhelmed.

Night bombing relieved bombers of the fighter threat (at least until effective radar was installed in planes), but it presented difficulties in finding and hitting targets. With visual navigation impossible except on the clearest moonlit nights, electronic aids became vital. In the blitz of London and other cities, the Luftwaffe used a system called *Knickebein,* in which bombers followed one radio beam broadcast from ground stations on the continent until that beam was intersected by another beam at a point over the target. Lead bombers dropped incendiary bombs, which set fires that guided other bombers carrying high explosives as well as more incendiaries.

The RAF used two radar-beam systems called Gee and Oboe to guide its Lancaster and Halifax bombers to cities on the Continent. In addition, the bombers carried a radar mapping device, code-named H_2S, that displayed reasonably detailed pictures of coastal cities such as Hamburg, where a clear contrast between land and water allowed navigators to find the target areas. In order to "spoof" the Germans' radar warning system, RAF planes dispensed "window," which consisted of clouds of tinfoil strips that masked the bombers' movements.

Because Japan had no radar, U.S. B-29 Superfortresses did not face dangerous opposition in their long-range assaults on the Japanese home islands beginning in November 1944. Nevertheless, unpredictable weather over the target areas, plus the action of the jet stream on bombs dropped from 30,000 feet, made high-altitude bombing imprecise. In response, LeMay ordered low-level bombing runs. Flying at night to avoid enemy defenses, B-29s dropping incendiary bombs from 5,000 to 9,000 feet devastated more than 60 cities between March and July 1945.

THE JET AGE

Toward the end of World War II, the first operational jet fighter, the German Me-262, outflew the best Allied escorts while attacking bomber formations. This introduced the jet age, in which aircraft soon flew at more than twice the speed of sound (741 miles per hour at sea level and 659 miles per hour at 36,000 feet) and easily climbed to altitudes of 50,000 feet. At the same time, advanced electronics removed the task of early warning from the pilot's eye, and guided missiles extended the range of aerial combat, at least in theory, to beyond visual range.

Air superiority. Flying at supersonic or near-supersonic speeds, often climbing into the thin air of the stratosphere, jet fighters were far less maneuverable than their propeller-driven predecessors. This made necessary a formation even more flexible than the finger-four. One solution was the fluid-four, in which two fighters flying 300 yards apart would be trailed by another pair flying 2,000 to 3,000 yards to the side, 600 yards back, and 1,000 yards above. Separation of a mile or more would allow the trailing pair to cover the lead pair from surprise attack. The basic fighting element remained the pair, often favouring a "loose deuce" formula in which either pilot, depending upon the combat situation, could adopt the role of lead fighter while the other covered as wingman.

The fluid-four and loose deuce

Because jet fighters had excellent climbing but poor turning ability, fighting in the vertical plane became more important than ever. The scissors maneuver acquired a vertical variation, in which two fighters would execute a series of climbing turns or barrel rolls, each with the aim of slipping behind the plane that climbed too fast. Speed—usually the greatest asset of the fighter—could easily become a liability, and many maneuvers were developed to preserve its advantage. One such maneuver was the "high-speed yo-yo," in which an attacking fighter, in pursuing a more maneuverable opponent in a tight circle, would pull up while turning; this would reduce his speed, allowing him to remain within the circle while placing him in a position to swoop down from above.

Supersonic speed actually accounted for a tiny fraction of flying time, since igniting the jet's afterburner could consume a fighter's fuel in minutes. Military cruising speed was almost always subsonic, with the afterburner being used only for pursuit or escape. In fact, fuel became such a pressing concern in jet warfare that fighters often could spend no more time flying combat air patrol than they spent flying to and from the patrol area.

Suppression of air defense. Beginning in the 1960s, radar-directed antiaircraft weapons proved so dangerous that they threatened to sweep aircraft from the sky. By flying low and fast, jinking (making quick, irregular changes in direction and speed), or diving in a steep spiral, aircraft often succeeded in evading these weapons, but only at the price of spoiling the mission. Air defenses had to be destroyed; in order to do this, aircraft had not only to outfly and outgun the weapons but also to foil their guidance mechanisms with electronic countermeasures (ECM).

Electronic countermeasures

In the Vietnam War the North Vietnamese deployed a formidable air-defense system based on Soviet-made antiaircraft guns and SA-2 surface-to-air missiles (SAMs). In response, the U.S. Navy and Air Force mounted complex air strikes employing aircraft of multiple types and capabilities. One such operation might begin with F-4 Phantom II fighter-bombers entering the target area first to drop clouds of radar-reflecting metallic fibres called chaff. These would be followed by F-105 Thunderchiefs modified into "Wild Weasels" by the addition of radar homing and warning devices designed to jam some enemy radars and locate others. The Wild Weasels would guide other F-105s armed with radar-homing missiles, which would destroy the radars and SAM sites and clear the target area for the main strike force.

That air warfare in the jet age had effectively become electronic warfare was confirmed by the Arab-Israeli War of October 1973. In the first two days of that conflict, Israel lost 40 aircraft to Egyptian and Syrian air defenses. In June 1982, however, the Israeli air force displayed a new mastery of tactics in the electronic age by destroying Syrian SAM sites in al-Biqāʿ Valley, Lebanon. The attack began with a wide array of ECM equipment—Boeing 707s modified into electronic warfare aircraft, E-2C Hawkeye early warning aircraft, and A-4 Skyhawks flying reconnaissance—to confuse and deceive Syrian communications and the radars of Syrian SA-2, SA-3, SA-6, and SA-8 SAM units. Small remotely piloted vehicles were sent over the valley; when the Syrians fired on these, Israeli F-4s spotted the SAM sites and destroyed them with radar-homing missiles and cluster bombs. Israeli F-15 Eagles and F-16 Fighting Falcons then destroyed the Syrian air force, downing more than 80 MiG-21s and MiG-23s.

Strategic bombing. The importance of ECM in long-range bombing became apparent in 1972, when U.S. B-52 Stratofortresses struck targets in North Vietnam. By flying under escort at night and at about 30,000 feet, the B-52s were reasonably safe from MiG fighters and antiaircraft guns, and Wild Weasel and chaff-dropping aircraft helped suppress the SA-2s. But the most important ECM was provided by jammers built into the bombers. These flew in cells of three in order to create "blankets" of radar suppression that largely foiled the SAMs.

The next generation of variable-wing bombers, such as the U.S. B-1 and the Soviet Tu-26 Backfire, were designed to avoid more sensitive electronic warning systems by penetrating enemy airspaces at extremely low altitude.

Flying in groups was to be abandoned, since the large radar cross section and radio communication of several bombers would be easily detected. Instead, the new bombers were designed for solo missions and carried standoff weapons such as nuclear-armed cruise missiles, which could be launched beyond the range of SAMs guarding the target areas. (D.MacI.)

Logistics

FUNDAMENTALS

In the conduct of war, war-making activity behind the cutting edge of combat has always defied simple definition. The military vocabulary offers only a few general descriptive terms (such as administration, services, and the French *intendance*), all corrupted by loose usage and none covering the entire area of noncombat activity. All carry additional, though related, meanings that make them ambiguous.

Logistics belongs to this group. Its archaic meaning, the science of computation (from the Greek *logistikos,* "skilled in calculating"), persists in mathematics as the logistic or logarithmic curve but seems unrelated to modern military applications. In the 18th century it crept into French military usage with a variety of meanings, including "strategy" and "philosophy of war." But the first systematic effort to define the word with some precision and to relate it to other elements of war was made by Antoine-Henri Jomini (1779–1869), the noted French military thinker and writer. In his *Summary of the Art of War* (1838), Jomini defined logistics as "the practical art of moving armies," by which he evidently meant the whole range of functions involved in moving and sustaining military forces—planning, administration, supply, billeting and encampments, bridge and road building, even reconnaissance and intelligence insofar as they were related to maneuver off the battlefield. In any case, Jomini was less concerned with the precise boundaries of logistics than with the staff function of coordinating these activities. The word, he said, was derived from the title of the *major général* (or *maréchal*) *des logis* in French 18th-century armies, who, like his Prussian counterpart, the *Quartiermeister,* had originally been responsible for the administrative arrangements for marches, encampments, and troop quarters (*logis*). These functionaries became the equivalent of chiefs of staff to the commanders of the day.

"The practical art of moving armies"

Jomini's discussion of logistics was really an analysis of the functions of the Napoleonic general staff, which he conceived as the commander's right arm, facilitating his decisions and seeing to their execution. The mobility and gargantuan scale of Napoleonic warfare had left the simple old logistics of marches and encampments far behind. The new logistics, said Jomini, had become the science of generals as well as of general staffs, comprising all functions involved in "the execution of the combinations of strategy and tactics."

This broad conception had some validity in Jomini's day. He left an engaging picture of Napoleon, his own logistician, sprawled on the floor of his tent, marking each division's route of march on the map with a pair of dividers. But as staff organization and activity became more complex, along with war itself, the term logistics soon lost its association with staff activity and almost disappeared from the military vocabulary. Jomini's great contemporary, the Prussian theorist Carl von Clausewitz, did not share his conception of logistics, which he called "subservient services" that were not part of the conduct of war. Jomini's own influence, which was enormous in his day, was mainly on strategic and tactical thought, particularly in the American Civil War.

In the late 1880s the American naval historian Alfred Thayer Mahan introduced logistics into U.S. naval usage and gave it an important role in his theory of sea power. In the decade or so before World War I the navy's concern with the economic foundations of its expansion began to broaden the conception of logistics to encompass industrial mobilization and the war economy. Reflecting this trend, a U.S. marine officer, Lieutenant Colonel Cyrus Thorpe, published his *Pure Logistics* in 1917, arguing that the logical function of logistics, as the third member of the strategy–tactics–logistics trinity, was to provide all the means, human and material, for the conduct of war, including not merely the traditional functions of supply and transportation but also war finance, ship construction, munitions manufacture, and other aspects of war economics.

Providing all the means for the conduct of war

After World War II the most notable effort to produce a theory of logistics was by a retired rear admiral, Henry E. Eccles, whose *Logistics in the National Defense* appeared in 1959. Expanding Thorpe's trinity to five (strategy, tactics, logistics, intelligence, communications), Eccles developed a conceptual framework that envisaged logistics as the military element in the nation's economy and the economic element in its military operations—that is, as a continuous bridge or chain of interdependent activities linking combat forces with their roots in the national economy. Eccles stressed the tendency of logistic costs to rise (the logistic "snowball") and, echoing Jomini, the essential role of command. Despite its logic and symmetry, however, Eccles' overarching conception of logistics was not widely accepted. Official definitions still vary widely, and most ordinary dictionaries adhere to the traditional "supply, movement, and quartering of troops," but neither has much influence on common usage, which remains stubbornly inconsistent and loose.

Components of logistics. It is useful to distinguish four basic elements or functions of logistics: supply, transportation, facilities, and services. (A fifth, management or administration, is common to all organized human activity.) All involve the provision of needed commodities or assistance to enable armed forces to live, move, communicate, and fight.

Supply. Supply is the function of providing the material needs of military forces. The supply process embraces all stages in the provision and servicing of military material, including those preceding its acquisition by the military—design and development, manufacture, purchase and procurement, storage, distribution, maintenance, repair, salvage, and disposal. (Transportation is, of course, an essential link in this chain.) The whole process can be divided into four phases: (1) the design–development–production process of creating a finished item, (2) the administrative process by which military agencies acquire finished items, (3) the distribution–servicing processes undergone by military material while "in the service," and (4) the planning–administrative process of balancing supply and demand—that is, the determination of requirements and assets and the planning of production, procurement, and distribution.

Military supply has always had the basic aim of providing military forces the material needed to live (food, water, clothing, shelter, medical supplies), to move (vehicles and transport animals, fuel and forage), to communicate (the whole range of communications equipment), and to fight (weapons, defensive armament and materials, and the expendables of missile power and firepower). In all these categories are items, such as clothing, vehicles, and weapons, that are used repeatedly and therefore need to be replaced only when lost, destroyed, or worn out; and materials, such as food, fuel, and ammunition, that are expended or consumed—that is, used only once—and therefore must be continuously or periodically resupplied. From these characteristics are derived the basic classifications of initial issue, replacement, and resupply. The technical classifications of supply vary among countries and services. The British army, for example, recognizes two broad classes: (1) supplies, which include all the expendables except ammunition, and (2) stores, which include ammunition and military hardware. The U.S. Army in World War II and for many years after used five main classifications: (1) subsistence and forage, (2) equipment and other items regularly issued to organizations and individuals, (3) fuels, (4) equipment and materials of irregular issue such as construction materials, and (5) ammunition. These five classes were subsequently expanded to 10 by designating as separate classes certain large categories, such as vehicles, medical material, repair parts, and sales items, which formerly were considered as subclasses.

Historically, food and forage made up most of the bulk

The two principal components of supply: food and fuel

and weight of supply until the 20th century, when, with mechanization and air power, fuel displaced forage and became the principal component of supply. However, the demand for food remains unremitting and undeferrable, the one constant of logistics. A man's daily ration makes a small package—seven pounds and often much less. But an army of 50,000 may consume in one month as much as 4,500 tons (4.1 million kilograms) of food.

Animals require much more. The standard grain and hay ration in the 19th century was about 25 pounds (11.4 kilograms), and the daily forage of a corps of 10,000 cavalry weighed as much (allowing for remounts) as the food for 60,000 men. Forage requirements tended, moreover, to be self-generating, since the animals needed to transport it also had to be fed. The number of animals accompanying an army varied widely. Napoleon's ideal, which he himself never attained, was a supply train of only 500 wagons in an army of 40,000; with a corps of 7,000 cavalry, this would amount to about 10,000 animals exclusive of remounts and spare draft animals. Northern armies in the American Civil War commonly numbered half as many animals as soldiers. A force of 50,000 men might thus require more than 300 tons (272,000 kilograms) of forage daily. This was more than twice the weight of gasoline that an equivalent force of three World War II infantry divisions, using motor vehicles exclusively, needed to operate for the same length of time. In the latter case, moreover, fuel requirements diminished markedly when an army was not moving, whereas the premechanized force had to feed its animals whether moving or not. It was the immense forage requirements of premechanized armies, more than any other single factor, that restricted warfare before the 20th century so generally to seasons and climates when animals and men could subsist mainly on the countryside.

In 20th-century warfare the expendables of movement include fuel for rail and water transport as well as for motor vehicles, and also the immense fuel requirements of modern air power. In World War II, without counting transoceanic shipment, fuel made up half the resupply and replacement needs of U.S. forces in Europe. Technologically advanced warfare has, in fact, vastly increased fuel consumption both absolutely and relatively to other supply needs. The continued development of mechanization and air power has increased by one and one-half times the fuel requirements of large-scale conventional military operations typical of World War II. Food, by contrast, is a small and diminishing fraction of the total burden.

Before the 20th century, equipment replacement and ammunition resupply were a relatively small part of an army's needs. Missile power before the gunpowder era was limited by the difficulty of bringing missiles in quantity to the battlefield. For the first five centuries of the gunpowder era the provision of ammunition was not a major logistic problem. Not until the use of field artillery on a large scale in the late 18th century, and the development of quick-firing shoulder arms in the 19th, did ammunition begin to constitute a substantial part of resupply needs. As late as 1864, in the Atlanta campaign of the American Civil War, the Union army's average daily ammunition requirements amounted to only one pound (0.45 kilogram) per man, as against three pounds for rations; Confederate forces in that war were reported to expend, on the average, only half a cartridge per man per day.

Supplying ammunition

The great increase in firepower in the 20th century upset the historic ratios. In World War II the average ammunition requirements of Western forces in combat zones were 12 percent of total needs. In the mainly positional Korean War, ammunition expenditures climbed higher, and a late-1980s U.S. Army planning factor rated ammunition requirements as more than one-quarter of total supply. Material replacement needs have also mounted in absolute terms; the great tank battles of World War II and of the Arab-Israeli Wars of 1967 and 1973 involved the destruction of hundreds of tanks within a few days. But as a percentage of total supply, replacement of material losses is a declining factor.

Transportation. Before the development of steam propulsion, armies depended for mobility on the muscles of men and animals and the force of the wind. On land they used men and animals to haul and carry; on water they used oar-driven and sail-propelled vessels. Among these various modes the balance of advantage was often delicate. A force moving by water was vulnerable to storm and enemy attack; navigation was an uncertain art; transports were expensive and of limited capacity. Large expeditions could be undertaken only by wealthy states or seafaring peoples, such as the Scandinavians of the 8th and 9th centuries, who combined the roles of mariner and warrior. Seaborne armies were rarely strong enough to overcome a resolute land-based foe.

The advantages of sea and air transport

On the other hand, armies have usually been able to move faster and with a better chance of avoiding enemy detection by water than by land. Shipment of bulky freight is cheaper and safer by river than by road, and good roads are rare in military history. In the 19th and 20th centuries the revolution in ship design and propulsion made water travel largely independent of wind and weather, permitting the overseas movement and support of larger forces than ever before. After the mid-19th century, however, more and better roads and, above all, railroads began to offset the historic advantages of water transportation to some degree. In the 20th century motor vehicles and more road building extended the conquest of rough terrain. The airplane finally freed military movement, for modest forces and limited cargo, from bondage to earth altogether. Yet the costs of mobility on land—in equipment, materials, and energy—remain high, and large military movements are still confined to narrow ribbons of rail and road, which in many parts of the world are still rare or lacking.

On land the soldier himself has been the basic burden carrier of armies. As a matter of simple economy, he represents large carrying capacity at no extra cost. His equivalent, in an army of 50,000 in the preindustrial era, would be 1,875 wagons drawn by 11,250 horses or mules, which might need additional wagons and animals to haul forage. A difference of only five pounds (2.3 kilograms) in the soldier's load could add or subtract a requirement for 125 wagons and 750 animals. Since the days of the Roman legion, the soldier has had to carry, on the average, about 55 or 60 pounds (25 or 27 kilograms). The ratio between weapons and other items in the soldier's load has varied widely, but the modern soldier has relegated most of his food to vehicle transport while still carrying a heavy burden of weapons and ammunition. Since World War II, however, some armies have made drastic reductions in the combat load.

Human and animal haulers

Before the age of mechanization, the soldier's carrying capacity was usually supplemented by additional carriers and haulers, human and animal. Each had advantages. A team of six horses ate about as much as 30 to 40 men, but the men could carry more on their backs than the horses could haul and considerably more than the horses could carry. Men could negotiate rougher terrain, and they required less care. On the other hand, loads placed on men had to be distributed in small packages, and men proved less efficient than animals when teamed to haul heavy and bulky loads. The horse and mule, however, have less strength and stamina, though more agility, than the ox, history's primary beast of burden. In many parts of the world, motor transport still has not displaced human and animal carriers and haulers in the movement of military supply.

Facilities. The provision of military facilities, as distinct from fortification, did not become a large and complex sphere of logistic activity until the transformation of warfare in the industrial era. In that transformation the traditional function of providing nightly lodgings or winter quarters for the troops dwindled to relative insignificance in the mushrooming infrastructure of fixed and temporary installations that became part of the military establishments of the major powers. Modern armies, navies, and air forces own and operate factories, arsenals, laboratories, power plants, railroads, shipyards, airports, warehouses, supermarkets, office buildings, hotels, hospitals, homes for the aged, schools, colleges, and many other types of structures used by advanced societies in the 20th century—as well as barracks, the original military facility. They are among the world's great landowners. The management of

all this improved real estate is one of the largest areas of modern logistic administration.

Services. Services may be defined as activities designed to enable personnel or material to perform more effectively. Usage recognizes no clear distinction between logistic and nonlogistic services, but a somewhat blurred one has grown out of the traditional and opprobrious identification of logistics with noncombat rear-area activities. Thus, intelligence and communications personnel and combat engineers in the U.S. Army have long claimed the label of "combat support" as distinct from the "service support" functions of supply, transportation, hospitalization and evacuation, military justice and discipline, custody of prisoners of war, civil affairs, personnel administration, and nontactical construction (performed by "construction" engineers). Training of combat troops is hardly ever considered a logistic service, whereas training of service troops sometimes is. Usage does not, however, always assign "service support" to logistics. Personnel administration is an old, institutionalized sector of the military establishment, and personnel administrators tend to reject the logistics label. Personnel services (medical, spiritual, educational, financial) are more heterogeneous and have varied origins; most definitions of logistics include them.

Most service activities, logistic and nonlogistic, are of recent origin and, as organized specialities, are peculiar to the military establishments of advanced nations. Over the long haul of military history, the services considered necessary to keep armed forces in fighting trim were generally of a rudimentary character. From the earliest times, however, they posed a serious logistic problem. To armies and their lines of communication they added numbers of people who did not, as a primary function, belong to the fighting force and who, if not properly organized, might weaken its capacity to fight. Soldiers seldom possessed the technical skills required to perform any but the simplest services; sometimes, as members of a warrior elite, they were prohibited by social prerogative from performing them. A classic feature of armies, consequently, has been its long train of noncombatants, often far outnumbering the fighting men.

Logistic problems in providing services

Logistic services also added to the baggage of armies a growing burden of specialized equipment, tools, and materials needed for the performance of the services. Services tended to generate more services: service equipment itself had to be serviced, sometimes by additional technicians, and service personnel themselves required services. Logistic services thus meant more people to be fed, clothed, and sheltered and more people and baggage to be transported. What the British call the "administrative tail" is as old as military history.

Special features of naval logistics. From early times, the substantial carrying capacity of the warship made it an indispensable element in its own logistic support, particularly in the era before steam power eliminated the problem of covering long distances between ports. (Oar-driven warships, such as the Greek trireme, sacrificed this feature in order to maximize fighting power.) For centuries the most critical item of supply was water, which sailing ships found difficult to carry in sufficient quantities and to keep potable for long voyages. Food was somewhat less of a problem, except for its notoriously poor quality in the days before refrigeration, the sealed container, and sterilization.

During the long reign of the sailing ship, the absence of a fuel requirement was a major factor in the superior mobility of fleets over armies. The shift to steam was, in a sense, a return to the principle of self-contained propulsion earlier embodied in the oar-driven ship. The gain in control was of course an immeasurable improvement for the long haul, but for a time the inordinate amount of space that had to be allocated to carry wood or coal seriously inhibited the usefulness of early warships. Eventually the maritime nations established networks of coaling stations, which became part of the fabric of empire in the late 19th century. The shift to oil a few years before World War I involved a major dislocation in naval logistics and changed the stakes of imperial competition.

The shift from sail to steam

For modern navies the importance of bases goes far beyond the need for periodic replenishment of fuel, although this remains essential. Ships must be repaired, overhauled, and resupplied with ammunition and food; and, an ancient requirement, the crews must be given shore leave. Within limits, these needs can be filled by specialized auxiliary ships either accompanying naval forces at sea or stationed at predetermined rendezvous points. Naval operations in World War II saw a proliferation of these auxiliary vessels; in 1945 only 29 percent of the U.S. Navy consisted of purely fighting ships. By using auxiliaries and by rotating ships and personnel, modern fleets can remain at sea indefinitely, especially if not engaged in combat. U.S. fleets in the Mediterranean and far Pacific have done so for years, although the feat is less impressive than that of the British admiral Lord Nelson's fleet, which lay off Toulon, Fr., continuously, without rotation, for 18 months from 1803 to 1805. With nuclear propulsion, thus far applied only to submarines and a handful of large warships, the basic logistic function of replenishing fuel may eventually disappear. But that day will be long in coming, and the other functions of naval logistics will remain.

Power versus movement. The potential effectiveness of a military force derives from three attributes: fighting power, mobility, and range of movement. Which of these attributes is stressed depends on the commander's objectives and strategy, but all must compete for available logistic support. Three methods have been used, in combination, in providing this support for forces in the field: self-containment, local supply, and supply from bases.

Self-containment. The idea of complete independence from external sources of supply—the hard-hitting, self-contained "flying column"—has always been alluring but has seldom fully materialized. Self-containment in weapons, equipment, and missiles or ammunition was common enough before the great expansion of firepower and resupply requirements in the last century. But few military forces have been able to operate for long or move far without frequent resupply of food and forage or fuel.

Self-containment is the least economical of all methods of supply. Accompanying transport is fully employed only at the beginning of the movement, serving thereafter as a rolling warehouse that is progressively depleted as the force moves. Fast-moving, self-contained forces typically left a trail of abandoned vehicles and dead animals. The basic trade-off in self-containment is between the speed gained by avoiding delays and detours for foraging and the speed lost by dragging a large baggage train. When Hannibal crossed the Alps into northern Italy in 218 BC, he bypassed the Roman army guarding the easier coastal route; but his movement through the mountain passes was painfully slow, and he lost almost half his force to cold, disease, and hostile tribes along the way.

Disadvantages of self-contained forces

Local supply. Until the 20th century, armies commonly lived off the country and, in enemy territory, from captured stores. In fertile regions an army could usually provision itself at low cost in transport and without sacrificing fighting power or range; when efficiently organized, local supply even permitted a high degree of mobility. Normally, however, an army living off the country tended to straggle and to load itself down with loot. If it moved too slowly or was pinned down, it might sweep the region bare and starve. In winter, in deserts and mountains, or in thinly populated areas, local supply offered meagre fare. And a hostile population, as Napoleon discovered in Russia and Spain, could bring disaster to an army that had to scrounge for its food. (British forces in the American colonies during the Revolution had to draw most of their supplies from overseas.) Animals, in any case, almost always had to shift for themselves. Cattle driven with an army could transform forage into food, a supply technique as ancient as the Bible and still common in the 19th century. Unwieldy and slow-moving though it was, the accompanying herd had the great merit of transporting itself and dwindling as it was consumed.

When mechanized transport replaced animals, one of the great continuities of military history was broken. Mechanized armies can operate in winter and desert areas as long as they have fuel; when that runs out, they grind to a

halt. Until fuel can be compressed into small capsules (as, in a sense, atomic energy is) or, like forage, be gathered along the way, the door to both self-containment and local supply will remain closed.

Supply from bases. The alternative to self-containment and local supply is continuous or periodic resupply and replacement from stores prestocked at bases or other accessible points. Supply from bases involves three serious disadvantages. First, supply routes are often vulnerable to attack. Second, an army shackled to its bases lacks flexibility and moves slowly—even more slowly as it advances. Finally, the transportation costs of maintaining a flow of supply over substantial distances are heavy and, beyond a point, prohibitive. The reason is twofold; first, because the transport of the supply train must operate a continuous shuttle—that is, for each day's travel time, two vehicles are needed to deliver a single load—and, second, because additional food and forage or fuel must be provided for the personnel, animals, or vehicles of the train itself. In the era of animal-drawn transport this multiplier factor set practical limits to the operating radius of an army, which the American Civil War general William T. Sherman fixed at about 100 miles (160 kilometres), or five days' march, from its base. The critical limitation was the provision of forage, the bulkiest supply item. For an army operating at any considerable distance from its bases, the in-transit forage requirements of its shuttling supply train, if supplied entirely from bases, would saturate any amount of transport, leaving none to supply the fighting force. Since premechanized armies usually found some local forage and food, supply from bases, in combination with local supply and an accompanying train, was the normal method, but Sherman's 100 miles was seldom exceeded.

The tendency of supply trains to multiply

With modern mechanized transport the theoretical maximum operating radius is so great that other limitations come into play. Nevertheless, the in-transit fuel needed to supply a force from distant bases adds major increments of transport cost, especially under conditions (*e.g.,* poor roads) that reduce speed or increase fuel consumption. It can also severely limit the speed of an advancing mechanized force, as shown by the bogdown of the U.S. 3rd Army's drive across France in the summer of 1944 for lack of fuel.

HISTORICAL DEVELOPMENT

Logistic systems before 1850. In ancient history the combination of local supply for food and forage and self-containment in hardware and services appears often as the logistic basis for operations by forces of moderate size. Some of these operations are familiar to many a schoolchild—the long campaign of Alexander the Great from Macedonia to the Indus, the saga of Xenophon's Ten Thousand, Hannibal's campaigns in Italy. The larger armies of ancient times—like the Persian invaders of Greece in 480 BC—seem to have been supplied by depots and magazines along the route of march. The Roman legion combined all three methods of supply in a marvelously flexible system. The legion's ability to march fast and far owed much to superb roads and an efficiently organized supply train, which included mobile repair shops and a service corps of engineers, artificers, armourers, and other technicians. Supplies were requisitioned from local authorities and stored in fortified depots; labour and animals were drafted as required. When necessary, the legion could carry in its train and on the backs of its soldiers up to 30 days' supply of provisions. In the First Punic War against Carthage (264–241 BC), a Roman army marched an average of 16 miles (26 kilometres) a day for four weeks.

Combining self-containment and local supply

One of the most efficient logistic systems ever known was that of the Mongol cavalry armies of the 13th century. Its basis was austerity, discipline, careful planning, and organization. In normal movements the Mongol armies divided into several corps and spread widely over the country, accompanied by trains of baggage carts, pack animals, and herds of cattle. Routes and campsites were selected for accessibility to good grazing and food crops; food and forage were stored in advance along the routes of march. On entering enemy country, the army abandoned its baggage and herds, divided into widely separated columns, and converged upon the unprepared foe at great speed from several directions. In one such approach march a Mongol army covered 180 miles (290 kilometres) in three days. Commissariat, remount, and transport services were carefully organized. The tough and seasoned Mongol warrior could subsist almost indefinitely on dried meat and curds, supplemented by occasional game; when in straits, he might drain a little blood from a vein in his mount's neck. Every man had a string of ponies; baggage was held to a minimum, and equipment was standardized and light.

In the early 17th century, King Gustav II Adolf of Sweden and Prince Maurice of Nassau, the military hero of the Netherlands, briefly restored to European warfare a measure of mobility not seen since the days of the Roman legion. This period saw a marked increase in the size of armies; Gustav and his adversaries mustered forces as large as 100,000, Louis XIV of France late in the century even more. Armies of this size had to keep on the move to avoid starving; as long as they did so, in fertile country they could usually support themselves without bases, even with their customary huge noncombatant "tail." Logistic organization improved, and Gustav also reduced his artillery train and the size of guns. In the Thirty Years' War (1618–48) strategy tended to become an appendage of logistics as armies, wherever possible, moved and supplied themselves along rivers exploiting the economies of water transportation, and operated in rich food-producing regions.

After the Thirty Years' War, European warfare became more sluggish and formalized, with limited objectives and an elaborate logistics that sacrificed both range and mobility. The new science of fortification made towns almost impregnable while enhancing their strategic value, making 18th-century warfare more an affair of sieges than of battles. Two logistic innovations were notable: the magazine, a strategically located prestocked depot, usually established to support an army conducting a siege; and its smaller, mobile version, the rolling magazine, which carried a few days' supply for an army on the march. Secure lines of communication became vital, and whole armies were deployed to protect them. The increasing size of armies and of artillery and baggage trains placed heavier burdens on transport. Also, a revulsion against the depredations and inhumanity of the 17th-century religious wars resulted in curbs on looting and burning and in regulated requisitioning or purchase of provisions from local authorities. Because of the high cost of mercenary soldiery, commanders tended to avoid battles, and campaigns tended to become sluggish maneuvers aimed at threatening or defending bases and lines of communication. "The masterpiece of a successful general," Frederick the Great remarked, "is to starve his enemy."

The magazine and rolling magazine

The era of the French Revolution and the Napoleonic domination of Europe (1789–1815) brought back both mobility and range of movement to European warfare, along with an immense further increase in the size of armies. Abandoning the siege warfare of the 18th century, Napoleonic strategy stressed swift offensives aimed at smashing the enemy's main force in a few decisive battles. The logistic system inherited from the Old Regime proved surprisingly adaptable to the new scale and pace of operations. Organization was made more efficient, baggage trains were pared down and some of their load shifted to the soldier's back, and much of the noncombatant tail was eliminated. The artillery train was increased, and the rolling magazine was used as the occasion demanded. The heavily burdened citizen-soldier marched faster and farther than his mercenary predecessor. In densely populated and fertile regions, moving armies continued to subsist, by purchase and requisition, on the countryside through which they marched, spreading out over parallel roads, each corps foraging to one side only. Even so, the numbers involved dictated greater dependence on magazines.

Napoleon made relatively few logistic innovations. He militarized some services formerly performed by contractors and civilian personnel, but the supply service (*intendance*) remained civilian though under military control. A significant change was the establishment in 1807 of a fully

militarized train service to operate over part of the line of communication; this was divided into sections that were each serviced by a complement of shuttling wagons—foreshadowing the staged resupply system of the 20th century. The 600-mile advance of Napoleon's Grande Armée of 600,000 men into Russia in 1812 involved logistic preparations on an unprecedented scale. Despite extensive sabotage by the Russian peasantry, the system brought the army victorious to Moscow.

Logistics in the industrial era. *The revolution in warfare.* Between the mid-19th and the mid-20th centuries the conditions and methods of logistics were transformed by a fundamental change in the tools and modes of making war—perhaps the most fundamental change since the beginning of organized warfare. The revolution had four facets: (1) the mobilization of mass armies; (2) a revolution in weapons technology involving a phenomenal increase in firepower; (3) an economic revolution that provided the means to feed, arm, and transport mass armies; and (4) a revolution in the techniques of management and organization, which enabled nations to operate their military establishments more effectively than ever before.

The huge growth of destructive power

These interrelated developments did not occur all at once. Armies of unprecedented size had appeared in the later years of the Napoleonic Wars. But for almost a century after 1815, the world saw no comparable mobilization of manpower except in the American Civil War. Meanwhile, the growth of population (in Europe, from 180 million in 1800 to 490 million in 1914) was creating a huge reservoir of manpower. By the end of the 19th century most nations were building large standing armies backed by even larger partially trained reserves. In the world wars of the 20th century the major powers mobilized armed forces numbering millions.

The revolution in weapons had started earlier but accelerated after about 1830. By the 1850s and '60s the rifled percussion musket, rifled and breech-loading artillery, large-calibre ordnance, and steam-propelled armoured warships were all coming into general use. The revolution proceeded with gathering momentum thereafter, but it remained for mass armies in the 20th century to realize its full potential for destruction.

By the mid-19th century the Industrial Revolution had already given Great Britain, France, and the United States the capacity to produce munitions, food, transport, and many other items in quantities no commissary or quartermaster had ever dreamed of. But except in the Northern states during the American Civil War, the wars of the 19th century hardly scratched the surface of the existing war-making potential. The nature of international rivalries of the period tended to limit war objectives and the mobilization of latent military power. Only in the crucible of World War I, at the cost of colossal blunders and wasted effort, did nations begin to learn the techniques of "total" war. Long before 1914, however, new instruments and techniques of logistics were emerging.

Transportation and communication. The railroad, the steamship, and the telegraph had a profound impact on logistic method during the last half of the 19th century. Beginning with the Crimean War (1854–56), telegraphic communication became an indispensable tool of command, intelligence, and operational coordination, particularly in controlling rail traffic. In the 20th century it yielded to more efficient forms of electronic communication—the telephone, radio, radar, television, telephotography, and the high-speed computer.

Railroads and mass mobilization

Railroads spread rapidly over western and central Europe and the eastern United States between 1850 and 1860. They were used—mainly for troop movements—in the suppression of central European revolutions in 1848–49, on a considerable scale in the Italian War of 1859, and extensively in the American Civil War, where they also demonstrated their capacity for long hauls of bulky freight in sustaining the forward movement of armies. In Europe, from 1859 on, railroads shaped the war plans of all the general staffs, the central features of which were the rapid mobilization and concentration of troops on a threatened frontier at the outbreak of war. In 1870, at the outset of the Franco-German War, the German states were able to concentrate 550,000 troops, 150,000 horses, and 6,000 pieces of artillery on the French border in 21 days. Germany's recognized efficiency in mobilizing influenced the war plans of all the European powers in 1914. In both world wars Germany's railroads enabled it to shift troops rapidly between the Eastern and Western fronts.

Steam propulsion and iron ship construction also introduced new logistic capabilities into warfare in the 19th century. Steamships moved troops and supplies in support of U.S. forces in the Mexican War of 1846–48 and of British and French armies in the Crimea. River steamboats played an indispensable role in the American Civil War.

Motor transport

The complement of the railroad was the powered vehicle that could travel on ordinary roads and even unprepared surfaces, within the operating zones of armies forward of railheads. This was a 20th-century development, a combination of the internal-combustion engine, the pneumatic tire, and the endless track. Motor transport was used on an increasing scale in both world wars, although animal-drawn transport and railroads still dominated land movement. Another innovation was the pipeline, used to move water in the Palestine campaign of World War I and extensively in World War II to move oil and gasoline to storage points near the combat zones. More revolutionary was the development of large-scale air transportation. In World War II, units as large as a division were carried in one movement by air over and behind enemy lines and resupplied by the same means. Cargo aircraft maintained an airlift for more than three years from bases in India across the Himalayas into China; during the last eight months of operation it averaged more than 50,000 tons per month. But the fuel costs of such an operation were exorbitant. Air transportation remained primarily a means of emergency movement when speed was an overriding consideration.

The growth in quantity. The most conspicuous logistic phenomenon of the great 20th-century wars was the enormous quantity of material used and consumed. One cause was the growth of firepower, which was partly a matter of increased rapidity of fire of individual weapons, partly a higher ratio of weapons to men—both multiplied by the vast numbers of troops now mobilized. An American Civil War infantry division of 3,000 to 5,000 men had an artillery complement of up to 24 pieces; its World War II counterpart, numbering about 15,000 men, had 328 artillery pieces, all capable of firing heavier projectiles far more rapidly. A World War II armoured division had nearly 1,000 pieces of artillery. Twentieth-century infantrymen, moreover, were armed with semiautomatic and automatic weapons.

The increasing consumption of ammunition

The upward curve of firepower was reflected in the immense amounts of ammunition required in large-scale operations. Artillery fire in the Franco-German War and in the Russo-Japanese War (1904–05), for example, showed a marked increase over that in the American Civil War. But World War I unleashed a firepower hardly hinted at in earlier conflicts. For the preliminary bombardment (lasting one week) in the First Battle of the Somme in 1916, British artillery was provided 23,000 tons of projectiles; 100 years earlier, Napoleon's gunners at Waterloo had about 100 tons. In World War II the United States procured only about four times as many small arms as it had in the Civil War but 43 times as much small-arms ammunition. (To the ammunition expenditures in World War II were added, moreover, the immense tonnages of explosives used in air bombardment.) The Confederacy fought through the four years of the Civil War on something like 5,000 or 6,000 tons of gunpowder, whereas U.S. factories in one average month during World War I turned out almost four times this quantity of smokeless powder. Again, in one year of World War II, seven million tons of steel went into the manufacture of tanks and trucks for the U.S. Army, four million tons into artillery ammunition, one million tons into artillery, and 1.5 million tons into small arms—as contrasted with less than one million tons of pig iron used by the entire economy of the Northern states during one year of the Civil War.

With quantitative growth went a parallel growth in the complexity of military equipment. The U.S. Army in

World War II used about 60 major types of artillery above .60-inch calibre; for 20 different calibres of cannon there were about 270 types and sizes of shells. The list of military items procured for U.S. Army ground forces added up to almost 900,000, each of which contained many separate parts—as many as 25,000 for some antiaircraft guns. To convert and expand a nation's peacetime industry to the production of such an arsenal posed staggering technical problems. Manufacturers of automobiles, refrigerators, soap, soft drinks, bed springs, toys, shirts, and microscopes had to learn how to make guns, gun carriages, recoil mechanisms, and ammunition.

Staged resupply. Long before mechanization relegated local supply to a minor role in logistics, growing supply requirements were making armies more dependent on supply from bases. The *Etappen* system of the Prussian army in 1866 resembled the Napoleonic train service of 1807. Behind each army corps trailed a lengthening series of shuttling wagon trains moving up supplies through a chain of magazines extending back to a railhead. A small train accompanied the troops, carrying a basic load of ammunition, rations, and baggage; each soldier also carried additional ammunition and three days' emergency rations. The system was geared to a steady, slow advance on a rigid schedule and a predetermined route.

Shuttling supplies along a chain of magazines

Before the advent of mechanization half a century later, the system did not work well, since the shuttling wagon trains were unable to keep up with a rapid advance. In both the Franco-German War and the German invasion of France in 1914, German forces outran their trains and had to live off the French countryside, one of the richest agricultural regions in Europe. In the latter campaign, however, the Germans' tiny motor transport corps played a vital role in supplying ammunition for the opening battles. In subsequent operations on the Western Front, the immobility of the opposing forces provided an ideal environment for the staged resupply system, reversing the ancient rule that a "sitting" army must starve. On the other hand, many offensives on that front bogged down, after gaining only a few miles, through failure to move up quickly the quantities of fuel, ammunition, and supplies needed to maintain momentum.

The staged resupply system, in practice, did not precisely resemble either a pipeline or a series of conveyor belts maintaining a continuous flow from ultimate source to consumer. Reserves were stocked as far forward as was safe and practicable, permitting a regular supply of food and fuel and an immediate provision of ammunition, equipment, and services as needed. Before a major operation, large reserves had to be accumulated close behind the front; the two-year Allied build-up in the British Isles before the Normandy invasion of 1944, for example, involved the shipment of 16 million tons of cargo across the Atlantic. After the invasion, behind the armies on the Continent spread the rear-area administrative zone, a vast complex of depots, traffic regulating points, railway marshaling yards, troop cantonments, rest areas, repair shops, artillery and tank parks, oil and gasoline storage areas, air bases, and headquarters—through which ran the lines of supply stretching back to ultimate sources.

In the Pacific, the administrative zone covered vast reaches of ocean and clusters of islands. Communication and movement in this theatre depended largely on shipping, supplemented by aircraft, and one of the major logistic problems was moving forward bases and reserves as the fighting forces advanced. Supply ships often sailed all the way from the U.S. West Coast, bypassing intermediate bases, to forward areas where they were held as floating warehouses until their cargoes were exhausted.

In a real sense, the basic logistic tools of land operations in World War II were the railroad, the motor truck, and, carried over from the premechanized era, the horse-drawn wagon. Motor transport, when available, served to move forward the mountains of material brought to railheads by the railroads—a feat that, as the late 19th-century wars and World War I had shown, could not be done by horse-drawn vehicles rapidly enough to sustain fast-moving forces. When supplied by motor transport, mechanized armies, particularly in the European theatre, achieved a mobility and striking power never before seen. Paradoxically, Germany, which dominated operations in this theatre until late in the war, suffered from a severe shortage of motor transport and rolling stock, only partially made good by levies on conquered nations. The Wehrmacht that invaded the Soviet Union in 1941 consisted mainly of slow-moving infantry divisions supplied by horse-drawn wagons and spearheaded by a few armoured and mechanized units racing ahead. In order to maximize the capacity of its meagre motor transport, the organic transport of the armoured spearheads actually backtracked over the route of advance to pick up containerized fuel from prepositioned dumps—a novel modification of the staged resupply system. Motor transport was also supplemented by use of captured Soviet railroads (which had to be converted from wide to narrow gauge to accommodate German rolling stock) extending into the combat zone and paralleling vehicle roads.

Motorized and horse-drawn transport

The logistics of the North African desert campaigns in World War II virtually eliminated local supply and intermediate bases and depots, in effect replacing staged resupply by a simple single-shuttle base-to-troops operation. In 1941–42 the German Afrika Korps in Libya was supplied across the Mediterranean through the small port of Tripoli and eastward over a single coastal road that had no bases or magazines and was exposed to enemy air attack—a distance of up to 1,300 miles, depending on the location of the front (200 miles was considered the normal limit for effective supply). This operation was occasionally supplemented by small coastal shipments into the ports of Banghāzī and Tobruk. The fuel cost of this overland operation was between one-third and one-half of all the fuel imported.

One of the striking lessons of World War II, often obscured by the tactical achievements of air power and mechanized armour, was the great power that modern logistics gave to the defense. In 1943 and 1944 the ratio of superiority enjoyed by Germany's enemies in output of combat munitions was about 2.5:1; the whole apparatus of Germany's war economy was subjected to relentless attack from the air and had to make good enormous losses of matériel in a succession of military defeats. Yet Germany was able, for about two years, to hold its own, primarily because its waning logistic strength could be concentrated on sustaining the firepower of forces that were stationary or retiring slowly toward their bases, instead of on the expensive effort required to support a rapid forward movement.

Logistic specialization. For many centuries the soldier was a fighting man and nothing else; he depended on civilians to provide the services that enabled him to live, move, and fight. Even the more technical combat and combat-related skills, such as fortification, siegecraft, and service of artillery, were traditionally civilian. After the mid-19th century, with the rather sudden growth in the technical complexity of warfare, the military profession faced the problem of assimilating a growing number and variety of noncombatant skills. Many of the uniformed logistic services date from this period; examples are the British army's Transport Corps (later the Royal Army Service Corps), Hospital Corps, and Ordnance Corps. In the American Civil War the Union army formed a railway construction corps, largely civilian but under military control. A little later, Prussia created a railway section in the Great General Staff and a combined military–civilian organization for controlling and operating the railroads in time of war.

Ratios of combat and noncombat personnel

Not until the 20th century, however, did organized military units performing specialized logistic services begin to appear in large numbers in the field. By the end of World War II, what was called "service support" comprised about 45 percent of the total strength of the U.S. Army. Only three out of every 10 soldiers had combat functions, and even within a combat division one man out of four was a noncombatant. Even so, the specialized services that the military profession succeeded in assimilating were only a small fraction of those on which the combat soldier depended. Throughout the vast administrative zones behind combat areas and in the national

base, armies of civilian workers and specialists manned depots, arsenals, factories, communication centres, ports, and the other apparatuses of a modern society at war. Military establishments employed growing numbers of civilian administrators, scientists, technicians, management and public relations experts, and other specialists. Within the profession itself, the actual incorporation of specialized skills was limited, in the main, to those directly related (or exposed) to combat, such as the operating and servicing of military equipment, though even there the profession had no monopoly. Soldiers also served as administrators and supervisors over civilian specialists with whose skills they had only a nodding acquaintance. On the whole, the fighting man at mid-20th century belonged to a shrinking minority in a profession made up largely of administrators and noncombatant specialists.

Logistics in the nuclear age. The dropping of the first atomic bombs in August 1945 seemed to inaugurate a new era in warfare, demanding radical changes in logistic systems and techniques. The bombs did, in truth, give birth to a new line of weaponry of unprecedented destructive power. Within a decade they were followed by the thermonuclear weapon, an even greater leap in destructive force. Development of intercontinental ballistic missiles and nuclear-powered, missile-firing submarines a few years later extended the potential range of destruction to targets anywhere on the globe. The following decades saw dramatic developments in the offensive capabilities of nuclear weapons and also, for the first time, in defenses against them. But the world moved into the late 20th century without any of the new nuclear weaponry having been used in anger. Most warfare, moreover, was limited in scale and made little use of advanced technology. It produced only nine highly mobilized war economies: the two Koreas (1950–53), Israel (1956, 1967, 1973), North Vietnam (1965–75), Biafra (1967–70), Iran and Iraq (1980–88)—all except Israel preindustrial Third World countries.

Wars of limited scale

The first major conflict in this period, the war in Korea (1950–53), seemed in many ways an extension of the positional campaigns in World War II. It was fought largely with World War II weapons, in some cases improved versions, and with stocks of munitions left over from that conflict. United Nations forces had an excellent base in nearby Japan, whose factories made a major contribution by rebuilding U.S. World War II material. UN air superiority kept both Japan and Pusan, South Korea's major port of entry, free from communist air attack. UN forces thus were able to funnel through Pusan supply tonnages comparable to those handled by the largest ports in World War II and to concentrate depots and other installations in the Pusan area to a degree that would have been suicidal without air superiority. The communist supply system, although technically primitive, functioned well under UN air attack, moving troops and supplies by night, organizing local labour, and exploiting the Chinese soldier's famous ability to fight well under extreme privation.

By World War II standards, the Korean War was a limited conflict (except for the two Korean belligerents, on whose soil it was fought). It involved only a partial, or "creeping," economic mobilization in the United States and a modest mobilization of reserves. Yet this was no small war. Over three years about 37.2 million measurement tons of cargo were poured into the South Korean ports, more than three-fourths of the amount shipped to U.S. Army forces in all the Pacific theatres in World War II. Combined UN forces reached a peak strength of almost one million men; communist forces were considerably larger.

New technology. Advances in the technology of supply and movement after 1945 were not commensurate with those in weaponry. On land, internal-combustion vehicles and railroads, with increasing use of diesel fuel in both, remained the basic instruments of large-scale troop and freight movement despite their growing vulnerability to attack. In the most modern systems, substantial amounts of motor transport were capable of crossing shallow water obstacles. In areas not yet penetrated by rail or metaled roads—areas where much of the warfare of the period occurred—surface movement necessarily reverted to the ancient modes of human and animal porterage, sometimes usefully supplemented by the bicycle. Some exotic types of vehicles capable of negotiating rough and soft terrain off the roads were designed and tested—the "hovercraft," or air-cushion vehicle, for instance. But none of these innovations came into general use. The most promising developments in overland movement were helicopters and vertical-takeoff-and-landing aircraft, along with techniques of rapid airfield construction, which enabled streamlined airmobile forces and their logistic tails to overleap terrain obstacles and greatly reduced their dependence on roads, airfields, and forward bases. Helicopters also permitted the establishment and maintenance of isolated artillery fire bases in enemy territory.

In air movement there was a spectacular growth in the range and payload capacity of transport aircraft. The piston-engine transports of World War II vintage that carried out the Berlin airlift of 1948–49 had a capacity of about four tons (3,640 kilograms) and a maximum range of 1,500 miles (2,400 kilometres). The U.S. C-141 jet transport, which went into service in 1965, had a 45-ton (40,900-kilogram) capacity and a range of 3,000 miles (4,800 kilometres); it could take an average payload of 24 tons from the U.S. West Coast to South Vietnam in 43 hours and evacuate wounded back to the East Coast (10,000 miles) in less than a day. By 1970 these capabilities were dwarfed by the new "global logistics" C-5A, with payloads up to 130 tons and ranges up to 5,500 miles. It is estimated that 10 C-5As could have handled the entire Berlin airlift, which employed more than 140 of the then-available aircraft. C-5As played a vital role in the U.S. airlift to Israel during the Arab-Israeli War of October 1973. Very large cargo helicopters were also developed, notably in the Soviet Union, as were new techniques for packaging and air-dropping cargo.

Jet aircraft

In this period, movement by sea was the only branch of logistics that tapped the huge potential of nuclear propulsion. Its principal application, however, was in submarines, which did not develop a significant logistic function. (Development of nuclear-powered aircraft proved abortive.) The Soviet Union produced a nuclear-powered icebreaker in 1957, and the United States launched the first nuclear-powered merchant ship in 1959. But high initial and operating costs and (in the West) vested mercantile interests barred extensive construction of nuclear merchant ships. Except for supertankers built after the Suez crisis in 1956, and again during the energy crisis of the 1970s, seaborne cargo movement still depended on ships not radically different from those used in World War II. The chief technical improvement in sea lift, embodied in a few special-purpose vessels, was the "roll-on-roll-off" feature, first used in World War II landing craft, which permitted loading and discharge of vehicles without hoisting. Containerization, the stowage of irregularly shaped freight in sealed, reusable containers of uniform size and shape, became widespread in commercial ship operations and significantly affected ship design.

This period saw further development, from World War II models, of large vessels capable of discharging landing craft and vehicles offshore or over a beach as well as transporting troops, cargo, and helicopters in amphibious operations. For follow-up operations, improved attack cargo ships were built, such as the British landing-ship logistic, with accommodations for landing craft, helicopters, vehicles and tanks, landing ramps, and heavy-cargo-handling equipment. More revolutionary additions to the technology of amphibious logistics were the American landing vehicle hydrofoil and the BARC, both amphibians with pneumatic-tired wheels for overland movement and, in the latter case, capacity for 100 tons of cargo. Hydrofoil craft, which skimmed at high speeds above the water on submerged inclined planes, developed a varied family of types by 1970.

The revolution in electronic communication after World War II lies beyond the scope of this article, but its profound impact on logistic administration should be noted. In advanced logistic systems the combination of advanced electronic communication with the high-speed electronic computer almost wholly replaced the elaborate processes

Electronic communication and computers

of message transmission, record search, and record keeping formerly involved in supply administration, making the response of supply to demand automatic and virtually instantaneous.

Strategic mobility. Because the leading military powers did not directly fight each other during the decades after World War II, none of them had to deal with the classic logistic problem of deploying and supporting forces over sea lines of communication exposed to enemy attack. The Soviet Union was able in 1962 to establish a missile base in Cuba manned by some 25,000 troops without interference by the United States until its offensive purpose was detected. Similarly, the large deployments of U.S. forces to Korea, Southeast Asia, and elsewhere, as well as the 8,000-mile movement of a British expeditionary force to the Falkland Islands in 1982, encountered no opposition.

Supplying forces at global distances

Yet the problem of strategic mobility was of major concern after 1945 to the handful of nations with far-flung interests and the capacity to project military power far beyond their borders. In the tightly controlled power politics of the period, each of these countries needed the capability to bring military force quickly to bear to protect its interests in local emergencies at remote points—as Great Britain and France did at Suez in 1956, the United States in Lebanon in 1958 and in the Taiwan Straits in 1959, Great Britain in Kuwait in 1961 and in the Falkland Islands in 1982, and France in Chad on several occasions in the 1980s. The most effective instruments for such interventions were small, powerful, mobile task forces brought in by air or sea as well as forward-deployed aircraft-carrier and amphibious forces. The United States developed strong and versatile intervention capabilities, with major fleets deployed in the far Pacific and the Mediterranean; a worldwide network of bases and alliances; large ground and air forces in Europe, Korea, and Southeast Asia; and, in the 1960s, a mobile strategic reserve of several divisions with long-range sea-lift and airlift capabilities. The Soviet Union, Great Britain, and France had more limited capabilities, although the Soviet Union began in the late 1960s to deploy strong naval and air forces into the eastern Mediterranean and also maintained a naval presence in the Indian Ocean. After the U.S. withdrawal from Vietnam in 1973, the Soviet navy extended its power into the South China Sea.

The logistics of strategic mobility was complex and was decisively affected by the changing technology of movement, especially by air and sea. During the 1950s the proponents of naval and land-based air power debated the relative cost and effectiveness of naval-carrier forces and fixed air bases as a tool of emergency intervention. Studies seemed to show that the fixed bases were cheaper if all related costs were considered but that the advantage of mobility and flexibility lay with the naval carriers. In the 1970s the growing range and capacities of transport aircraft provided an increasingly effective tool for distant intervention and were a large factor in the reduction of the American and British overseas base systems. In practice, emergency situations called for using the means available and involved a great deal of improvisation, especially for second-rank powers.

Management. Both during and after World War II the United States operated the largest and most advanced logistic system in the world. Its wartime operations stressed speed, volume, and risk-taking more than efficiency and economy. The postwar years, with accelerated technological change, skyrocketing costs, and diminished public interest in defense, brought a revulsion against military prodigality, manifested by calls for reduced defense budgets and a growing demand for more efficient management of the military establishment. This demand culminated in a thorough overhaul of the whole system in the 1960s.

Introduction of modern managerial techniques and systems

One result was the reorganization of logistic activities in the three military services, generally along functional lines, with large logistic commands operating under functional staff supervision. In each service, however, each major weapon system was centrally managed by a separate project officer, and central inventory control was maintained for large commodity groups. In 1961 a new defense supply agency was established to manage on a wholesale basis the procurement, storage, and distribution of common military supplies and the administration of certain common services.

The most far-reaching managerial reforms of the period were instituted by the U.S. defense secretary, Robert S. McNamara (1961–68), in the resource allocation process. A unified defense planning–programming–budgeting system provided for five-year projections of force, manpower, and dollar requirements for all defense activities, classified into eight or nine major programs (such as strategic forces) that cut across the lines of traditional service responsibilities. The system was introduced in other federal departments after 1965, and elements of it were adopted by the British and other governments. In 1966 a program was inaugurated to integrate management accounting at the operating level with the programming–budgeting system. At the end of the 1960s a new administration restored some of the initiative in the planning–budgeting–programming cycle to the Joint Chiefs of Staff and the military services.

The reforms of the 1960s exploited the whole range of current managerial methodology. The basic techniques, such as systems and operations analysis, all stressed precise, scientific, usually quantitative formulations of problems and mathematical approaches to rational decision making. Systems analysis, the technique associated with defense planning and programming, was a method of economic and mathematical analysis useful in dealing with complex problems of choice under conditions of uncertainty. The technological foundation of this improved logistic management was the high-speed electronic computer, which was being used chiefly in inventory control; in automated operations at depots, bases, and stations; in transmitting and processing supply data; in personnel administration; and in command-and-control networks.

War in Vietnam. One of the most significant developments in logistics after 1945 was the pitting of advanced high-technology systems against well-organized low-technology systems operating on their own ground. The Korean War and the anticolonial wars in French Indochina and Algeria were the principal conflicts of this kind in the 1950s. The war in Vietnam following large-scale U.S. intervention in 1965 brought into conflict the most effective of both types of systems.

Scale of the U.S. logistic effort

Because South Vietnam lacked most of the facilities on which modern military forces depend, the massive U.S. deployment that began in the spring of 1965, reaching 180,000 men by the end of that year and more than 550,000 in 1969, was accompanied, rather than preceded, by a huge ($4 billion) construction program, carried out partly by army, navy, and air force engineer units and partly by a consortium of engineering contractors. Under this program were built seven deepwater and several smaller ports, eight jet air bases with 10,000-foot (3,050-metre) runways, 200 smaller airfields, and 200 heliports, besides millions of square feet of covered and refrigerated storage, hundreds of miles of roads, hundreds of bridges, oil pipelines and tanks, and all the other apparatuses of a modern logistic infrastructure. Deep-draft shipping brought in all but scarce items of airlifted supplies and came mainly from the U.S. directly.

The soldier in the field received lavish logistic support. By means of helicopter supply, troops in contact with the enemy were often provided with hot meals; most of the wounded were promptly evacuated to hospitals and serious cases were moved by air to base facilities in the Pacific or the United States. Medical evacuation, combined with advances in medicine, helped to raise the ratio of surviving wounded to dead to 6:1, in contrast to a World War II ratio of 2.6:1. Logistic support of army forces was organized under a single logistic command having a strength of 30,000 and employing 50,000 Vietnamese, U.S., and foreign civilians. Ultimately there were four or five support personnel for every infantryman who bore the brunt of contact fighting with the enemy.

The communist logistic system centred in the highly mobilized society of North Vietnam. In its integration, efficiency, and resilience under concentrated and prolonged bombing it rivaled the war economy of Germany

in World War II. Its resilience owed much, however, to its being a village-centred agricultural society, with modest material needs and a limited industrial base, which produced no steel, very little pig iron, and only one-fifth as much electric power as a single power plant in a small American town.

By late 1967 the communist war effort in South Vietnam depended heavily on the flow of troops, equipment, and supplies from North Vietnam, supplied mainly by the Soviet Union. The troops and most of the supplies moved over the Ho Chi Minh Trail, originally a network of footpaths and dirt roads (often paved after 1967) through communist-controlled areas in Laos and Cambodia. Supplies also came into South Vietnam by sea, directly across the northern border, and, especially after 1967, through the Cambodian port of Kompong Som and overland into the Mekong delta.

The Ho Chi Minh Trail

The Ho Chi Minh Trail was a long, slow-moving pipeline, requiring from three to six months in transit by truck, barge, ox cart, bicycle, and foot, but its capacity was ample for the modest demands placed upon it. In mid-1967, U.S. intelligence estimated the total nonfood requirements of all communist forces in South Vietnam, except in the northernmost provinces, to be as low as 15 tons (13,640 kilograms) per day (about 1.5 ounces, or 43 grams, per man); food was procured locally and in nearby Cambodia and Laos. In 1968, when the pace of the war quickened and communist forces were substantially augmented, estimated nonfood requirements rose to about 120 tons per day. (A single U.S. division required about five times this amount.)

American bombing had little effect on the flow of troops to the south, and the communist logistic system stood up remarkably well—and ultimately victoriously—under the weight of American air power. Its strength lay primarily in its austerity, but also in efficient organization, lavish use of manpower, availability of sanctuary areas in Laos and Cambodia, and a steady flow of imported supplies.

Afghanistan. The Soviet Union's Afghan war (1979–89), though on a scale smaller than Vietnam, embodied similar political, social, and economic dynamics and a similar contest between high-technology and low-technology logistic systems. Soviet forces, concentrated in the principal cities and towns, relied heavily on airlift and convoyed motor transport to move troops and supplies. Afghan guerrillas (called mujahideen), holding most of the countryside, used mainly animal transport and brought much of their supplies and weapons across the border from Pakistan. In an agriculturally poor country, significantly depopulated by Soviet bombing and forced flight into Pakistan, mass hunger and disease were widespread. For most of the war an approximate stalemate prevailed, in logistics as well as in tactical operations. But in 1986 the acquisition from the United States and Great Britain of substantial numbers of shoulder-fired surface-to-air missiles enabled the mujahideen to challenge Soviet control of the air—a significant factor in the Soviets' withdrawal early in 1989.

Trends and prospects. For logisticians the fundamental dilemma posed by the quantum leap in weapons technology after World War II was the absence of any comparable development in logistics. The electronic computer had, indeed, a dramatic impact on logistic planning and administration, as well as on military administration in general. The computer enabled planners to visualize problems concretely, often in quantitative terms; it accelerated the transmission of demand and the administrative response to it; and it enabled the military services for the first time to control their inventories. But the computer could not touch the ancient problem, compounded by the new weaponry, of actually providing and moving supplies to their users.

Logistics in a nuclear war

Conversely, nuclear weapons threatened to sweep away every vestige of the logistic system of the industrial era. None of the elaborate apparatuses of rear-area administration, lines of communication, or even sources of supply seemed likely to survive the nuclear firepower that could be brought to bear against it. The problem was studied and restudied, and a great deal of hopeful doctrine was developed for logistic operations in a nuclear war. It revolved about such concepts as dispersion, mobility, small targets, duplication, multiplicity, austerity, concealment, and automaticity, yet all of it was little more than a planner's dream, and a fading dream at that. At best it promised to reduce somewhat the inherent vulnerability of the surface-bound installations and transport on which military forces for the foreseeable future were likely to depend. Dispersion and duplication were enemies of economy and efficiency. The net effect could only be to increase the costs of logistic support and diminish the yield of delivered supplies and services.

In any case, nuclear war seemed the least likely of prospects. The most likely appeared to be a continuation of the confused patterns of limited conventional war and quasi-war that had filled the decades since the end of World War II. Under these conditions the central problems of logistics would be the historic ones of weight and bulk, which inhibited mobility and range of movement and were the primary causes of vulnerability to the new firepower. The technologies of these decades had accelerated the basic logistic trends of the industrial era: increasing complexity and cost in military hardware, increasing overall weight and volume of material (despite a reverse trend toward reduced numbers in some major items, such as aircraft), and, above all, an enormous increase in expenditures of ammunition and fuel. Logisticians in the postwar decades had to face the probability that in another large-scale conventional conflict between advanced powers the new vehicles would consume about half again as much fuel and the new weapons would expend more than four times as much ammunition as had been consumed and expended in World War II.

Some of the new tools of logistics were highly effective in specialized environments, notably the growing family of helicopters used in conjunction with conventional and short-takeoff-and-landing air transports, which permitted a mobility and a range of movement over difficult terrain far beyond the capabilities of surface transport. Whether an airmobile logistic system could survive the firepower likely to be encountered in a conflict with an adversary disputing command of the air was a question to which experience had not yet given an answer. In any case, the system purchased its mobility and range at a fuel cost several times higher than that involved in surface transport.

The unprecedented threat to supply lines

How well the "sophisticated" systems, with their growing burden of weight and bulk, would function under a threat to their previously immune supply lines was perhaps the most serious challenge facing modern logisticians. Nuclear propulsion offered a theoretical solution, but there seemed little hope for its early application to large sectors of military movement. A nuclear-powered sea transport service was a reasonable prospect, though not an early one, and it would not suffice for a major overseas war. More fundamentally, fuel consumption on the sea lanes was not the crux of the problem, and nuclear propulsion offered no solution to the vulnerability of surface vessels to air and submarine attack. The massive fuel consumers were aircraft and ground vehicles, and serious technical obstacles barred the application of nuclear energy to their power plants.

The reckoning, if there was to be one, might be long postponed. Given the existing distribution and equilibriums of power among the advanced nations, on the one hand, and the high cost and slow diffusion of sophisticated military technology to the less-developed two-thirds of the world, on the other, limited warfare seemed likely for a long time to come to remain at relatively low technical levels. Meanwhile, sophisticated logistic systems were becoming more entangled in their own complexity and absorbed in the endless pursuit of efficient management and in the struggle to control the waste and friction involved in delivering the tools of war to their users. (R.M.Le.)

Intelligence

Military intelligence is as old as warfare itself. Even in biblical times, Moses sent spies to live with the Canaanites in order to learn about their ways and about their

strengths and weaknesses. In the American Revolution George Washington relied heavily on information that was provided by an intelligence net based in New York City, and in World War II the results of a lack of good intelligence were realized in the destruction of the U.S. Pacific fleet at Pearl Harbor.

Today, nations have at their disposal information collection and processing systems that permit gathering and producing intelligence more rapidly and more accurately than ever before. Satellites, ultramodern aircraft, electronic systems, human sources, cameras, imaging and electronic devices, and a host of other systems permit the amassing of information on a scale that was unheard of in the past.

LEVELS OF INTELLIGENCE

Strategic and tactical intelligence

Intelligence is conducted at two levels, strategic and tactical. Strategic intelligence is information that is needed to formulate policy and military plans at the international and national policy levels. Tactical intelligence is intended primarily to respond to the needs of military field commanders so they can plan for and, if necessary, conduct combat operations. Essentially, tactical intelligence and strategic intelligence differ only in scope, point of view, and level of employment.

Whether tactical or strategic, military intelligence attempts to respond to or satisfy the needs of the operational leader, the person who has to act or react to a given set of circumstances. The process begins when the commander determines what information is needed to act responsibly. Several terms are used when discussing these requirements. On the national level they are usually called the essential elements of information and are defined as those items of intelligence information about a foreign power, armed force, target, or physical environment that are absolutely vital for timely and accurate decision making. On the tactical level intelligence needs are defined in a similar manner; often called information requirements, they are those items of information concerning the enemy and his environment that must be collected and processed in order to meet the intelligence needs of the military commander.

SOURCES OF INTELLIGENCE

It is critical for the intelligence analyst to know the source of information. Depending on the nature of a problem, certain sources are of great value and are therefore considered of high quality, while other sources, although contributing to the production of intelligence, are supportive rather than critical in nature.

Following are the major sources of intelligence.

Acoustics. This is information derived from analyzing acoustic waves that are radiated either intentionally or unintentionally. In naval intelligence, underwater acoustic waves from surface ships and submarines are detected by sonar arrays. These sensors are extremely accurate and are a major source of information on submarines in the world's oceans.

Imagery. This is information gleaned from analyzing all types of imagery, including photography as well as infrared and ultraviolet imagery. The examination of imagery, called imagery interpretation, is the process of locating, recognizing, identifying, and describing objects, activities, and terrain that appear on imagery.

Imagery collected by satellites and high-altitude aircraft is one of the most important sources of intelligence. It not only provides information for a huge number of intelligence categories (such as order of battle, military operations, scientific and technical developments, and economics), but it is indispensable for successfully monitoring compliance with arms-limitation treaties. The Intermediate Nuclear Forces Treaty of 1987 allows the United States periodically to request that the Soviet Union open certain intercontinental ballistic missile sites so that U.S. satellites (referred to as "national technical means") can verify that the sites do not house intermediate-range missiles banned by the treaty.

Monitoring compliance with treaties

Tactical infrared imaging devices can often identify camouflaged tanks and armour because the materials used to cover them—trees, branches, and leaves—often register different infrared signatures than does the surrounding foliage. Infrared satellites can register heat through clouds, producing imagery on enemy forces, equipment, and movements.

Signals. Gained from intercepting, processing, and analyzing foreign electrical communications and other signals, signals intelligence (often called SIGINT) comprises three elements: communications, electronics, and telemetry.

Communications intelligence is gleaned from foreign communications that are intercepted by other than the intended recipients. Such intelligence can be of the greatest value to a nation's fighting forces because it allows them to be privy to the strategies, weaknesses, and attitudes of the enemy. For example, before and during World War II, the U.S. Navy's breaking of the Japanese PURPLE code allowed the United States to know of Japanese moves in advance. It even provided warning of the attack on Pearl Harbor, although this intelligence was not sent to Hawaii quickly enough to prevent the debacle.

Electronics intelligence (also called ELINT) is technical and intelligence information obtained from foreign electromagnetic emissions that are not radiated by communications equipment or by nuclear detonations and radioactive sources. By analyzing the electronic emissions from a given weapon or electronic system, an intelligence analyst can very often determine the purpose of the device.

Telemetry intelligence is technical information that is derived from intercepting, processing, and analyzing foreign telemetry data. For example, by intercepting the telemetry signals emitted during foreign ballistic missile tests, an intelligence agency can calculate the range, accuracy, and number of warheads of the weapon.

Radiation. This source of intelligence does not include energy emanating from nuclear detonations or radioactive sources. Rather, it concerns unintentional emissions of energy from electronic systems (while ELINT is based on intentional radiations from the same systems). Inadequate shielding of electronic systems, or the following of incorrect procedures, may result in inadvertent energy emissions, which, when analyzed, may reveal a great deal about a system's purpose or capabilities.

Foreign matériel. In 1976 a Soviet air force lieutenant, wishing to defect to the West, flew a MiG-25 Foxbat to Japan, where Japanese and U.S. technicians pored over every detail of the supersonic fighter before reassembling it and handing it back to its owners. Such analysis of a foreign weapon system can prove invaluable in producing systems to defeat it, and intelligence derived from any foreign matériel is of great value in assessing enemy capabilities.

Analyzing foreign weapons

Human agents. Often called HUMINT, human intelligence is provided by people rather than by technical means and is very often provided by spies and covert agents. Spies are often a prime source of information about a nation's political leaders, strategies, and political decisions. The Soviet colonel Oleg Penkovsky, for example, was a very important source for British and U.S. intelligence until he was arrested and executed in 1963. The political, scientific, and technical information he provided included data on the capabilities of Soviet intermediate-range missiles during the Cuban missile crisis. Likewise, the Philby–Burgess–Maclean spy ring, which penetrated the highest circles of Britain's MI-6 intelligence agency, provided the Soviets with a tremendous amount of information on British and Allied military and counterintelligence operations during and after World War II. In the United States, the Walker family sold the Soviet Union classified reports on the tracking of Soviet submarines and surface ships. Operating from 1968 until it was broken up in 1985, this spy ring did irreparable damage to the submarine warfare capabilities of the U.S. Navy.

TYPES OF INTELLIGENCE

In most situations, intelligence production involves the assessment of conflicting pieces of incomplete information, the attempt to determine the correct items, and then the processing and assembly of these accurate items into a complete, understandable document that responds to the needs of the operational leader. More often than not the resulting product, which is usually called an intelligence

appraisal or intelligence assessment, contains some incorrect information.

In order to structure this production, analysts divide intelligence into types. While all types of intelligence are valuable, in any given situation some may be of greater worth than others, may be more accurate, and may provide a more complete view of the situation. By dividing intelligence into types, analysts and commanders arrive at a better understanding of the value and accuracy of a given piece of information.

Following are some important types of intelligence.

Armed forces. Information on a potential enemy's armed forces—that is, personnel, training, equipment, bases, capabilities, manpower levels, disposition, readiness, and other factors pertaining to strength and effectiveness—is crucial for a nation that is about to enter combat. If the weaknesses can be exploited, then the conflict may be won more quickly and with fewer casualties. Toward the end of World War II, owing to incomplete intelligence it was predicted that Japan would fight resolutely against a U.S. invasion and that the United States might suffer up to one million casualties. This was a major factor in the decision to drop the atomic bomb on Hiroshima and Nagasaki. In reality, though, Japanese resolve was grossly overestimated, and Japan could probably have been conquered with far fewer Allied casualties.

Assessing an enemy's strengths and weaknesses

Biographical. This is information collected on the views, traits, habits, skills, importance, relationships, health, and professional history of the leaders and important individuals of a nation. Biographical intelligence is important to those who must decide whether to support a foreign leader. For example, when Fidel Castro first came to power in Cuba in 1959, he claimed to be a nationalist and was even allowed to conduct a speaking tour in the United States. Subsequently, however, Castro revealed that he was a communist who intended to transform Cuba into a Soviet-style state. More accurate intelligence on Castro might have revealed his intentions more promptly, and U.S. foreign policy could have been revised accordingly.

In clandestine operations, one of the most difficult problems is assessing the validity of an individual who volunteers his services to an intelligence organization. Very often, information on the family life, education, travels, and professional and political affiliations of such a person provides great insight into motivation and can help in verifying authenticity.

Cartographic. Derived from maps and charts, cartographic intelligence is crucial for all military operations. During the Falkland Islands War, for example, British forces depended heavily on cartography. They also interviewed schoolteachers and scientists who had recently left the islands so that they had the most accurate information possible on road conditions, towns, and facilities. This prepared invading troops to meet the obstacles caused by rough terrain and poor roads, and as a result the invasion went remarkably well.

Economic. This is information concerning the production, distribution, and consumption of goods and services, as well as labour, finance, taxation, and other aspects of a nation's economy or of the international economic system. Economic intelligence allows a nation to estimate the magnitude of possible military threats and is also valuable in estimating the intentions of a potential enemy. In wartime, economic intelligence is a prime indicator of an enemy's ability to sustain a war. This is particularly important when analyzing small nations, such as Israel, where a conflict requires total mobilization and cannot be sustained for long without creating severe economic problems.

Energy. Energy intelligence specifically addresses the location and size of foreign energy resources; how these resources are used and allocated; foreign governments' energy policies, plans, and programs; new or improved foreign energy technologies; and the economic and security aspects of foreign energy supply, demand, production, distribution, and use.

Energy requirements can be an important factor in military planning. For example, as German forces were advancing on Moscow during World War II, Hitler, on being informed that the German military was short of fuel, sent several of the advancing units southward to capture the oil complexes at Baku on the Caspian Sea. This move so depleted the forces advancing on Moscow that they failed to capture the city, dealing the German war effort a fatal setback. Later, on the Western Front, advancing Allied forces were so short of fuel that U.S. general George Patton's 3rd Army was forced to stop and await replenishment. This allowed the retreating Germans to dig in and prolong the war.

Counterintelligence. Counterintelligence is intended to detect, counteract, and prevent espionage and other clandestine intelligence activities, sabotage, terrorist attacks, or assassinations conducted on behalf of foreign powers, organizations, or persons. It is especially vital that nations identify the capabilities and intentions of international terrorist organizations so that their operations can be thwarted; in the event that a terrorist attack is successful, identifying the culprit allows for reprisals, which are crucial to combating terrorism. In December 1988 an American commercial aircraft was destroyed over Scotland, and neither the United States nor Great Britain initially could identify the terrorist organization involved. As a result, the act was successful from the perspective of the terrorists, who had injured their enemy without suffering retaliation.

Preventing or reacting to attacks

Geographic. Gained from studying natural characteristics including terrain, climate, natural resources, transportation, boundaries, and population distribution, military geographic intelligence involves evaluating all such factors that in any way influence military operations.

Geographic intelligence was crucial to the success of Israel's rescue mission at the Entebbe airport in Uganda in 1976. Because they had reliable information on the exact location of the buildings at the airport, of the roads leading to Entebbe, and of military bases in the region, Israeli soldiers were able to land in three transport planes, kill many of the terrorists holding Israeli hostages, and depart with most of the hostages before the Ugandan military could react. A significant factor in the disastrous U.S. attempt to rescue its hostages in Iran in 1980 was a failure to anticipate and prepare for seasonal sandstorms, which disabled several helicopters and forced the rescuers to abort their mission.

Medical. This is intelligence gained from studying every aspect of foreign natural and man-made environments that could affect the health of military forces. This information can be used not only to predict the medical weaknesses of an enemy but also to provide one's own forces with adequate medical protection. For example, in the Spanish-American War the majority of U.S. casualties in the Caribbean resulted from disease rather than combat, because U.S. forces were not prepared to deal with the environment of that region.

Sociological. Information on a nation's social stratification, value systems, beliefs, and other social characteristics are of crucial value in assessing nations such as South Africa, the Soviet Union, or Israel, where national, racial, or social factions can have a great impact on a nation's military capability.

A lack of good sociological intelligence was a major cause of U.S. blunders in dealing with revolutionary Iran. When Mohammad Reza Shah Pahlavi was overthrown in 1979, the United States had only the most superficial understanding of Islām and Iranian society, and the situation improved only slightly in subsequent years. As a result, the United States often remained ignorant about Iranian officials, calling them "radical" or "moderate" even when such terms did more to cloud a situation than to make it clear.

Transportation and telecommunication. This type of intelligence can be crucial to correctly assessing a nation's ability to wage war, as it concerns a nation's highways, railroads, inland waterways, and civil airways as well as its telephone, telegraph, and civil broadcast capabilities. When China sent troops across the border into Vietnam in 1979, many observers assumed that China would win the conflict. This estimate was based on the huge size of the Chinese army and on its excellent performance against United Nations forces in the Korean War. After China

The enemy's ability to wage war

failed to score a decisive victory, the same commentators examined China's transportation and telecommunication networks and found that, while they were very highly developed in the Northeast, they were quite primitive in the South. It was concluded that the advanced northeastern systems and the primitive southern systems were prime factors in China's success in Korea and in its lackluster performance in Vietnam. (B.W.Wa.)

Guerrilla warfare

Guerrilla warfare is characterized by irregular forces employing unorthodox military tactics to fight small-scale, limited actions against orthodox civil and military forces. Traditionally, it is a method of protest employed to rectify real or imagined wrongs levied on a people either by a foreign invader or by an incumbent government, but it has also been used in an offensive role, in both ancient and modern times.

Spanish *guerrilleros*

Although this type of warfare is as old as history, the word guerrilla (the diminutive of Spanish *guerra,* "war") stems from the Duke of Wellington's campaigns during the Peninsular War (1808–14), in which Spanish and Portuguese irregulars, or *guerrilleros* (also referred to at the time as partisans, insurgents, and bandits), helped to drive the French from Iberia. In World War II the word partisan became synonymous with guerrilla; in later years the word insurgent came into vogue, followed by the (often contradictory) term freedom fighter.

HISTORY

In 512 BC the Persian warrior-king Darius I, who ruled the largest empire and commanded the best army in the world, bowed to the hit-and-run tactics of the nomadic Scythians and left them to their lands beyond the Danube. Alexander the Great (356–323 BC) also fought serious guerrilla opposition, which he overcame by modifying his tactics and by winning important tribes to his side. In 218 BC Hannibal faced considerable guerrilla opposition in crossing the Alps into Italy; he was later brought to bay by the delaying military tactics of Quintus Fabius Maximus, from whom the term Fabian tactics is derived and who earned the surname Cunctator (meaning "Delayer"). The Romans themselves fought against guerrillas in their conquest of Spain for more than 200 years before the foundation of the Empire.

Guerrilla and quasi-guerrilla operations were employed in an aggressive role in ensuing centuries by such predatory barbarians as the Goths and Huns, who forced the western Roman Empire onto the defensive; by the Magyars, who conquered Hungary; by the hordes of northern barbarians who attacked the eastern Roman Empire for more than 500 years; by the Vikings who overran Ireland, England, and France; by the Mongols, who conquered China and terrified central Europe. In the 12th century the Crusader invasion of Syria was at times stymied by the guerrilla tactics of the Seljuq Turks, an experience shared by the Normans in their conquest of Ireland (1169–75). A century later, Kublai Khan's army of Mongols was driven from Vietnam by Tran Hung Dao, who had trained his army to fight guerrilla warfare. King Edward I of England struggled through long, hard, and expensive campaigns to subdue Welsh guerrillas; that he failed to conquer Scotland was due largely to Robert Bruce's brilliant guerrilla operations. Bertrand du Guesclin, a Breton guerrilla leader in the Hundred Years' War (1337–1453), all but pushed the English from France by using Fabian tactics of harassment, surprise, ambush, sudden assault, and slow siege.

Guerrillas allied with conventional armies

Guerrilla warfare in time became a useful adjunct to larger political and military strategies—a role in which it complemented orthodox military operations both inside enemy territory and in areas seized and occupied by an enemy. Early examples of this role occurred in the first two Silesian Wars (1740–45) and in the Seven Years' War (1756–63), when Hungarian, Croatian, and Serbian irregulars (called *Grenzerer,* "border people"), fighting in conjunction with the Austrian army, several times forced Frederick the Great of Prussia to retreat from Bohemia and Moravia after suffering heavy losses. Toward the end of the American Revolution (1775–83), Francis ("Swamp Fox") Marion's ragtag band of South Carolina irregulars relied heavily on terrorist tactics to play a major role in driving the British general Charles Cornwallis from the Carolinas and to defeat at Yorktown, Va. Wellington's operations in Spain were frequently supported by effectively commanded regional bands of guerrillas—perhaps 30,000 in all—who made life miserable for the French invaders by blocking roads, intercepting couriers, and at times even waging conventional war. In 1812 Napoleon's columns suffered thousands of casualties in the long retreat from Moscow inflicted by bands of Russian peasants working with mounted Cossacks.

Guerrilla wars flourished in the following two centuries as native irregulars in India, Algeria, Morocco, Burma, New Zealand, and the Balkans tried, usually in vain, to prevent colonization by the great powers. Indian tribes in North America viciously fought the opening of the West; Cuban guerrillas fought the Spanish, and Filipino guerrillas fought the Spanish and Americans. In the South African War (1899–1902), 90,000 Boer irregulars held off a large British army for two years before succumbing.

The political factor

As these vicious campaigns continued, political motivations became more and more important. The Taiping Rebellion in China (1850–64), a peasant uprising against the Ch'ing Dynasty, killed an estimated 20 million Chinese before it was suppressed. Mexican peasants, fighting under such leaders as Emiliano Zapata, used guerrilla warfare to achieve a specific political goal in the Mexican Revolution (1910–20). Arab tribesmen under Fayṣal I employed T.E. Lawrence's brilliant guerrilla strategy and tactics in their campaign to liberate their lands from the Turks in World War I. In 1916 the Easter Rising in Ireland led to a ferocious guerrilla war fought by the Irish Republican Army (IRA)—a war that ceased only with the uneasy peace and partition of Ireland in 1921. In 1927 a 33-year-old communist leader in China, Mao Zedong (Mao Tse-tung), raised the flag of a rural rebellion that continued for 22 years. This experience resulted in a codified theory of protracted revolutionary war, Mao's *On Guerrilla Warfare* (1937), which was later called "the most radical, violent and extensive theory of war ever put into effect."

The political factor became more pronounced in the numerous guerrilla campaigns of World War II. In most of the countries invaded by Germany, Italy, and Japan, communists either formed their own guerrilla bands or joined other bands, such as the French and Belgian maquis. While consolidating their hold on the country, some of these groups spent as much time eliminating indigenous opposition as they did in fighting the enemy, but most of them contributed sufficiently to the Allied war effort to be sent shipments of arms, equipment, and gold, which helped to strengthen them for their postwar challenges to existing governments.

In Yugoslavia and Albania the communist takeover of government was simple and immediate; in China it was complicated and delayed; in Vietnam it succeeded after nearly three decades; in Greece, Malaya, and the Philippines it was foiled—but only after prolonged and costly fighting. Non-communist insurgents simultaneously used guerrilla warfare, with heavy emphasis on terrorist tactics, to help end British rule in Palestine in 1948 and Dutch rule in Indonesia in 1949.

Mao Zedong's victory in China in 1949 established him as the prophet of "revolutionary warfare" who had transferred Marxism-Leninism from the industrial areas to the countryside. A spate of new insurgencies, both communist and non-communist, followed to end French rule in Algeria and Indochina and British rule in Kenya, Cyprus, and Rhodesia. Fidel Castro's overthrow of the tottering and corrupt Batista regime in Cuba in 1959 provoked other rural insurgencies throughout Latin America, Southeast Asia, the Middle and Far East, and Africa. Ho Chi Minh's guerrillas ejected the French from Indochina, then held off U.S. forces in South Vietnam until the orthodox armies of North Vietnam claimed final victory in 1975. That same year, insurgencies in Angola and Mozambique freed those lands from Portuguese rule. South Africa similarly was forced to relinquish control of South West Africa/Namibia

in 1989. Rudely clad Muslim guerrillas, the famed mujahideen, fought a decade-long campaign that finally forced a powerful Soviet army to leave Afghanistan in 1989.

From the country to the cities

In the early 1970s the general failure of rural insurgencies in Central and South America caused some frustrated revolutionaries to shift from rural to urban guerrilla warfare with emphasis on terrorist tactics. Fired by the quasi-anarchistic teachings of Herbert Marcuse, Régis Debray, and others, and armed with a do-it-yourself manual of murder (Carlos Marighela, *For the Liberation of Brazil* [1970]), New Left revolutionaries embraced assassination, robbery, indiscriminate bombing, and kidnapping to attain their ends—crimes that became the order of the day as did, on an international scale, airplane hijackings, kidnappings, and mass murder.

Such was the media-heightened impact of urban guerrilla warfare, and such its potential danger to civilized society, that some observers believed "urban terrorism" should be classified as a new genre of warfare. But terrorist tactics, urban or rural, had always been integral to guerrilla and counterguerrilla warfare, indeed to all warfare. ("Kill one, frighten 10,000," wrote the Chinese general Sun-tzu in 350 BC.) In addition, urban guerrilla warfare by itself accomplished virtually nothing except to bring on a host of repressive measures to combat it. A more serious menace to society came from fringe advocates and practitioners of urban and international terrorism, groups of social dropouts far removed from guerrilla insurgencies—for example, the Black Panthers, the Weathermen, and the Symbionese Liberation Army in the United States, the Japanese Red Army, the Red Army Faction in West Germany, the Angry Brigade in Great Britain, and such Middle Eastern splinter groups as the Popular Front for the Liberation of Palestine–General Command. Such people came to be regarded by society as international criminals and were treated as such by law enforcement agencies around the world.

Meanwhile, guerrilla warfare continued to flourish. Guerrilla insurgencies got under way in Peru, Colombia, El Salvador, Nicaragua, the Philippines, Thailand, Sri Lanka, India, Lebanon, Syria, Morocco, Angola, Mozambique, and Northern Ireland. The most successful of these campaigns took place in Nicaragua, where the urban and rural guerrilla tactics of the Sandinist National Liberation Front led, in 1979, to the overthrow of the corrupt Somoza regime. Not without reason did one expert write that "guerrilla warfare and terrorism, rural or urban, internal or international, has undoubtedly now become the primary form of conflict for our time."

PRINCIPLES

Purpose and motivation. Fundamental to militant revolution is a cause, which unfortunately has never been difficult to find in a less than perfect world. The guerrillas' cause may assume several guises: to the world it may be presented as liberating a country from a colonial yoke; to the peasant being converted to communism it may be freedom from serfdom, from oppressive taxation, or from payment of oppressive rents to absentee landlords; to a middle-class citizen it may be establishment or restoration of representative government as opposed to a military- or Marxist-sponsored dictatorship.

Whether real or artificial, whether inspired by communism or by virulent nationalism or by a genuine desire for a better life, this cause is fundamental in motivating people to armed action. Mao Zedong leaves no doubt of its importance:

Mao Zedong and the role of politics

> Without a political goal, guerrilla warfare must fail, as it must if its political objectives do not coincide with the aspirations of the people and their sympathy, cooperation, and assistance cannot be gained.

Popular support. Revolutionary writings constantly stress the guerrillas' affiliation with the people. Guerrillas spring from the people, who in turn support their spawn, not only by furnishing their sons and daughters to the cause but also by furnishing money, food, shelter, refuge, transport, medical aid, and intelligence—support that must simultaneously be denied to the enemy. Although T.E. Lawrence called for no more than "a friendly population, not actively friendly, but sympathetic to the point of not betraying rebel movements to the enemy," he also wrote that his guerrillas "had won a province when the civilians in it had been taught to die for the ideal of freedom." Georgios Grivas, the non-communist Greek professional soldier who led the Cypriot rebellion in the 1950s, wrote that a guerrilla war stands no chance of success unless it has "the complete and unreserved support of the majority of the country's inhabitants." Mao repeatedly stressed the importance of proper troop behaviour: the Chinese guerrilla was required to pay a peasant for food, to respect his property, and not to offend propriety by undressing in front of a peasant woman.

Essential to maintaining popular support is intelligent propaganda to advertise success and stifle news of failure. The Yugoslav partisan leader Josip Broz Tito spread the word by newspaper, the Algerians by newspaper and radio, thereby enforcing Lawrence's dictum that the press is the greatest weapon in the army of a modern commander. Perhaps unfortunately, the printed word has since been supplemented by the television camera, defined by one expert as "a weapon lying in the street, which either side can pick up and use—and is more powerful than any other."

Leaders and recruits. Such are the vicissitudes of guerrilla warfare that outstanding leadership is necessary at all levels if a guerrilla force is to survive and prosper. A leader must not only be endowed with intelligence and courage but must be buttressed by an almost fanatical belief in himself and his cause. T.E. Lawrence, V.I. Lenin, Leon Trotsky, Mao Zedong, Josip Tito, the Filipino Luis Taruc, Menachem Begin, the Kenyan Jomo Kenyatta, the Malayan Ch'en P'ing, Fidel Castro, the Algerian Ahmed Ben Bella, Ho Chi Minh, Vo Nguyen Giap—these and dozens of their lieutenants were all unusual persons, generally with civilian backgrounds. But each attracted followers to a cause, organized them, and instilled a disciplined zeal matched only by the most elite military organizations.

The guerrilla recruit must be resourceful and enduring, committed totally to the cause if he is to withstand the hardships and dangers of guerrilla fighting. A prolonged and difficult campaign may force guerrilla leaders to abandon selectivity and resort to terrorist tactics in order to gain recruits—as was the case in Vietnam, where rigorous political indoctrination only partially compensated for lack of voluntary zeal.

Organization and unity of command. Protracted revolutionary warfare demands a complicated organization on both the political and military levels. Mao early developed a clandestine political-military hierarchy that began with the cadre or cellular party structure at the hamlet–village level and proceeded to the top via district, province, and regional command structures. This was roughly the concept followed by guerrilla forces in Malaya and Indochina. Tito was careful to build a parallel political organization in "liberated" areas as a foundation for his future government. Other guerrilla leaders formed civil organizations to provide money, supplies, intelligence, and propaganda. The Viet Cong and the Algerian rebels even established provisional governments in order to win external recognition, including representation at the United Nations.

Parallel political and military structures

Divided commands have plagued guerrilla leaders through the centuries and are probably more responsible for failed insurgencies than any other factor. The Algerian rebellion suffered severely until the National Liberation Front absorbed or neutralized rival guerrilla groups, but it never did settle feuds between the Arabs and the Berbers or between its own internal and external commands. Internal guerrilla rivalries dangerously weakened the post-colonial governments in Mozambique and Angola and seriously undermined the anti-Sandinista guerrilla effort in Nicaragua in the 1980s.

The tactical organization of guerrilla units varies according to size and operational demands. Mao called for a guerrilla squad of nine to 11; his basic unit was the company, about 120 strong. Grivas initially employed sabotage–terrorist teams of four or five members. The Greek insurgency opened with about 4,000 guerrillas divided into units of 150 fighters each that, as strength increased, grew to "battalions" 250 strong. Tito began his campaign

against the German invaders with about 15,000 guerrillas organized into small cadres; he ended the war with some 250,000 troops organized into brigades. Vietnamese guerrillas initially were organized in small squads that expanded to battalion and even regimental strengths.

Urban guerrilla units are small as a rule and are more tightly organized, generally in a cellular structure that from a security standpoint has proved valid over the decades but has suffered—as in the abortive Tupamaro insurgency in Uruguay (1963–73)—from an inability to take the war to the masses.

Arms. The guerrilla by necessity must fight with a wide variety of weapons, some homemade, some captured, and some supplied from outside sources. In the earlier stages of an insurgency, the weapons are often primitive. The Mau Mau in Kenya initially relied on knives and clubs (soon replaced by stolen British arms). French and American soldiers in Vietnam frequently encountered homemade rifles, hand grenades, bombs, booby traps, mines, and trails studded with punji stakes soaked in urine (to ensure infection). Nearly every guerrilla campaign has relied on improvisation, both from necessity and to avoid a cumbersome logistic "tail." Molotov cocktails and *plastique* (plastic explosive) bombs are cheap, yet under certain conditions nothing can be more effective. Stolen and captured arms are also a favourite source of supply, not least because army and police depots also stock ammunition to fit the weapons.

As a guerrilla war escalates and involves the interests of other nations, the guerrilla is often supplied with more sophisticated weapons, including modern rifles, machine guns, mortars, rockets, and even antitank and antiaircraft missiles, as was the case in Vietnam, Afghanistan, Nicaragua, Mozambique, and Angola.

Fighting on familiar ground

Terrain and sanctuary. It was axiomatic to Mao and his followers that revolution begins in familiar terrain. Once sufficient base and operational areas are established, guerrilla operations can be extended to include cities and vulnerable lines of communication. This rural strategy may be influenced by such factors as political goal, geography, and insurgent and government strengths.

If a guerrilla force is to survive, let alone prosper, it must control safe areas to which it can retire for recuperation and repair of arms, clothing, and equipment and where recruits can be indoctrinated, trained, and equipped. Such areas are traditionally located in remote, rugged terrain, usually mountains, forests, and jungles.

Sympathetic neighbouring countries may also provide sanctuary. Ho Chi Minh's guerrillas, in the later stages of the war against France, relied on China for refuge, training, and supply of arms and equipment; later in the war against the United States they used Laos and Cambodia for sanctuary. Still later Thai guerrillas found sanctuary and support in Cambodia, as did Nicaraguan guerrillas in Honduras. For years Spanish Basque guerrillas found sanctuary in France, as Northern Irish guerrillas did in the Irish Republic.

People offer a final form of sanctuary, one especially important to an urban guerrilla employing terrorist tactics. During the Cypriot war Grivas was surrounded by a British force for nearly two months without being captured. An Algerian rebel leader installed himself within 200 yards of the army commandant's office in Algiers. As a member of Irgun Zvai Leumi, a Zionist guerrilla group, Menachem Begin survived a massive British raid in Tel Aviv by hiding in a tiny cupboard.

Rapid transportation offers still another form of sanctuary, particularly for the international terrorist who frequently leaves the target country by jet aircraft before the time bomb explodes in the marketplace.

The uses of terror

Terror. Euphemistically termed "armed propaganda" by Marxist terrorists, terror is one of the most hideous characteristics of guerrilla warfare. It is used for several reasons: to focus world attention on the rebel cause with the hope of winning international support; to eliminate opposition leaders and, in the countryside, officials loyal to the government; to paralyze normal governmental activities; to intimidate the general populace in order to gain support and recruits (while denying them to the government); to keep one's own followers from defecting; and to raise funds by collecting ransoms for kidnapped victims.

Not all guerrilla leaders have favoured the use of terrorist tactics. In Palestine the Haganah broke with the Irgun and Stern gangs over the issue. IRA leaders in Ireland also disagreed on the use of terror, which resulted in a movement divided between "official" and "provisional" wings. After decades of advocating the use of terror, the leader of the Palestine Liberation Organization, Yāsir ʿArafāt, denounced international terrorism (more likely from a desire to gain political legitimacy than from genuine distaste).

It is difficult to assess the psychological impact of terrorist tactics on the general population, but it seems that even those persons originally sympathetic to a guerrilla cause may be alienated by the indiscriminate use of terrorism, such as planting bombs in shopping centres or blowing commercial aircraft out of the sky. They also may be disillusioned when orthodox forces reply in kind, so that the population is subject to terror from both sides and the original insurgency turns into virtual civil war. This happened in Argentina, Northern Ireland, Angola, Mozambique, and Nicaragua.

STRATEGY AND TACTICS

The broad strategy underlying successful guerrilla warfare is that of protracted harassment accomplished by extremely subtle, flexible tactics designed to wear down the enemy. The time gained is necessary either to develop sufficient military strength to defeat the enemy in orthodox battle (as did Mao Zedong in China and Ho Chi Minh in Vietnam) or to subject him to internal and external military and political pressures sufficient to cause him to seek peace favourable to the guerrillas (as did the Algerian guerrillas to France and the Angolan and Mozambican guerrillas to Portugal). This strategy embodies political, social, economic, and psychological factors to which the military element is often subordinated—without, however, lessening the ultimate importance of the military role.

That role greatly varies, as does the way it is carried out. T.E. Lawrence's Arabian campaign (1916–18) was strategically vital in protecting the British general Edmund Allenby's flank during his orthodox advance in Palestine, yet its success hinged on carrying out the Arabs' political aim, which was to expel the Turks from tribal lands. Lawrence's acceptance of this goal, combined with his linguistic ability, imagination, perception, and immense energy, helped him to establish and maintain unity of command. Popular support was ensured in part by tribal loyalties and hatred of the Turks, in part by effective propaganda and decent treatment of the people. There were too many Turks to risk doing battle, but in any case killing the enemy was secondary to killing his line of communication. In Lawrence's words, "the death of a Turkish bridge or rail" was more important than attacking a well-defended garrison. Lawrence kept discipline and organization (Arab, not Western, style) simple and effective. He drilled his men in the employment of light machine guns and in rudimentary demolitions. Camels provided transport. The terrain was desert and desert was sanctuary, and the guerrillas were "an influence, a thing invulnerable, intangible, without front or back, drifting about like a gas." Demanding "perfect intelligence, so that plans could be made in complete certainty," Lawrence "used the smallest force in the quickest time at the farthest place." Mobility and surprise were everything. Hit-and-run tactics on a broad front cut communication, eventually causing enemy garrisons to wither on the vine. By war's end the Arabs had gained control of some 100,000 square miles while holding 600,000 Turks in passive defense. They had killed or wounded 35,000 enemy at little loss to themselves. They had protected Allenby's vital flank in Palestine and had proved the truth of Lawrence's later dictum: "Guerrilla warfare is more scientific than a bayonet charge."

The "hit-and-run" tactics of T.E. Lawrence

Mao Zedong's political goal was the communist takeover of China. Guerrilla warfare alone, he realized, could not achieve this, but in a prolonged war it was an indispensable weapon, particularly in holding off the enemy (Chinese and Japanese) until orthodox armies could take to the field.

Mao's guerrilla campaign of over two decades stressed the flexible tactics based on surprise and deception that the ancient writer Sun-tzu had called for in *The Art of War*. Mao later wrote that "guerrilla strategy must be based primarily on alertness, mobility, and attack." He demanded tactics based on surprise and deception: "Select the tactic of seeming to come from the east and attacking from the west; avoid the solid, attack the hollow; attack, withdraw; deliver a lightning blow, seek a lightning decision." Mao instructed his subordinates to accept battle only under favourable conditions, otherwise avoid it and retreat: "We must observe the principle, 'To gain territory is no cause for joy, and to lose territory is no cause for sorrow.' " Careful planning was vital: "Those who fight without method do not understand the nature of guerrilla action."

Ho Chi Minh and his able military commander Vo Nguyen Giap were disciples of Mao's teachings, as was shown in their remarkably successful campaigns against the French and, later, against the U.S. and South Vietnamese armies. Ho and Giap did not, however, hesitate to extend guerrilla operations to the cities when occasion warranted. Vietnamese organization and leadership were generally effective, albeit expensive in lives. The use of terrain was often masterful, both tactically and for sanctuary. When popular support lagged, terrorist tactics were used, particularly the murder of pro-government village headmen, to coerce peasants into furnishing recruits, food, and information while denying these to the enemy. Operations were carefully planned and audaciously executed. As cruel as it was, the guerrilla portion of the Indochinese Wars must rank as one of the most successful in history.

The cost of abandoning guerrilla principles

The leader who does not respect the principles of guerrilla warfare soon finds himself in trouble, particularly against effective counterguerrilla forces. Greek communist guerrillas lost their war (1946–49) for a variety of reasons, not so much because Tito deprived them of sanctuary in and supply from Yugoslavia, but more because they forfeited popular support in northern Greece by their barbarous treatment of civilian hostages, by their rapacious behaviour in villages, and by kidnapping children and sending them to be raised in communist countries.

Filipino, Malayan, and Indonesian guerrillas of the 1940s and '50s suffered from poor organization and leadership as well as from lack of external support, and later "liberation" movements failed for similar reasons. Uruguayan and Guatemalan insurgents lost control over terrorist tactics and suffered heavily for it. Basque guerrillas became unpopular in Spain because of their brutal assassinations. Polisario fighters, inadequately supported by Algeria and Libya, faced continuing stalemate in their war against Morocco over Western Sahara. Angolan and Mozambican guerrillas split into several factions and became pawns of Cuba (by extension, the Soviet Union), South Africa, and the United States. The use of indiscriminate terrorist tactics by the provisional wing of the IRA brought general opprobrium on their movement, including a partial loss of heavy financial support from previously sympathetic Irish-Americans.

COUNTERGUERRILLA WARFARE

Perhaps the most important challenge confronting the military commander in fighting guerrillas is the need to modify orthodox battlefield thinking. This was as true in ancient, medieval, and colonial times as it is today. Alexander the Great's successful campaigns resulted not only from mobile and flexible tactics but also from a shrewd political expedient of winning the loyalty of various tribes (Alexander recruited one guerrilla leader into his army and then married his daughter). The few Roman commanders in Spain—Tiberius Sempronius Gracchus, Cato the Elder, the Scipios, and Pompey—who introduced more mobile and flexible tactics often succeeded in defeating large guerrilla forces, and their victories were then exploited by decent treatment of the vanquished in order to gain a relatively peaceful occupation.

Fighting guerrillas with guerrilla tactics

In their conquest of Ireland, the Normans borrowed the enemy guerrilla tactics of feigned retreat, flanking attack by cavalry, and surprise. (These tactics were countered by the Irish retreat to impenetrable bog country.) Early settlers in Virginia and New England tried to adopt the best features of Indian guerrilla tactics: small-unit operations, loose formations, informal dress, swift movement, fire discipline, terror, ambush, and surprise attack. As frontiers expanded, colonists reverted to European methods of formal warfare with disastrous results until a Swiss mercenary, Henry Bouquet, trained his new light-infantry regiment to fight Indian-style in the French and Indian War (1754–63). Bouquet's treatise on tactics, clothing, arms, training, logistics, and decentralized tactical formations is reminiscent of Caesar's work on Gaul. British generals fighting in the New World never quite understood Bouquet's teachings and suffered accordingly. A similar blindness cost Napoleon and his generals disastrous defeats in Spain and Russia.

The French conquest of Algeria (1830–44) might well have failed had it not been for tribal discord and the tactical innovations of Thomas-Robert Bugeaud, who understood the value of the ruse, the raid, and the ambush. Bugeaud dispensed with heavy columns in favour of small, fast-moving task forces, or "flying columns," which pursued and brought the Berbers to battles that were usually won by disciplined troops using superior arms. Although Bugeaud believed in constructive occupation—"the sword only prepared the way for the plough"—he nonetheless depended more on fear than on persuasion, relying on the razzia, or scorched-earth policy, to starve the natives into submission. Bugeaud's offensive tactics of clearing, holding, and expanding became the model for subsequent pacification campaigns around the globe, including the United States' winning of the West and its forays into colonialism in Cuba and the Philippines.

Such were the string of colonial successes that occasional serious reverses due to inept leadership and ill-trained troops were shrugged off. Orthodox commanders were generally quite content to put unquestioning faith in sheer military weight with little consideration given either to the poor organization and leadership of native forces or to the lack of modern arms and allies. Blockhouses and garrisons kept the peace in pacified areas. If natives rebelled, they were put down with force.

This simplistic concept was challenged by a French general, Louis-Hubert-Gonsalve Lyautey. He had been taught by Joseph-Simon Gallieni in Indochina in 1895 that military success, in Gallieni's words, meant "*nothing* unless combined with a simultaneous work of organization—roads, telegraphs, markets, crops—so that with the pacification there flowed forward, like a pool of oil, a great belt of civilization." Lyautey later employed this *tache d'huile,* or oil-spot, strategy in Algeria, where he used the army not as an instrument of repression but, in conjunction with civil services, as a positive social force—"the organization on the march." Lyautey's success went generally unheeded, as did the potency of the guerrilla weapon in World War I and subsequent decades. Native rebellions continued to be put down with force, no one paid much attention to Mao Zedong's guerrilla war, nor were orthodox commanders greatly impressed with the guerrilla weapon in World War II.

The cost of fighting guerrillas with conventional tactics

The greater the postwar shock, then, when these commanders and their subordinates were called upon to quell organized insurgencies by ideologically motivated, combat-trained guerrillas equipped with modern weapons and often politically allied with and, on occasion, supplied by the Soviet Union and its satellite countries.

Most governments and commanders simply floundered while calling for more soldiers and more weapons. The Greek army originally tried to suppress the "bandits" by static defense tactics that soon failed. Once the army had received massive reinforcements of U.S. arms and equipment, it launched large-scale offensives, or "search-and-clear" operations, which met with only limited success. Chinese Nationalist commanders moved vast armies hither and yon in futile efforts to capture Mao's guerrillas before finally holing up in towns and cities, where they eventually fell prey to Mao's own army divisions. During the Hukbalahap Rebellion (1946–54), U.S. Army advisers in the Philippines trained and equipped splendid Filipino

combat teams supported by armour, aircraft, artillery, and even dogs. Large-scale search-and-destroy operations—the "ring of steel" tactic similar to that unsuccessfully employed by German commanders against Tito's guerrillas—produced minimal results, as did free-fire areas (zones in which troops may fire at anything and everything), massive and sometimes brutal interrogations of villagers, and the employment of terrorist tactics, all of which further alienated the rural people whose support was necessary to defeat the guerrillas. Wiser commanders replaced conventional tactics with small-unit patrols and a variety of ruses that largely neutralized overt guerrilla action, then turned the army to the vital task of winning civil cooperation. With this the Huk insurgency died, but by the 1970s the failure to carry out promised reforms, mainly land distribution, had brought on a guerrilla insurgency by the New People's Army.

British commanders in Malaya also performed ineffectually in the early phases of the communist insurgency that began in 1948. Eventually, however, they realized that the support of the rural natives was vital to their goal of eliminating the entire guerrilla apparatus. Once they had achieved a reasonable civil-military chain of command, their first priority became the reestablishment of law and order, which meant revitalizing the rural police function. The military effort concentrated on breaking up and dispersing large guerrilla formations, then depriving them of the initiative by small-unit tactics, mainly frequent patrols and ambushes based on valid intelligence often gained from natives. The subsequent civil effort was designed to win "the hearts and minds" of the people, first by providing security in the form of village police and local militias working with government forces, second by providing social reforms (land reform, schools, hospitals) that identified the government with the people's best interests. Harsh measures were necessary: compulsory census, an identity-card system, suspension of habeas corpus (with carefully publicized safeguards), search of private property without a warrant, the death sentence for persons caught with unauthorized weapons, harsh sentences for collaborators, curfews, resettlement of entire villages, and other extraordinary measures. These were somewhat palliated by the British government's promise of eventual independence and by the general unpopularity of the guerrillas among the majority Malay population as well as the urban Chinese business community.

Winning "the hearts and minds" of the people

Despite the lessons of history, including those learned in Malaya and the Philippines, orthodox commanders have continued to employ a wide range of weapons and tactics that, judged by results, have been more appropriate to conventional warfare. These have included wholesale bombings and mass artillery interdictions of suspected sanctuary areas, division- and corps-strength "sweeping" operations in which only a few guerrillas are captured or killed while entire villages are destroyed, free-fire areas, the building of defended but isolated chains of military outposts, the construction of massive walls that can be outflanked, mass arrests, and brutal interrogations. The result: an expenditure in lives and money that in time have lost whatever support the government enjoyed at home. France backed down in Algeria, Portugal in Mozambique and Angola, the United States in Vietnam, the Soviet Union in Afghanistan. These campaigns failed because commanders chose quantity over quality to fight "little wars," in which victory is attainable only by patient application of intelligent, skillful, and extremely subtle strategy and tactics.

Precisely the same command qualities are necessary in fighting urban guerrilla warfare and international terrorism. Although a number of Western countries have established special antiterrorist military units, which at times have functioned effectively, the problem in the long run can be solved only by competent police work and by governments that refuse to allow the payment of ransoms (no matter the human cost) and refuse to be intimidated by terrorist threats into denying extradition of captured terrorists to other countries. Also necessary to put an end to terrorist depredations is a sharing of resources and intelligence among nations. (R.B.A.)

BIBLIOGRAPHY

Theory of war: CARL VON CLAUSEWITZ, *On War,* 3 vol. in 1 (1873; originally published in German, 3 vol., 1832–34), remains a classic and is available in many later editions both in the original and in English translation, of which the one edited by and with commentary by MICHAEL HOWARD and PETER PARET (1976) can be recommended. Other outstanding classical philosophical treatments are IMMANUEL KANT, *Perpetual Peace: A Philosophical Essay* (1903, reissued 1972; originally published in German, 1795), also available in later editions, some entitled *Eternal Peace;* JOHN STUART MILL, *On Liberty* (1859) and *Considerations on Representative Government* (1861), both available in the selection published as *Political Writings* (1981), in "The Great Books of the Western World" series; and JEAN-JACQUES ROUSSEAU, *A Lasting Peace Through the Federation of Europe, and The State of War,* trans. from French (1917).

Explanations of war as a product of imperialism are found in J.A. HOBSON, *Imperialism,* 3rd rev. ed. (1938, reprinted 1988); and in Marxist writings. Psychological theories of war are best expressed in SIGMUND FREUD, *Civilization, War, and Death,* new ed., ed. by JOHN RICKMAN (1953, reprinted 1968); and ethological theories in ROBERT ARDREY, *The Territorial Imperative* (1966, reissued 1978); and KONRAD LORENZ, *On Aggression* (1966, reprinted 1974; originally published in German, 1963). NORMAN ANGELL, *The Great Illusion: A Study of the Relation of Military Power in Nations to Their Economic and Social Advantage* (1910), is an early economic study, available in many later editions. REINHOLD NIEBUHR, *Moral Man and Immoral Society* (1932, reprinted 1960), analyzes war from the ethical point of view.

A comprehensive modern analysis is QUINCY WRIGHT, *A Study of War,* 2nd ed. (1964). Other major contemporary general analyses are RAYMOND ARON, *The Century of Total War* (1954, reprinted 1985); KENNETH N. WALTZ, *Man, the State, and War* (1959, reprinted 1965); and HANS J. MORGENTHAU and KENNETH W. THOMPSON, *Politics Among Nations,* 6th ed. (1985). On nuclear war, see HERMAN KAHN, *On Escalation: Metaphors and Scenarios* (1965, reprinted 1986). INIS L. CLAUDE, JR., *Swords into Plowshares,* 4th ed. (1971, reprinted 1984), discusses United Nations approaches to peace. Later studies include EVAN LUARD, *War in International Society: A Study in International Sociology* (1986), a broad and systematic view of war in history; J. DAVID SINGER *et al., Explaining War: Selected Papers from the Correlates of War Project* (1979); and MELVIN SMALL and J. DAVID SINGER, *Resort to Arms: International and Civil Wars, 1816–1980,* rev. ed. (1982). (J.Fr.)

Law of war: Among comprehensive works of reference quality are L. OPPENHEIM, *International Law, a Treatise,* 8th ed., ed. by H. LAUTERPACHT, vol. 2, *Disputes, War, and Neutrality* (1962); MYRES S. MCDOUGAL and FLORENTINO P. FELICIANO, *Law and Minimum World Public Order: The Legal Regulation and International Coercion* (1961); JULIUS STONE, *Aggression and World Order: A Critique of United Nations Theories of Aggression* (1958, reprinted 1976), and *Legal Controls of International Conflict: A Treatise on the Dynamics and Disputes and War-Law* (1959, reprinted 1974); INGRID DETTER DE LUPIS, *The Law of War* (1987); and YORAM DINSTEIN, *War, Aggression, and Self-Defence* (1988).

For the history of the laws of war, see GEOFFREY BEST, *Humanity in Warfare* (1980). For the development of the law of war as it exists today, see JEAN S. PICTET (ed.), *The Geneva Conventions of 12 August 1949: Commentary,* 4 vol. (1952–60); CLAUDE PILLOUD *et al.* (eds.), *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (1987; originally published in French, 1986); G.I.A.D. DRAPER, *The Red Cross Conventions* (1958); and WALDEMAR A. SOLF and J. ASHLEY ROACH (eds.), *Index of International Humanitarian Law* (1987).

Special treatments of developments brought about by particular military campaigns include ROBERT R. BOWIE, *Suez 1956* (1974); GEORGES ABI-SAAB, *The United Nations Operation in the Congo, 1960–1964* (1978); ABRAM CHAYES, *The Cuban Missile Crisis* (1974, reprinted 1987); RICHARD A. FALK (ed.), *The Vietnam War and International Law,* 4 vol. (1968–76); ALLAN GERSON, *Israel, the West Bank, and International Law* (1978); ALBERTO R. COLL and ANTHONY C. AREND (eds.), *The Falklands War: Lessons for Strategy, Diplomacy, and International Law* (1985); and SCOTT DAVIDSON, *Grenada: A Study in Politics and the Limits of International Law* (1987). For internal conflicts, see RICHARD A. FALK (ed.), *The International Law of Civil War* (1971); and HEATHER A. WILSON, *International Law and the Use of Force by National Liberation Movements* (1988).

Laws on particular weapons, means of conflict, and areas of war are explored in HOWARD S. LEVIE, *The Code of International Armed Conflict,* 2 vol. (1986); MORRIS GREENSPAN, *The Modern Law of Land Warfare* (1959); ROBERT TUCKER, *The Law of War and Neutrality at Sea* (1957); D.P. O'CONNELL, *The International Law of the Sea,* 2 vol. (1982–84); D.H.N. JOHNSON,

Rights in Air Space (1965); ANN VAN WYNEN THOMAS and A.J. THOMAS, JR., *Development of International Legal Limitations on the Use of Chemical and Biological Weapons,* 2 vol. (1968); GEORG SCHWARZENBERGER, *The Legality of Nuclear Weapons* (1958); R.C. HINGORANI, *Prisoners of War,* 2nd ed. (1982); and PETER ROWE, *Defence: The Legal Implications: Military Law and the Laws of War* (1987).

War crimes as a branch of military law are analyzed in ROBERT K. WOETZEL, *The Nuremberg Trials in International Law: With a Postlude on the Eichmann Case* (1962); ANN TUSA and JOHN TUSA, *The Nuremberg Trial* (1983); PHILIP R. PICCIGALLO, *The Japanese on Trial: Allied War Crimes Operations in the East, 1945–1951* (1979); L.C. GREEN, *Superior Orders in National and International Law* (1976); and PETER KARSTEN, *Law, Soldiers, and Combat* (1978).

Current research is found in useful articles in various journals, including *The American Journal of International Law* (quarterly); *The British Year Book of International Law; The International and Comparative Law Quarterly; Military Law Review* (quarterly); *The Military Law and Law of War Review* (semiannual); and *International Review of the Red Cross* (bimonthly). (P.J.R.)

Military law: Development of military law in the United States is traced in UNITED STATES. ARMY. JUDGE ADVOCATE GENERAL'S CORPS, *The Army Lawyer: A History of the Judge Advocate General's Corps, 1775–1975* (1975); and *Military Law Review Bicentennial Issue* (1975), an anthology of articles by 17 authors. Both are useful for the general reader as well as the specialist. A concise history of British military law is found in Peter Rowe's work cited above. FREDERICK B. WIENER, *Civilians Under Military Justice: The British Practice Since 1689, Especially in North America* (1967), offers a general history of the subject, with a focus on the British administration of 18th-century America. A special aspect of military law is treated in NICO KEIJZER, *Military Obedience* (1978). INTERNATIONAL SOCIETY FOR MILITARY LAW AND THE LAW OF WAR, *The Present Evolution of Military Justice,* 2 vol. (1981), studies systems of military law throughout the world on the basis of national reports from 25 countries. See also JAMES STUART-SMITH, "Military Law: Its History, Administration, and Practice," *Law Quarterly Review* 85:478–504 (October 1969). Journals cited in the previous section also contain many articles on the military law of many countries. (J.S.-Sm.)

Defense economics: GAVIN KENNEDY, *Defence Economics* (1983), is a general survey written for the nonspecialist and providing information on a broad spectrum of questions of military appropriations and expenditures. Historical studies include J.M. WINTER (ed.), *War and Economic Development* (1975), addressing the economic consequences of war on the preindustrial economies of Europe; GAUTAM SEN, *The Military Origins of Industrialization and International Trade Rivalry* (1983), exploring the influence of military technology on the industrial development of Europe; and EMILE BENOIT, *Defense and Economic Growth in Developing Countries* (1973), challenging the assumption that defense spending is incompatible with economic growth and attempting to show that defense spending has actually assisted economic development in some Third World countries. SAADET DEGER and ROBERT WEST (eds.), *Defence, Security, and Development* (1987), is a collection of articles by scholars from several countries who criticise Benoit's method and conclusions and provide a more complex assessment of the connection between economic growth and national security. A detailed study of governmental defense planning in poorer economies is presented in STEPHANIE G. NEUMAN (ed.), *Defense Planning in Less-Industrialized States: The Middle East and South Asia* (1984). For a view that defense spending inhibits economic performance, based on international comparative data and statistical analysis, see ROBERT W. DEGRASSE, JR., *Military Expansion, Economic Decline: The Impact of Military Spending on U.S. Economic Performance,* expanded ed. (1983). CHRISTIAN SCHMIDT (ed.), *The Economics of Military Expenditures: Military Expenditures, Economic Growth, and Fluctuations* (1987), collects papers by most of the world's leading specialists in defense economics, providing a rich agenda in the contribution of economic analysis to defense dilemmas.

A rare comparative study of defense industries in both capitalist and communist economies of such countries as the United States, the Soviet Union, France, Germany, Sweden, Czechoslovakia, Italy, China, Israel, the United Kingdom, and some representatives of the Third World is provided in NICOLE BALL and MILTON LEITENBERG (eds.), *The Structure of the Defense Industry: An International Survey* (1983). (G.Ke.)

Strategy: EDWARD MEAD EARLE *et al.* (eds.), *Makers of Modern Strategy: Military Thought from Machiavelli to Hitler* (1943, reprinted 1971), remains a most competent anthology on the development of the military mind and art. Its worthy successor is PETER PARET (ed.), *Makers of Modern Strategy: From Machiavelli to the Nuclear Age* (1986), which repeats three of the essays and provides more than 20 new ones.

For works that explore the complex correlation between military campaigns and such nonmilitary factors as economic, technological, and human resources, psychological conditions of people involved, and political decision making by statesmen and generals, the reader will have to turn to broad military histories, such as JAMES R.M. BUTLER (ed.), *Grand Strategy,* 6 vol. in 7 (1956–76), an analysis of the high-level conduct of war by the British High Command; GORDON WRIGHT, *The Ordeal of Total War, 1939–1945* (1968); HENRI MICHEL, *The Second World War,* 2 vol. (1975; originally published in French, 1968–69); and PETER CALVOCORESSI, GUY WINT, and JOHN PRITCHARD, *Total War: Causes and Courses of the Second World War,* 2nd rev. ed. (1989).

The role of the U.S. Army and the War Department in World War II is studied in MAURICE MATLOFF and EDWIN M. SNELL, *Strategic Planning for Coalition Warfare,* 2 vol. (1953–59, reprinted 1968); KENT ROBERTS GREENFIELD (ed.), *Command Decisions* (1960, reprinted 1984); and RONALD H. SPECTOR, *Eagle Against the Sun: The American War with Japan* (1985). Russian and Soviet strategists are discussed in JOHN ERICKSON, *Stalin's War with Germany* (1975), and *The Road to Berlin: Continuing the History of Stalin's War with Germany* (1983). (Ma.M./M.E.Ho.)

Strategy in the nuclear age is surveyed in JOHN BAYLIS *et al., Contemporary Strategy,* 2nd rev. ed., 2 vol. (1987); INSTITUTE FOR STRATEGIC STUDIES, *Problems of Modern Strategy* (1970); and LAWRENCE MARTIN (ed.), *Strategic Thought in the Nuclear Age* (1979). LAWRENCE FREEDMAN, *The Evolution of Nuclear Strategy,* 2nd ed. (1989), summarizes the history of thought on nuclear strategy. FRED KAPLAN, *The Wizards of Armageddon* (1983), is a look at the thinkers and strategists themselves.

The classics of postwar strategic thinking include BERNARD BRODIE, *Strategy in the Missile Era* (1959, reprinted 1965); HERMAN KAHN, *On Thermonuclear War,* 2nd ed. (1961, reprinted 1978), and *On Escalation: Metaphors and Scenarios* (1965, reprinted 1986); THOMAS C. SCHELLING, *The Strategy of Conflict* (1960, reprinted 1980), and *Arms and Influence* (1966, reprinted 1976). Among later contributions, EDWARD N. LUTTWAK, *Strategy: The Logic of War and Peace* (1987), is a complex scholarly treatment of strategy as a system of thought; and MICHAEL HOWARD, *The Causes of Wars and Other Essays* (1983), is a collection of previously published works by an influential author.

The theory of deterrence is considered in PATRICK M. MORGAN, *Deterrence: A Conceptual Analysis,* 2nd ed. (1983); and carefully analyzed in ALEXANDER L. GEORGE and RICHARD SMOKE, *Deterrence in American Foreign Policy: Theory and Practice* (1974). PAUL BRACKEN, *The Command and Control of Nuclear Forces* (1983), is another authoritative treatment; and JOHN J. MEARSHEIMER, *Conventional Deterrence* (1983), applies the deterrence theory to land warfare.

Special branches of strategic thought are examined in GEOFFREY TILL, *Maritime Strategy and the Nuclear Age,* 2nd ed. (1984); and M.J. ARMITAGE and R.A. MASON, *Air Power in the Nuclear Age, 1945–84: Theory and Practice,* 2nd ed. (1985). On Soviet strategic thinking, see JOHN BAYLIS and GERALD SEGAL (eds.), *Soviet Strategy* (1981). Influence of the Vietnam experience is studied in HARRY G. SUMMERS, JR., *On Strategy: A Critical Analysis of the Vietnam War* (1982). (L.D.Fr.)

Tactics of land warfare: Fundamentals of tactics, laid out in early wars, are discussed in such general histories as J.F.C. FULLER, *A Military History of the Western World,* 3 vol. (1954–56; also published as *The Decisive Battles of the Western World, and Their Influence upon History*); LYNN MONTROSS, *War Through the Ages,* 3rd rev. ed. (1960); and RICHARD A. PRESTON and SYDNEY F. WISE, *Men in Arms: A History of Warfare and Its Interrelationships with Western Society,* 4th ed. (1979). The role of weapons in tactical decisions is studied in TOM WINTRINGHAM, *The Story of Weapons and Tactics from Troy to Stalingrad* (1943, reprinted 1971); and JAMES R. NEWMAN, *The Tools of War* (1942). For tactical theory, see CLAUSEWITZ, *op. cit.;* CHARLES ARDANT DU PICQ, *Battle Studies* (1921, reprinted 1983; originally published in French, 1880, 8th French. ed., 1914); and J.F.C. FULLER, *The Reformation of War* (1923).

Tactics of ancient warfare are explored in THEODORE AYRAULT DODGE, *Alexander* (1890), *Hannibal* (1891), and *Caesar* (1892, reissued 1963). For medieval wars, see CHARLES OMAN, *A History of the Art of War in the Middle Ages,* 2nd rev. ed., 2 vol. (1924, reissued 1978), and *A History of the Art of War in the Sixteenth Century* (1937, reprinted 1979). Early modern wars are surveyed in C.V. WEDGWOOD, *The Thirty Years War* (1938, reprinted 1981); and WINSTON S. CHURCHILL, *Marlborough: His Life and Times,* 6 vol. (1933–38, reissued in 2 vol., 1974). Discussions of the wars of the Napoleonic era include THEODORE AYRAULT DODGE, *Napoleon,* 4 vol. (1904–07, reprinted 1970); B.H. LIDDELL HART, *The Ghost of Napoleon* (1933, reprinted 1980); and CHARLES OMAN, *A History of the Peninsular War,*

7 vol. (1902–30, reprinted 1980). American wars of the 19th century are looked at in DOUGLAS SOUTHALL FREEMAN, *R.E. Lee,* 4 vol. (1934–35); and J.F.C. FULLER, *The Generalship of Ulysses S. Grant,* 2nd ed. (1958, reprinted 1977); for Europe, see COLMAR VON DER GOLTZ, *The Nation in Arms,* (1903; originally published in German, 1883), available in many later editions, as is WINSTON S. CHURCHILL, *The World Crisis,* 4 vol. in 5 (1923–29); and B.H. LIDDELL HART, *A History of the World War, 1914–1918* (1934, reissued in 1970 as *The Real War, 1914–1918).* Surveys of the World War II campaigns include CHESTER WILMOT, *The Struggle for Europe* (1952, reissued 1986); J.F.C. FULLER, *The Second World War, 1939–1945* (1948, reprinted 1968); and JETER A. ISELY and PHILIP A. CROWL, *The U.S. Marines and Amphibious War* (1951, reprinted 1979). For the Korean War, see RUSSELL A. GUGELER, *Combat Actions in Korea,* rev. ed. (1970, reprinted 1987).

See also the official histories of the U.S., British, Canadian, and Australian armed forces. (L.Mo./Ed.)

Naval tactics: The most extensive work devoted exclusively to naval tactics is S.S. ROBISON and MARY L. ROBISON, *A History of Naval Tactics from 1530 to 1930: The Evolution of Tactical Maxims* (1942). GIUSEPPE FIORAVANZO, *A History of Naval Tactical Thought* (1979; originally published in Italian, 1973), is an authoritative later analysis. WAYNE P. HUGHES, JR., *Fleet Tactics: Theory and Practice* (1986), includes history but is more attentive to the art and science of tactics.

WILLIAM LEDYARD RODGERS, *Naval Warfare Under Oars, 4th to 16th Centuries* (1939, reprinted 1967), interprets galley warfare. The history of tactics in the age of fighting sail as presented by JULIAN S. CORBETT in his numerous works, from *Drake and the Tudor Navy: With a History of the Rise of England as a Maritime Power,* 2 vol. (1898, reissued 1988), to *England in the Mediterranean: A Study of the Rise and Influence of British Power Within the Straits, 1603–1713* (1904, reprinted 1987), remains unsurpassed. A.T. MAHAN, *The Influence of Sea Power upon History, 1660–1783* (1890), available in many later editions, is a classic naval history. Studies of tactics in the age of steam and steel include BRADLEY A. FISKE, *The Navy as a Fighting Machine,* 2nd ed. (1918, reprinted 1988); ROMEO BERNOTTI, *The Fundamentals of Naval Tactics* (1912; originally published in Italian, 1910); and W. BAINBRIDGE-HOFF, *Elementary Naval Tactics* (1894). RICHARD HOUGH, *The Great War at Sea, 1914–1918* (1983), explores the tactics of naval operations in World War I. The transition from battleship to aircraft carrier is discussed in BERNARD BRODIE, *A Layman's Guide to Naval Strategy* (1942). Studies of naval tactics in World War II are found in comprehensive histories, such as SAMUEL E. MORISON, *History of United States Naval Operations in World War II,* 15 vol. (1947–62); S.W. ROSKILL, *The War at Sea, 1939–1945,* 3 vol. in 4 (1954–61); and FRIEDRICH RUGE, *Der Seekrieg: The German Navy's Story, 1939–1945* (1957; originally published in German, 1954; also published as *Sea Warfare, 1939–1945: A German Viewpoint).* CLARK G. REYNOLDS, *The Fast Carriers: The Forging of an Air Navy* (1968, reprinted 1978), studies carrier operations. (W.P.H.)

Air tactics: CHARLES HARVARD GIBBS-SMITH, *Aviation: An Historical Survey from Its Origins to the End of World War II,* 2nd ed. (1985), and *Flight Through the Ages: A Complete, Illustrated Chronology from the Dreams of Early History to the Age of Space Exploration* (1974), provide basic introduction to the development of airplanes as a military force. ROBIN HIGHAM, *Air Power: A Concise History,* 3rd rev. ed. (1988); and BILL GUNSTON (ed.), *Aviation: The Complete Story of Man's Conquest of the Air* (1978), are general surveys. EDWARD H. SIMS, *Fighter Tactics and Strategy, 1914–1970,* 2nd ed. (1980); and ROBERT L. SHAW, *Fighter Combat: The Art and Science of Air-to-Air Warfare,* 2nd ed. (1988), focus on fighter planes. R.A. MASON (ed.), *War in the Third Dimension: Essays in Contemporary Air Power* (1986), emphasizes continuity, innovation, and convergence in development of military aeronautics. LON O. NORDEEN, *Air Warfare in the Missile Age* (1985), provides an excellent overview of air tactics from 1964 to the 1980s. (D.MacI.)

Logistics: MARTIN VAN CREVELD, *Supplying War: Logistics from Wallenstein to Patton* (1977), is an insightful groundbreaking history of logistics. Classic studies of the subject include GEORGE CYRUS THORPE, *George C. Thorpe's Pure Logistics: The Science of War Preparation,* new ed., with an introduction by STANLEY L. FALK (1986); G.C. SHAW, *Supply in Modern War* (1938), mainly on subsistence; S.L.A. MARSHALL, *The Soldier's Load and the Mobility of a Nation* (1950, reprinted 1980); and HENRY E. ECCLES, *Logistics in the National Defense* (1959, reprinted 1981), with emphasis on theory.

The 18th-century logistics systems are examined in LEE KENNETT, *The French Armies in the Seven Years' War: A Study in Military Organization and Administration* (1967, reprinted 1986); and ERNA RISCH, *Supplying Washington's Army* (1981). The U.S. Army experience is surveyed in JAMES A. HUSTON, *The Sinews of War: Army Logistics 1775–1953* (1966, reprinted 1988); EDWIN A. PRATT, *The Rise of Rail-Power in War and Conquest, 1833–1914* (1916), an old but still useful survey; ROBERT GREENHALGH ALBION and JENNIE BARNES POPE, *Sea Lanes in Wartime: The American Experience 1775–1945,* 2nd ed. (1968); and C.B.A. BEHRENS, *Merchant Shipping and the Demands of War,* rev. ed. (1978), on the overseas supply. ROLAND G. RUPPENTHAL, *Logistical Support of the Armies,* 2 vol. (1953–59, reprinted 1985–87); and RICHARD M. LEIGHTON and ROBERT W. COAKLEY, *Global Logistics and Strategy,* 2 vol. (1955–68), provide the U.S. Army's official history of logistics in World War II, in the European theatre and in the framework of coalition strategy, respectively. R. ELBERTON SMITH, *The Army and Economic Mobilization* (1959, reprinted 1985), analyzes the U.S. Army's World War II economic mobilization.

CHARLES J. HITCH and ROLAND N. MCKEAN, *The Economics of Defense in the Nuclear Age* (1960, reissued 1978), is the "bible" of the managerial reforms in the U.S. Defense Department; and NEVILLE BROWN, *Strategic Mobility* (1963), explores a facet of post-World War II international strategy and logistics. (R.M.Le.)

Intelligence: RICHARD DEACON, *Spyclopedia: The Comprehensive Handbook of Espionage* (1987), is an original reference work providing concise information on intelligence organizations of more than 30 countries within a chronology of 25 centuries of intelligence activity. JOHN RANELAGH, *The Agency: The Rise and Decline of the CIA,* rev. ed. (1987), offers a well-documented nonpartisan historical analysis of the organization and personalities; it can be complemented by JOHN PATRICK QUIRK *et al., The Central Intelligence Agency: A Photographic History* (1986). JAMES BAMFORD, *The Puzzle Palace: A Report on America's Most Secret Agency* (1982), explores in a well-researched investigative framework the National Security Agency's operations, using unpublished archival and official information. A general survey is given in MARK M. LOWENTHAL, *U.S. Intelligence: Evolution and Anatomy* (1984). GERALD W. HOPPLE and BRUCE W. WATSON (eds.), *The Military Intelligence Community* (1986); and SCOTT D. BRECKINRIDGE, *The CIA and the U.S. Intelligence System* (1986), examine the organizations and operations of intelligence professionals and the relevant legal and ethical problems. Further discussion of the latter is available in BRUCE W. WATSON and PETER M. DUNN (eds.), *Military Intelligence and the Universities: A Study of an Ambivalent Relationship* (1984). Useful reference information is found in JOINT CHIEFS OF STAFF, *Department of Defense Dictionary of Military Terms* (1988); and GEORGE C. CONSTANTINIDES, *Intelligence and Espionage: An Analytical Bibliography* (1983), an annotated list of about 500 important nonfiction works on a group of related topics. (B.W.Wa.)

Guerrilla warfare: ROBERT B. ASPREY, *War in the Shadows: The Guerrilla in History,* 2 vol. (1975), surveys guerrilla and counterguerrilla warfare from its origin to the 1970s. For specific guerrilla studies see REGINALD HARGREAVES, *Beyond the Rubicon: A History of Early Rome* (1967); FERDINAND SCHEVILL, *The History of the Balkan Peninsula, from the Earliest Times to the Present Day* (1922, reprinted 1971); R.C. SMAIL, *Crusading Warfare, 1097–1193* (1956, reprinted 1972); JOHN E. MORRIS, *The Welsh Wars of Edward I* (1901, reprinted 1969); C.E. CALLWELL, *Small Wars: Their Principles and Practice* (1906, reprinted 1976); JOHN R. ALDEN, *The South in the Revolution, 1763–1789* (1957, reprinted 1976); RUSSELL F. WEIGLEY, *History of the United States Army,* enlarged ed. (1984); M.R.D. FOOT, *SOE in France: An Account of the Work of the British Special Operations Executive in France, 1940–1944* (1966, reprinted 1984); OTTO HEILBRUNN, *Warfare in the Enemy's Rear* (1963); and DAVID GALULA, *Counterinsurgency Warfare: Theory and Practice* (1964).

Studies of specific campaigns include LEON WOLFF, *Little Brown Brother: How the United States Purchased and Pacified the Philippine Islands at the Century's Turn* (1961, reissued 1970); NAPOLEON D. VALERIANO, *Counter-Guerrilla Operations: The Philippine Experience* (1962); RONALD ATKIN, *Revolution!: Mexico 1910–1920* (1969); EDGAR HOLT, *Protest in Arms: The Irish Troubles, 1916–1923* (1960); FITZROY MACLEAN, *Disputed Barricade: The Life and Times of Josip Broz-Tito, Marshal of Jugoslavia* (1957); JOHN A. ARMSTRONG, *Soviet Partisans in World War II* (1964); EDGAR O'BALLANCE, *The Greek Civil War, 1944–1949* (1966); DAVID S. WOOLMAN, *Rebels in the Rif: Abd el Krim and the Rif Rebellion* (1968); RICHARD L. CLUTTERBUCK, *The Long, Long War: Counterinsurgency in Malaya and Vietnam* (1966); PETER PARET, *French Revolutionary Warfare from Indochina to Algeria: The Analysis of a Political and Military Doctrine* (1964); and BERNARD B. FALL, *The Two Viet-Nams: A Political and Military Analysis,* 2nd rev. ed. (1967, reprinted 1984). (R.B.A.)

Warsaw

Warsaw (Polish: Warszawa), the capital and largest city of Poland, is notable among Europe's capital cities not for its size, its age, or its beauty but for its indestructibility. It is a phoenix that has risen repeatedly from the ashes of war. Having suffered fearful damage during the Swedish and Prussian occupation of 1655–56, it was again assaulted in 1794, when the Russian army massacred the population of the right-bank suburb of Praga. In 1944, after the Warsaw Uprising failed, by Hitler's order the city was razed; the left-bank suburbs, controlled by the Germans, were emptied of their remaining population; and the buildings were systematically reduced to rubble by fire and dynamite. In 1945, however, the people of Warsaw, the Varsovians, returned, and the city resumed its role as the capital of Poland and the country's centre of social, political, economic, scientific, and cultural life. Many of the historic streets, buildings, and churches have been restored exactly according to their original forms. Since the second half of the 18th century, the emblem of Warsaw (originally a siren) has been a mermaid with sword and shield in hand, representing the creature who in legend led a prince to the site of Warsaw and ordered him to found the city. The city's motto is, appropriately, "Contemnit procellas" ("It defies the storms").

This article is divided into the following sections:

Physical and human geography 696
- The landscape 696
  - The city site
  - Climate
  - The city layout
- The people 696
- The economy 697
  - Industry and commerce
  - Transportation
- Administration and social conditions 697
  - Government
  - Services and health
  - Education
- Cultural life 697

History 697
- Foundation and early development 697
- Growth of the modern city 697
  - The 18th century
  - The 19th and early 20th centuries
  - World War II and contemporary Warsaw

Bibliography 698

Physical and human geography

THE LANDSCAPE

The city site. Warsaw lies on the Vistula (Wisła) River, about 240 miles (386 kilometres) southeast of the Baltic coast city of Gdansk. It is situated in the middle of the Warsaw Plain, a glacier-formed basin that ranges from 250 to 380 feet (76 to 116 metres) above sea level. Divided into right- and left-bank portions by the river, the city extends about 18 miles from north to south and 16 miles from east to west. The river is some 3,900 feet (1,190 metres) wide at this point, although the riverbed has been artificially narrowed by embankment to a third of this width.

Climate. The climate is moderate and rather cool, the prevailing westerly winds bringing frequent changes of weather. The average yearly temperature is 46° F (8° C), with a July average of 66° F (19° C) and a January average of 26° F (−3° C). Yearly rainfall averages 21 inches (541 millimetres), most of which falls in the summer. Snow cover persists for 50 to 64 days a year.

The city layout. The size of Warsaw reflects the historical fortunes of the city. From about 0.5 square mile (1.25 square kilometres) in the 17th century, it expanded to 50 square miles by 1937 and, in the postwar period, to 172 square miles by 1957. Growth has continued since. The subdivision into seven districts—Śródmieście (the city's centre), Żoliborz, Wola, Ochota, Mokotów, Praga-Południe (Praga South), and Praga-Północ (Praga North)—reflects centuries-old local names, but Warsaw is now virtually a new creation, with a layout that only partly resembles the historic city. The changes reflect a conscious planning of social and economic functions. Industries and warehouses are located on the outskirts or between modern housing developments; park areas have tripled in size; and streets, though still largely based on the old network, have been widened. The Old and New towns, Nowy Świat Avenue, and the city churches and palaces, on the other hand, have all been carefully reconstructed.

Growth of the city

Warsaw possesses a wide variety of architectural monuments, whether as replicas or originals. In the Old Town, the Gothic St. John's Cathedral and the red-brick fortifications known as the Barbican remain from the medieval period. The houses of the Old Town Market Square have been rebuilt in the splendour of their 15th-century style. There are many Baroque churches of the Counter-Reformation period, including the Jesuit Church next to the cathedral and the Church of the Holy Cross, which contains Chopin's heart. The magnificently reconstructed Royal Castle, decorated in late 18th-century style, is on Zamkowy Square. Other royal and aristocratic palaces are at Łazienkowski Park and at John III Sobieski's Wilanów. Remnants of the tsarist era are evident in the Church of St. Alexander in the middle of Trzech Krzyży Square and in the vast Alexander Citadel on the riverside, north of the New Town. The grandest of tsarist monuments, the colossal Orthodox Cathedral (1911), was demolished by the Polish government in the 1920s, but its symbolic role in the city has been assumed by the massive Palace of Culture and Science (1949), built by the Soviets south of the Old Town. The city's modern architecture is generally regarded as undistinguished. Although the prewar garden suburbs of Żoliborz and Saska Kępa have survived, the vast sprawl of contemporary suburbia is made up in large part of seemingly endless expanses of uniform, prefabricated concrete apartment blocks.

Historical monuments

Numerous historical monuments adorn Warsaw, some of which have been the object of political conflict. The postwar government was sensitive toward monuments and tended to discourage unapproved representations of people and events. For instance, the national shrine of the Tomb of the Unknown Soldier, which contains the body of a Polish youth killed in the battle for Lwów in 1919, bears only inscriptions starting with the Spanish Civil War in 1937. There is an imposing monument unveiled in 1948 in the Muranów area honouring the Heroes of the Ghetto Rising (1943), but not until 1989, the year in which Solidarity formed the country's first noncommunist national government since World War II, was there a memorial honouring the Home Army, which fought the Germans in the 1944 Warsaw Uprising. The statue of Feliks Dzerzhinsky (Polish: Dzierżyński), founder of the Soviet security police, which stood on Saxon Square on the spot where Tsar Nicholas I had raised a statue to his loyalist Polish generals, was removed in 1990. Other monuments affected by politics include the Nicolaus Copernicus statue on Ulica (street) Krakowskie Przedmieście, which was the object of struggles with the Nazi occupiers, and the statue of Frédéric Chopin, in Łazienkowski Park, which was destroyed by the Nazis in 1940 but has been reconstructed.

THE PEOPLE

The multinational population of Warsaw was transformed as a result of World War II, and today the city is com-

posed almost entirely of Poles. For centuries, though, Warsaw had been a place where the Polish-speaking Roman Catholic majority lived alongside Jews, Germans, and Russians. Early in the 20th century the largely Yiddish-speaking Jewish community accounted for almost 50 percent of the population, although it declined somewhat after 1918. The old German community, originally connected with trade and commerce, was being assimilated, however, and the Russian community, influential in the 19th century, had dwindled. From 1939 to 1945, what remained of the former diversity was destroyed. Warsaw's Jews were virtually annihilated by the Nazis, and their few remaining numbers have continued to decrease. The intelligentsia also was decimated and the administrative class connected with the prewar republic dispersed and the working class diminished by deaths and deportations. After the war Warsaw had to be completely repopulated by returning refugees, by a vast influx of peasants from the countryside, and by the families of officials connected with the new communist state. The overwhelming majority of the people are Roman Catholic.

THE ECONOMY

Industry and commerce. After 1948, when Poland's communist government was established, the largest segment of the city's labour force was employed by state-owned and cooperatively owned sectors of the national economy; manufacturing accounted for about one-third of the workers in the 1980s. Shortages of some consumer goods and food items were a problem, symbolized by the rather common sight of people standing in lines to buy goods. Electrical engineering, metallurgy (including the Warsaw steelworks), machine production (including automobile manufacturing), and toolmaking, chemical, printing, textile and clothing, and food enterprises dominate the economy. The Warsaw region is important to Poland agriculturally, for it constitutes a highly specialized gardening and vegetable centre.

Scarcity of consumer goods

Transportation. Warsaw is the hub of main rail, road, and air routes that are of importance to eastern Europe. Expressways have been built through the city along both banks of the Vistula and in the form of a ring road through the inner suburbs. Motor traffic, however, is light in comparison with that of western European cities, and it still shares the capital's main streets with a surface tramway system. The Okęcie airport, with two terminals, international and domestic, is in the Ochota district to the south.

ADMINISTRATION AND SOCIAL CONDITIONS

Government. As the capital of the Republic of Poland, Warsaw houses all the central institutions of the national government. The Sejm, Poland's national legislature, is not far from the crossroads of Nowy Świat and Aleje Jerozolimskie. South of Łazienkowski Park is Belweder (Belvedere) Palace, the office of the president of the national government. The government of Warsaw is run by an elected City Council, headed by a city president. Warsaw's seven subdivisions also have their own elected legislatures. Until 1990 the city administration was only nominally elective and subject to the Warsaw Committee of the Polish United Workers' Party, the country's communist party. The city is also the administrative centre of the Warsaw Capital Province (Województwo Stołeczne Warszawskie).

Services and health. Like most large cities, Warsaw is continually expanding its infrastructure to keep pace with its growth. A postwar housing shortage was alleviated with prefabricated housing units, and housing construction in the suburbs and future planning have continued. As in many European countries, medical and health services are virtually free for all citizens. Hospitals, outpatient clinics, and medical research facilities are widespread.

Education. Education in Warsaw benefits from the presence of the headquarters of the Polish Academy of Sciences, which coordinates research in both physical and social sciences through a number of institutes and industrial establishments. The Technical University of Warsaw and the University of Warsaw are notable institutions. Major libraries include the library (established in 1817) of the University of Warsaw and the National Library (1919); there are also a number of specialist libraries.

CULTURAL LIFE

Warsaw's writers, artists, and musicians play a major role in creating the cultural values of the nation. The majority of the members of such bodies as the Polish Union of Writers and the Polish Union of Composers work in the city, which is also the seat of such prominent institutions as the National Museum and the Zachęta Art Gallery. Poland's leading theatre and radio and television operations are centred in Warsaw. The National Philharmonic orchestra draws large crowds. There are numerous specialist museums, and, among the many social, cultural, and educational associations, the Friends of History in Warsaw and the Chopin societies are prominent.

A centre of Polish culture

Extensive recreational facilities exist in and around the city. Several large and many smaller parks provide open space to accommodate a variety of outdoor activities. Indoor and outdoor swimming pools, sport and physical culture centres, and ice rinks are prevalent. Major sporting and other events take place in a number of stadiums, the largest being the 100,000-capacity Tenth Anniversary Stadium, built to commemorate a decade of reconstruction following World War II. (Ja.D./A.H.D./N.D./Ed.)

History

FOUNDATION AND EARLY DEVELOPMENT

The origins of Warsaw remain obscure. Excavations within present urban limits have confirmed the existence of Stare Bródno, a small trading settlement of the 10th and early 11th centuries AD. Its functions were taken over successively by Kamion (*c.* 1065) and Jazdow (first recorded in 1262). About the end of the 13th century, Jazdow was moved about two miles to the north, to a village named Warszowa (Warsaw), and the community was strengthened by the protection of a castle. From 1339, authority was invested in a bailiff and, from 1376, in a city council. By the end of the 14th century, the growing settlement had a double line of protective ramparts.

In the 15th century the town became the capital of the duchy of Mazovia, and the New Town sprang up to the north of the original, constricted site, afterward known as the Old Town. In 1526 both city and province became incorporated into the kingdom of Poland; from 1569 the Sejm met in Warsaw, and from 1573 the elections of the kings took place there. The first permanent bridge was built across the Vistula in 1573, and in 1596 King Sigismund III Vasa began to remodel the castle as a royal residence. In 1611 the king and his court finally moved from Kraków (Cracow) to Warsaw, making it the capital of the Polish state. Powerful persons built residences in Warsaw, and autonomous settlements sprang up around its periphery. This growth proved short-lived, for a Swedish invasion (1655–56) devastated the flourishing city. Afterward the War of the Polish Succession (1733–38) brought economic decay and pestilence.

The Old and New towns

GROWTH OF THE MODERN CITY

The 18th century. The development of manufacturing, banks, and other enterprises during the early 18th century provided a firm economic base for a number of early exercises in urban planning. During the reign (1764–95) of King Stanisław II August Poniatowski, Warsaw became the centre of the Polish Enlightenment. The first lay school (Szkoła Rycerska) was opened in 1765, and in 1773 the Committee on National Education began its activity in the city. The Polish theatre and numerous printing establishments also flourished. Warsaw played an important role in the striving for Polish political rebirth (following the first partition of Poland by Austria, Prussia, and Russia in 1772) when a parliamentary constitution was proclaimed in the city on May 3, 1791. The national insurrection of 1794, led by Tadeusz Kościuszko against the Russo-Prussian second partition, was brutally crushed; the ensuing third partition of Poland among Russia, Prussia, and Austria left Warsaw a provincial town of South Prussia.

The 19th and early 20th centuries. The occupation of

Houses in the Old Town Market Square, which were rebuilt after World War II in the style of their 15th-century originals.
© Hubertus Kanus—Photo Researchers, Inc.

Warsaw by the French Army in 1806 was followed a year later by Napoleon's creation of the duchy of Warsaw, a measure that reinstated the city as a capital. The emperor's downfall, however, led to occupation by Russia.

Nationalist uprisings and industrial growth

Under the Congress Kingdom of Poland (1815–30), formed by the Congress of Vienna and ruled by the Russian tsar, real national independence remained an elusive goal, and Warsaw figured prominently in struggles to throw off foreign domination. The Russo-Polish War of 1831, in which the Poles held the initial advantage, ended with the storming of Warsaw's defenses and the initiation of 30 years of military rule. Economic growth, however, continued, and by the mid-19th century textile, metal, and tannery industries were well established and the city had become the centre of a continentwide rail network. A fresh cycle of insurrection, which broke out in January 1863, led to the brief existence of an underground National Government in Warsaw. The suppression of the revolt led in 1864 to the abolition of the Congress Kingdom. Warsaw became a provincial city of the Russian Empire's "Vistula Land." A period of Russification was launched, and Varsovians abandoned national politics for a period of "positivism," which aimed to preserve Polish culture at all costs. Population soared, reaching 756,000 by 1903, and urban services underwent extensive modification. Despite revolutionary activity between 1905 and 1907, censorship was alleviated, Polish schools and other cultural institutions were established, and Warsaw experienced a cultural renaissance. After World War I the city regained its status as the national capital.

World War II and contemporary Warsaw. The population of Warsaw passed the 1,000,000 mark in 1925. The period between the world wars was marked by further advances despite a period of inflation, depressions and slumps in trade, and political instability. Automobile and aircraft manufacture were introduced; city services underwent further expansion, and Warsaw's emergence as a European cultural centre was symbolized by the beginning of such international competitions as the International Chopin Competition for Pianists (1927) and the Henryk Wieniawski International Violin Competition (1935).

During the Nazi siege of Warsaw in 1939, more than 10,000 citizens perished and more than 50,000 were wounded before the lack of supplies forced a surrender. The subsequent Nazi occupation was aimed at reducing Warsaw to a provincial city; its cultural treasures were systematically plundered and its inhabitants carried away to German labour camps or to concentration camps.

Warsaw's Jewish community was devastated. At the turn of the century the city had contained the largest urban concentration of Jews in the world. Warsaw's Jews were heavily represented in the city's bourgeoisie and intelligentsia, especially in the commercial classes and the professions. In addition to full religious and political freedom, Warsaw's Jewry enjoyed its own press, its own Yiddish theatre, and its own Jewish schools. Conflict with the rest of Warsaw society existed, but it was not alarming until immediately before the war.

Following German entry into Warsaw, a Jewish ghetto was established, surrounded by a high wall. Disease, starvation, and overcrowding caused thousands to perish before deportations to the Nazi death camps, especially to Treblinka, began in late 1941. Some 312,000 Jews were sent to the gas chambers in 1942 alone. The Warsaw Ghetto Uprising of April 1943, in which more than 60,000 Jews died, signaled a last, heroic act of defiance in the face of impending annihilation. The demolition by the Nazis of the Great Synagogue (now restored) near Aleja (avenue) General Karol Świerczewskiego symbolized the end of six centuries of Jewish Warsaw.

The Warsaw resistance

Near the war's end Warsaw became the national resistance centre. With the Soviet forces nearing the city, the Warsaw Home Army rose up against the weakened German garrison and was near victory when German reinforcements arrived and wiped out the Polish resistance. From 150,000 to 180,000 of the city's people died in the Warsaw Uprising of 1944. In all, some 600,000 to 800,000 Varsovians are estimated to have died between 1939 and 1944, and the Soviet armies in 1945 found the city in a state of almost total devastation. On Feb. 1, 1945, the provisional Polish government made Warsaw its capital, and an office for urban reconstruction was set up. In the decades that followed the city was rebuilt; the scars of war almost entirely healed over, and the city expanded beyond its prewar size in both area and population.

(Ja.D./A.H.D./N.D.)

BIBLIOGRAPHY. The first three published volumes of the Polish-language series *Dzieje Warszawy*, a comprehensive survey of Warsaw published by the Polish Academy of Sciences, are MARIA BOGUCKA *et al., Warszawa w latach 1526–1795* (1984); STEFAN KIENIEWICZ, *Warszawa w latach 1795–1914* (1976), and KRZYSZTOF DUNIN-WASOWICZ, *Warszawa w latach 1939–1945* (1984). ROBERT E. DICKINSON, *The West European City,* 2nd ed. (1961), contains a description of the site and layout of the medieval town, the postmedieval extensions, and the pre-1940 city; JAN ZACHWATOWICZ and PIOTR BIEGAŃSKI, *The Old Town of Warsaw,* trans. from the Polish (1956), is a pictorial record; A. CIBOROWSKI, *Warsaw: A City Destroyed and Rebuilt,* 2nd ed. (1969, originally published in Polish, 1964), is a history of Warsaw and includes a description of the city as it was in 1939 and the story of its destruction. See also STANISLAW JANKOWSKI and ADOLF CIBOROWSKI, *Warsaw 1945, Today, and Tomorrow* (1978). OLGIERD BUDREWICZ and EDWARD FALKOWSKI, *Homo Varsoviensis* (*Warsaw Man*), trans. from the Polish (1970), emphasizes postwar reconstruction. EDWARD D. WYNOT, *Warsaw Between the World Wars: Profile of the Capital City in a Developing Land, 1918–1939* (1983), is a historical survey. JOANNA K.M. HANSON, *The Civilian Population and the Warsaw Uprising of 1944* (1982); JAN M. CIECHANOWSKI, *The Warsaw Rising of 1944* (1984; originally published in Polish, 1971); and ISRAEL GUTMAN, *The Jews of Warsaw, 1939–1943: Ghetto, Underground, Revolt,* trans. from the Hebrew (1982), are scholarly historical discussions. A more comprehensive historical overview, in Polish, is JAN STANISLAW BYSTROŃ, *Warszawa,* 2nd ed. (1977). JANINA RUTKOWSKA, *A Guide to Warsaw and Environs,* 3rd ed. (1981; originally published in Polish, 1976), is a tourist book with illustrations and maps. The following are pictorial works, with texts in English translation: EDMUND KUPIECKI, *Warsaw: Landscape and Architecture,* 3rd ed. (1970); JULIUSZ A. CHROSCICKI and ANDRZEJ ROTTERMUND, *Atlas of Warsaw's Architecture* (1978); KRZYSZTOF JABLONSKI, *Warsaw, a Portrait of the City* (1979); DOBROSLAW KOBIELSKI, *Warsaw from a Bird's-Eye View* (1971).

(Ed.)

George Washington

General, statesman, and first president of the United States, George Washington was born in Westmoreland County, Virginia, on February 22 (February 11, old style), 1732. His father was Augustine Washington, who had gone to school in England, had tasted seafaring life, and was then managing his growing Virginia estates. His mother was Mary Ball, whom Augustine, a widower, had married early the previous year. The paternal lineage had some distinction; an early forebear was described as "gentleman," Henry VIII later gave the family lands, and its members held various offices. But family fortunes fell with the Puritan revolution in England, and John Washington, grandfather of Augustine, migrated in 1657 to Virginia. The ancestral home at Sulgrave, Northamptonshire, is maintained as a Washington memorial. Little definite information exists on any of the line until Augustine. He was an energetic, ambitious man who acquired much land, built mills, took an interest in opening iron mines, and sent his two oldest sons to England for schooling. By his first wife, Jane Butler, he had four children; by his second, six. Augustine died April 12, 1743.

Childhood and youth. Little is known of George Washington's early childhood, spent largely on the Ferry Farm on the Rappahannock River, opposite Fredericksburg, Virginia. Mason L. Weems's stories of the hatchet and cherry tree and of young Washington's repugnance to fighting are apocryphal efforts to fill a manifest gap. He attended

Schooling

school irregularly from his seventh to his 15th year, first with the local church sexton and later with a schoolmaster named Williams. Some of his schoolboy papers survive. He was fairly well trained in practical mathematics—gauging, several types of mensuration, and such trigonometry as was useful in surveying. He studied geography, possibly had a little Latin, and certainly read some of *The Spectator* and other English classics. The copybook in which he transcribed at 14 a set of moral precepts or *Rules of Civility and Decent Behaviour in Company and Conversation* was carefully preserved. His best training, however, was given him by practical men and outdoor occupations, not by books. He mastered tobacco growing and stock raising, and early in his teens he was sufficiently familiar with surveying to plot the fields about him.

Lawrence Washington

At his father's death, the 11-year-old boy became the ward of his eldest half brother, Lawrence, a man of fine character who gave him wise and affectionate care. Lawrence inherited the beautiful estate of Little Hunting Creek, which had been granted to the original settler, John Washington, and which Augustine had done much since 1738 to develop. Lawrence married Anne (Nancy) Fairfax, daughter of Col. William Fairfax, cousin and agent of Lord Fairfax, one of the chief proprietors of the region. Lawrence also built a house and named the 2,500-acre holding Mount Vernon, in honour of the admiral under whom he had served in the siege of Cartagena. Living there chiefly with Lawrence (though he spent some time with his other half brother, Augustine, called Austin, near Fredericksburg), George entered a more spacious and polite world. Anne Fairfax Washington was a woman of charm, grace, and culture; Lawrence had brought from his English school and his naval service much knowledge and experience. A valued neighbour and relative, George William Fairfax, whose large estate, Belvoir, was about four miles distant, and other relatives by marriage, the Carlyles of Alexandria, helped form George's mind and manners.

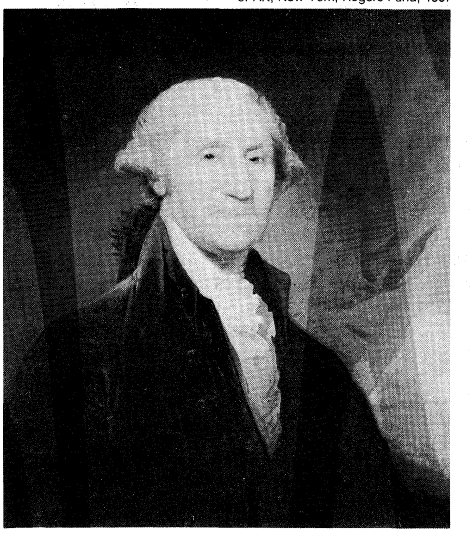
The youth turned first to surveying as a profession. Lord Fairfax, a middle-aged bachelor who owned more than 5,000,000 acres in northern Virginia and the Shenandoah Valley, came to America in 1746 to live with his cousin George William at Belvoir and to look after his properties. Two years later he sent to the Shenandoah Valley a party to survey and plot his lands to make regular tenants of the squatters moving in from Pennsylvania. With the official surveyor of Prince William County in charge, Washington went along as assistant. The 16-year-old lad kept a disjointed diary of the trip, which shows skill in observation. He describes the discomfort of sleeping under "one thread Bear blanket with double its Weight of Vermin such as Lice Fleas & c"; an encounter with an Indian war party bearing a scalp; the Pennsylvania-German emigrants, "as ignorant a set of people as the Indians they would never speak English but when spoken to they speak all Dutch"; and the serving of roast wild turkey on "a Large Chip," for "as for dishes we had none."

The following year (1749), aided by Lord Fairfax, Washington received an appointment as official surveyor of Culpeper county, and for more than two years he was kept almost constantly busy. Surveying not only in Culpeper but also in Frederick and Augusta counties, he made journeys far beyond the tidewater region into the western wilderness. The experience taught him resourcefulness and endurance and toughened both body and mind. Coupled with his half brother Lawrence's ventures in land, it also gave him an interest in western development that endured throughout his life. He was always disposed to speculate in western holdings and to view favourably projects for colonizing the West, and he greatly resented the limitations that the crown in time laid on the westward movement. In 1752 Lord Fairfax determined to take up his final residence in the Shenandoah Valley and settled there in a log hunting lodge, which he called Greenway Court, after a Kentish manor of his family. There Washington was sometimes entertained and had access to a small library that Fairfax had begun accumulating at Oxford.

Mount Vernon

The years 1751–52 marked a turning point in Washington's life, for they placed him in control of Mount Vernon. His half brother Lawrence, stricken by tuberculosis, went to Barbados in 1751 for his health, taking George along. From this sole journey beyond the present borders of the United States, Washington returned with the light scars of an attack of smallpox. In July of the next year, Lawrence died, making George executor and residuary heir of his estate in the event of the decease of his daughter, Sarah, without issue. As she died within two months, Washington at the age of 20 became head of one of the best Virginia estates. He always thought farming the "most delectable" of pursuits. "It is honorable," he wrote, "it is amusing,

By courtesy of the Metropolitan Museum of Art, New York, Rogers Fund, 1907



Washington, oil painting by Gilbert Stuart, *c.* 1795. In the Metropolitan Museum of Art, New York.

and, with superior judgment, it is profitable." And of all the spots for farming, he thought Mount Vernon the best. "No estate in United America," he assured an English correspondent, "is more pleasantly situated than this." His greatest pride in later days was to be regarded as the first farmer of the land.

He gradually increased the estate until it exceeded 8,000 acres. He enlarged the house in 1760 and made further enlargements and improvements on the house and its landscaping in 1784–86. He tried to keep abreast of the latest scientific advances.

For the next 20 years the main background of Washington's life was the work and society of Mount Vernon. He had to manage the 18 slaves that came with the estate and others he bought later; by 1760 he paid tithes on 49 slaves—though he strongly disapproved of the institution and hoped for some mode of abolishing it. He gave assiduous attention to the rotation of crops, fertilization of the soil, and the management of livestock.

For diversion he was fond of riding, fox hunting, and dancing; of such theatrical performances as he could reach; and of duck hunting and sturgeon fishing. He liked billiards and cards and not only subscribed to racing associations but ran his own horses in races. In all outdoor pursuits, from wrestling to colt breaking, he excelled. A friend of the 1750s describes him as "straight as an Indian, measuring six feet two inches in his stockings"; as very muscular and broad shouldered, but though large boned, weighing only 175 pounds; and as having long arms and legs. His penetrating blue-gray eyes were overhung by heavy brows, his nose was large and straight, and his mouth was large and firmly closed. "His movements and gestures are graceful, his walk majestic, and he is a splendid horseman." He soon became prominent in community affairs, was an active member and later vestryman of the Episcopal Church, and as early as 1755 expressed a desire to stand for the Virginia House of Burgesses.

PREREVOLUTIONARY MILITARY AND POLITICAL CAREER

Early military career. Traditions of John Washington's feats as Indian fighter and Lawrence Washington's talk of service days helped imbue George with military ambition. Just after Lawrence's death, Lieut. Gov. Robert Dinwiddie appointed George adjutant for the southern district of Virginia at £100 a year (November 1752). The next year he became adjutant of the Northern Neck and Eastern Shore. Then in 1753 Dinwiddie found it necessary to warn the French to desist from their encroachments on Ohio Valley lands claimed by the crown; and after sending one messenger who failed to reach the goal, he determined to dispatch Washington. On the day he received his orders, October 31, 1753, Washington set out for the French posts. His party consisted of a Dutchman to serve as interpreter, the expert scout Christopher Gist as guide, and four others, two of them experienced traders with the Indians. Theoretically, Great Britain and France were at peace; but actually war impended, and Dinwiddie's message was an ultimatum: the French must get out or they would be put out.

Ohio Valley expedition

The journey proved rough, perilous, and futile. Washington's party left what is now Cumberland, Maryland, in the middle of November and despite wintry weather and wilderness impediments reached Fort-Le Boeuf, at what is now Waterford, Pennsylvania, 20 miles south of Lake Erie, without delay. The French commander was courteous but adamant. As Washington reported, his officers "told me, That it was their absolute Design to take possession of the Ohio, and by God they would do it." Eager to carry this alarming news back, Washington pushed off hurriedly with Gist. He was lucky to get back alive. An Indian fired at them at 15 paces but missed; when they crossed the Allegheny River on a raft, Washington was jerked into the ice-filled stream but saved himself by catching one of the timbers. That night he almost froze in his wet clothing. He reached Williamsburg on January 16, 1754, where he hastily penned a record of the journey. Dinwiddie, who was labouring to convince the crown of the seriousness of the French threat, had it printed; and when he sent it to London, it was reprinted in three different forms.

The enterprising governor forthwith planned an expedition to hold the Ohio country. He made Joshua Fry colonel of a provincial regiment, appointed Washington lieutenant colonel, and set them to recruiting troops. Two agents of the Ohio Company, which Lawrence Washington and others had formed to develop lands on the upper Potomac and Ohio rivers, had begun building a fort at what later became Pittsburgh. Dinwiddie, ready to launch into his own war, sent Washington with two companies to reinforce this post. In April 1754 the lieutenant colonel set out from Alexandria with about 160 men at his back. He marched to Cumberland only to learn that the French had anticipated the British blow; they had taken possession of the fort of the Ohio Company and had renamed it Fort-Duquesne. Happily, the Indians of the area offered support. Washington therefore struggled cautiously forward to within about 40 miles of the French position and erected his own post at Great Meadows, near what is now Confluence, Pennsylvania. With this as base, he made a surprise attack (May 28, 1754) upon an advance detachment of 30 French, killing the commander, Coulon de Jumonville, and nine others and making the rest prisoners. The last French and Indian War had begun.

The last French and Indian War

Washington at once received promotion to a full colonelcy and was reinforced, commanding a considerable body of Virginia and North Carolina troops, with Indian auxiliaries. But his attack soon brought the whole French force down upon him. They drove his 350 men into the Great Meadows fort (Ft. Necessity) on July 3, besieged it with 700 men, and, after an all-day fight, compelled him to surrender. The construction of the fort had been a blunder, for it lay in a waterlogged creek bottom, was commanded on three sides by forested elevations approaching it closely, and was too far from Washington's supports. The French agreed to let the disarmed colonials march back to Virginia with the honours of war, but they compelled Washington to promise that Virginia would not build another fort on the Ohio for a year and to sign a paper acknowledging responsibility for "*l'assassinat*" of de Jumonville, a word which Washington later explained he did not rightly understand. He returned to Virginia, chagrined but proud, to receive the thanks of the House of Burgesses, and to find that his name had been mentioned in the London gazettes. His remark in a letter to his brother that "I have heard the bullets whistle; and believe me, there is something charming in the sound" was commented on humorously by Horace Walpole and sarcastically by George II.

The arrival of Gen. Edward Braddock and his army in Virginia in February 1755, as part of the triple plan of campaign that called for his advance on Fort-Duquesne, Gov. William Shirley's capture of Niagara, and Sir William Johnson's capture of Crown Point, brought Washington new opportunities and responsibilities. He had resigned his commission in October 1754 in resentment of the slighting treatment and underpayment of colonial officers and particularly because of an untactful order of the British war office that provincial officers of whatever rank should be subordinate to any officer holding the king's commission. But he ardently desired a part in the war; "my inclinations," he wrote a friend, "are strongly bent to arms." When Braddock showed appreciation of his merits and invited him to join the expedition as personal aide-de-camp, with the courtesy title of colonel, he therefore accepted. His self-reliance, decision, and masterful traits soon became apparent.

At table he had frequent disputes with Braddock, who when contractors failed to deliver their supplies attacked the colonials as supine and dishonest while Washington defended them warmly. His freedom of utterance is proof of Braddock's esteem. Braddock accepted from him the unwise advice that he divide his army, leaving half of it to come up with the slow wagons and cattle train and taking the other half forward against Fort-Duquesne at a rapid pace. Washington was ill with fever during June but joined the advance guard in a covered wagon on July 8, begged to lead the march on Fort-Duquesne with his Virginians and the Indian allies, and was by Braddock's side when on July 9 the army was ambushed and bloodily defeated.

In this defeat Washington displayed the combination of coolness and determination, the alliance of unconquerable energy with complete poise, that was the secret of so many of his successes. So ill that he had to use a pillow instead of a saddle and that Braddock ordered his body servant to keep special watch over him, he was everywhere at once. At first he followed Braddock as the general bravely tried to rally his men to push either forward or backward, the wisest course the circumstances permitted. Then he rode back to bring up the Virginians from the rear and rallied them with effect on the flank. To him was largely due the escape of the force. His exposure of his person was as reckless as Braddock's, who was fatally wounded on his fifth horse; Washington had two horses shot under him and his clothes cut by four bullets without being hurt. He was at Braddock's deathbed, helped bring the troops back, and was repaid by being appointed, in August 1755, while still only 23 years old, commander of all the Virginia troops. But no part of his later service was conspicuous. Finding that a Maryland captain who held a royal commission would not obey him, he rode north in February 1756 to Boston to have the question settled by the commander in chief in America, Governor Shirley, and, bearing a letter from Dinwiddie, had no difficulty in carrying his point. On his return he plunged into a multitude of vexations. He had to protect a weak, thinly settled frontier nearly 400 miles in length with only some 700 ill-disciplined colonial troops; to cope with a legislature unwilling to support him; to meet attacks on the drunkenness and inefficiency of the soldiers; and to endure constant wilderness hardships. It is not strange that in 1757 his health failed and in the closing weeks of that year he was so ill of a "bloody flux" that his physician ordered him home to Mount Vernon.

Commander of the Virginia troops

In the spring of 1758 he recovered sufficiently to return to duty as colonel in command of all Virginia troops. As part of the grand sweep of several British armies organized by Pitt, Gen. John Forbes led a new advance upon Fort-Duquesne. This time Forbes resolved not to use Braddock's road but to cut a new one west from Raystown, Pennsylvania. Washington disapproved of the route but played an important part in the movement. Late in the autumn the French evacuated and burned Fort-Duquesne, and Forbes reared Ft. Pitt on the site. Washington, who had just been elected to the House of Burgesses, was able to resign with the honorary rank of brigadier general.

But though his officers expressed regret at the "loss of such an excellent Commander, such a sincere Friend, and so affable a Companion," he quit the service with a sense of frustration. He had thought the war excessively slow. The Virginia legislature had been niggardly in voting money; the Virginia recruits had come forward reluctantly and had proved of poor quality—he had hanged a few deserters and flogged others heavily. Virginia gave him less pay than other colonies offered their troops. Desiring a regular commission such as his half brother Lawrence had held, he applied in vain to the British commander in North America, Lord Loudoun, to make good a promise that Braddock had given him. Ambitious for both rank and honour, he showed a somewhat strident vigour in asserting his desires and in complaining when they were denied. He returned to Mount Vernon somewhat disillusioned.

Resignation from the colonial army

Marriage and plantation life. Immediately on resigning his commission he was married (January 6, 1759) to Martha Dandridge, the widow of Daniel Parke Custis. She was a few months older than he, was the mother of two children living and two dead, and possessed one of the considerable fortunes of Virginia. Washington had met her the previous March and had asked for her hand before his campaign with Forbes. Though it does not seem to have been a romantic love match, the marriage united two harmonious temperaments and proved happy. Martha was a good housewife, an amiable companion, and a dignified hostess.

Some estimates of the property brought him by this marriage have been exaggerated, but it did include a number of slaves and about 15,000 acres, much of it valuable for its proximity to Williamsburg. More important to Washington were the two stepchildren, John Parke ("Jacky") and Martha Parke ("Patsy") Custis, who at the time of the marriage were six and four, respectively. He lavished great affection and care upon them, worried greatly over Jacky's waywardness, and was overcome with grief when Patsy died just before the Revolution. Jacky died during the war, leaving four children. Washington adopted two of them, a boy and a girl, and even signed his letters to the boy as "your papa." Himself childless, he thus had a real family.

From the time of his marriage Washington added to the care of Mount Vernon the supervision of the Custis estate at the White House on the York River. As his holdings expanded they were divided into farms, each under its own overseer; but he minutely inspected operations every day and according to one visitor often pulled off his coat and performed ordinary labour. As he once wrote, "middling land under a man's own eyes, is more profitable than rich land at a distance." To the eve of the Revolution he devoted himself to the duties and pleasures of a great landholder, varied by several weeks' attendance every year in the House of Burgesses in Williamsburg. During 1760–74 he was also a justice of the peace for Fairfax County, sitting in court in Alexandria.

In no light does Washington appear more characteristically than as one of the richest, largest, and most industrious of Virginia planters. For six days a week he rose early and worked hard; on Sundays he irregularly attended Pohick Church (16 times in 1760), entertained company, wrote letters, made purchases and sales, and sometimes went fox hunting. In these years he took snuff and smoked a pipe; throughout life he liked Madeira wine and punch. Though wheat and tobacco were his staples, he practiced crop rotation on a three-year or five-year plan. He had his own waterpowered flour mill, blacksmith shop, brick and charcoal kilns, carpenters, and masons. His fishery supplied shad, bass, herring, and other catches, salted as food for the Negroes. Coopers, weavers, and his own shoemaker turned out barrels; cotton, linen, and woollen goods; and brogans for all needs. In short, his estates, in accordance with his orders to overseers to "buy nothing you can make yourselves," were largely self-sufficient communities. But he did send large orders to England for farm implements, tools, paint, fine textiles, hardware, and agricultural books and hence was painfully aware of British commercial restrictions.

He experimented in breeding cattle; acquired at least one buffalo, with the hope of proving its utility as a meat animal; and kept stallions at stud. He also took pride in a peach and apple orchard. His care of slaves was exemplary. He carefully clothed and fed them, engaged a doctor for them by the year, refused to sell them—"I am principled against this kind of traffic in the human species"—and administered correction mildly. They showed so much attachment that few ran away.

In the social life of the tidewater region he meanwhile played a prominent role. The members of the council and House of Burgesses, a roster of influential Virginians, were all friends. He visited the Byrds of Westover, the Lees of Stratford, the Carters of Shirley and Sabine Hall, and the Lewises of Warner Hall; Mount Vernon often was busy with guests in return. He liked house parties and afternoon tea on the Mount Vernon porch; he was fond of picnics, barbecues, and clambakes; and throughout life he enjoyed dancing, frequently going to Alexandria for balls. Cards were a steady diversion, and his accounts record sums lost at them, the largest reaching nearly £10. In bad weather his diary sometimes states, "at home all day, over cards." Billiards was a rival amusement. Not only the theatre, when available, but concerts, cockfights, the circus, puppet shows, and exhibitions of animals received his patronage.

He insisted on the best clothes—coats, laced waistcoats, hats, coloured silk hose—bought in London. The Virginia of the Randolphs, Corbins, Harrisons, Tylers, Nicholases, and other prominent families had an aristocratic quality, and Washington liked to do things in a large way. It has been computed that in the seven years prior to 1775, Mount Vernon had 2,000 guests, most of whom stayed to dinner if not overnight.

Prerevolutionary politics. Washington's contented life was interrupted by the rising storm in imperial affairs. The

British ministry, facing a heavy postwar debt, high home taxes, and continued military costs in America, decided in 1764 to obtain revenue from the colonies. Up to that time, Washington, though regarded by associates, in Col. John L. Peyton's words, as "a young man of an extraordinary and exalted character," had shown no signs of personal greatness and few signs of interest in state affairs. The Proclamation of 1763 interdicting settlement beyond the Alleghenies irked him, for he was interested in the Ohio Company, the Mississippi Company, and other speculative western ventures. He nevertheless played a silent part in the House of Burgesses and was a thoroughly loyal subject.

Anglo-American tensions

But he was present when Patrick Henry introduced his resolutions against the Stamp Act in May 1765 and shortly thereafter gave token of his adherence to the cause of the colonial Whigs against the Tory ministries of England. In 1768 he told George Mason at Mount Vernon that he would take his musket on his shoulder whenever his country called him. The next spring, April 4, 1769, he sent Mason the Philadelphia nonimportation resolutions with a letter declaring that it was necessary to resist the strokes of "our lordly masters" in England; that courteous remonstrances to Parliament having failed, he wholly endorsed the resort to commercial warfare; and that as a last resort no man should scruple to use arms in defense of liberty. When, the following May, the royal governor dissolved the House of Burgesses, he shared in the gathering at the Raleigh tavern that drew up nonimportation resolutions, and he went further than most of his neighbours in adhering to them. At that time and later he believed with most Americans that peace need not be broken.

Late in 1770 he paid a land-hunting visit to Ft. Pitt, where George Croghan was maturing his plans for the proposed 14th colony of Vandalia. Washington directed his agent to locate and survey 10,000 acres adjoining the Vandalia tract, and at one time he wished to share in certain of Croghan's schemes. But the Boston Tea Party of December 1773 and the bursting at about the same time of the Vandalia bubble turned his eyes back to the East and the threatening state of Anglo-American relations. He was not a member of the Virginia committee of correspondence formed in 1773 to communicate with other colonies, but when the Virginia legislators, meeting irregularly again at the Raleigh tavern in May 1774, called for a Continental Congress, he was present and signed the resolutions. Moreover, he was a leading member of the first provincial convention or revolutionary legislature late that summer, and to that body he made a speech that was much praised for its pithy eloquence, declaring that "I will raise one thousand men, subsist them at my own expense, and march myself at their head for the relief of Boston."

Delegate to the first Continental Congress

The Virginia provincial convention promptly elected Washington one of the seven delegates to the first Continental Congress. He was by this time known as a radical rather than a moderate, and in several letters of the time he opposed a continuance of petitions to the British crown, declaring that they would inevitably meet with a humiliating rejection. " Shall we after this whine and cry for relief when we have already tried it in vain?" he wrote. When the congress met in Philadelphia on September 5, 1774, he was in his seat in full uniform, and his participation in its councils marks the beginning of his national career.

His letters of the period show that while still utterly opposed to the idea of independence, he was determined never to submit "to the loss of those valuable rights and privileges, which are essential to the happiness of every free State, and without which life, liberty, and property are rendered totally insecure." If the ministry pushed matters to an extremity, he wrote, "more blood will be spilled on this occasion than ever before in American history." Though he served on none of the committees, he was a useful member, his advice being sought on military matters and weight being attached to his advocacy of a nonexportation as well as nonimportation agreement. He also helped to secure approval of the "Suffolk Resolves," which looked toward armed resistance as a last resort and which did much to harden the king's heart against America.

Returning to Virginia in November, he took command of the volunteer companies drilling there and served as chairman of the committee of safety in Fairfax County. The unanimity with which the Virginia troops turned to him, though the province contained many experienced officers and Col. William Byrd of Westover had succeeded Washington as commander in chief, was a tribute to his reputation and personality; it was understood that Virginia expected him to be its general. At the March 1775 session of the legislature he was elected to the second Continental Congress and again set out for Philadelphia.

REVOLUTIONARY LEADERSHIP

Head of the colonial forces. Washington's choice as commander in chief of the military forces of all the colonies followed immediately upon the first fighting, though it was by no means inevitable and was the product of partly artificial forces. The Virginia delegates differed upon his appointment. Washington himself recommended Gen. Andrew Lewis for the post, and Edmund Pendleton was, according to John Adams, "very full and clear against it." It was chiefly the fruit of a political bargain by which New England offered Virginia the chief command as its price for the adoption and support of the New England army. This army had gathered hastily and in force about Boston immediately after the clash of British troops and American minutemen at Lexington and Concord on April 19, 1775. When the second Continental Congress met in Philadelphia on May 10, one of its first tasks was to find a permanent leadership for this force. On June 15, Washington, whose military counsel had already proved invaluable on two committees, was nominated and chosen by unanimous vote. Beyond the considerations noted, he owed his choice to the facts that Virginia stood with Massachusetts as one of the most powerful colonies; that his appointment would augment the zeal of the southern people; that he had made an enduring reputation in the Braddock campaign; and that his poise, sense, and resolution had impressed all the delegates. The scene of his election, with Washington darting modestly into an adjoining room and John Hancock flushing with jealous mortification, will alway impress the historical imagination; so also will the scene of July 3, 1775, when wheeling his horse under an elm in front of the troops paraded on Cambridge common he drew his sword and took command of the army investing Boston. News of Bunker Hill had reached him before he was a day's jouney from Philadelphia, and he had expressed confidence of victory when told how the militia had fought. In accepting the command he refused any payment beyond his expenses and called upon "every gentleman in the room" to bear witness that he disclaimed fitness for it. At once he showed characteristic decision and energy in organizing the raw volunteers, collecting provisions and munitions, and rallying Congress and the colonies to his support.

Election as colonial commander

The first phase of Washington's command covered the period from July 1775 to the British evacuation of Boston in March 1776. In those eight months he imparted discipline to the army, which at maximum strength slightly exceeded 20,000; he dealt with subordinates who, as John Adams said, quarrelled "like cats and dogs"; and he kept the siege vigorously alive. Having himself planned an invasion of Canada by Lake Champlain, to be entrusted to Gen. Philip Schuyler, he heartily approved of Benedict Arnold's proposal to march north along the Kennebec River and take Quebec. Giving Arnold 1,100 men, he instructed him to do everything possible to conciliate the Canadians. He was equally active in encouraging privateers to atack British commerce. As fast as means offered, he strengthened his army with ammunition and siege guns, bringing heavy artillery from Ticonderoga over the frozen roads early in 1776. His position was at first precarious, for the Charles River pierced the centre of his lines investing the town; and if the British general, Sir William Howe, had moved his 20 veteran regiments boldly up the stream, he might have pierced Washington's army and rolled either wing back to destruction. But all the generalship was on Washington's side. Seeing that Dorchester heights, just south of Boston, commanded the city and harbour and that Howe had unaccountably failed to occupy it, he seized it on the night of March 4, 1776, placing his Ticonderoga guns in

position. The British naval commander declared that he could not remain if the Americans were not dislodged, and Howe, after a storm disrupted his plans for an assault, evacuated the city on March 17. He left 200 cannon and invaluable stores of small arms and munitions. After stamping out the smallpox in Boston and collecting his booty, Washington hurried south to take up the defense of New York.

Washington had won the first round, but there remained five years of the war, during which the American cause was repeatedly near complete disaster. It is unquestionable that Washington's strength of character, his ability to hold the confidence of army and people and to diffuse his own courage among them, his unremitting activity, and his strong common sense constituted the chief factors in achieving American victory. He was not a great tactician: as Jefferson said later, he often "failed in the field"; he was sometimes guilty of grave military blunders, the chief being his assumption of a position on Long Island in 1776 that exposed his entire army to capture the moment it was defeated. At the outset he was painfully inexperienced, the wilderness fighting of the French war having done nothing to teach him the strategy of manoeuvring whole armies. One of his chief faults was his tendency to subordinate his own judgment to that of the generals surrounding him; at every critical juncture, before Boston, before New York, before Philadelphia, in New Jersey, he called a council of war and in almost every instance accepted its decision. Naturally bold and dashing, as he proved at Trenton, Princeton, and Germantown, he repeatedly adopted evasive and delaying tactics on the advice of his associates; however, he did succeed in keeping a strong army in existence and maintaining the flame of national spirit; and when the auspicious moment arrived, he planned the rapid movements that ended the war.

Sources of Washington's military strength

One element of Washington's strength was his sternness as a disciplinarian. The army was continually dwindling and refilling; politics largely governed the selection of officers by Congress and the states; and the ill-fed, ill-clothed, ill-paid forces were often half-prostrated by sickness and ripe for mutiny. Troops from each of the three sections, New England, the middle states, and the South, showed a deplorable jealousy of the others. Washington was rigorous in breaking cowardly, inefficient, and dishonest men and boasted in front of Boston that he had "made a pretty good sort of slam among such kind of officers." Deserters and plunderers were flogged, and Washington once erected a gallows 40 feet high, writing that "I am determined if I can be justified in the proceeding, to hang two or three on it, as an example to others." At the same time, the commander in chief won the devotion of many of his men by his earnestness in demanding better treatment for them from Congress. He complained of their short rations, declaring once that they were forced to "eat every kind of horse food but hay."

Battle of New York

The darkest chapter in Washington's military leadership was opened when, reaching New York in April 1776, he placed half his army, about 9,000 men, under Israel Putnam, on the perilous position of Brooklyn Heights, Long Island, where a British fleet in the East River might cut off their retreat. He spent a fortnight in May with the Continental Congress in Philadelphia, then discussing the question of independence; and though no record of his utterances exists, there can be no doubt that he advocated complete separation. His return to New York preceded but slightly the arrival of the British army under Howe, which made its main encampment on Staten Island until its whole strength of nearly 30,000 could be mobilized. On August 22, 1776, Howe moved about 20,000 men across to Gravesend Bay on Long Island. Four days later, sending the fleet under command of his brother Adm. Richard Howe to make a feint against New York City, he thrust a crushing force along feebly protected roads against the American flank. The patriots were outmanoeuvred, defeated, and suffered a total loss of 5,000 men, of whom 2,000 were captured. Their whole position might have been carried by storm, but fortunately for Washington, General Howe delayed. While the enemy lingered, Washington succeeded under cover of a dense fog in ferrying the remaining force across the East River to Manhattan, where he took up a fortified position. The British, suddenly landing on the lower part of the island, drove back the Americans in a clash marked by disgraceful cowardice on the part of Connecticut and other troops. In a series of actions, Washington was forced northward, more than once in danger of capture, until the loss of his two Hudson River forts, one of them with 2,600 men, compelled him to retreat from White Plains across the river into New Jersey. He retired toward the Delaware while his army melted away, until it seemed that armed resistance to the British was about to expire.

Battles of Trenton and Princeton

It was at this darkest hour of the Revolution that Washington struck his brilliant blows at Trenton and Princeton, reviving the hopes and energies of the nation. Howe, believing the American army soon would dissolve totally, retired to New York, leaving strong forces in Trenton and Burlington. Washington, at his camp west of the Delaware, planned a simultaneous attack on both posts, using his whole command of 6,000 men. But his subordinates in charge of both wings failed him, and he was left on the night of December 25, 1776, to march on Trenton with about 2,400 men. He completely surprised the unprepared Hessians and after confused street fighting killed the commander, Johann Rall, and captured 1,000 prisoners with arms and ammunition. The immediate result was that General Cornwallis hastened with 8,000 men to Trenton, where he found Washington strongly posted behind the Assunpink Creek, skirmished with him, and decided to wait overnight "to bag the old fox."

During the night, the wind shifted, the roads froze hard, and Washington was able to steal away from camp, leaving his fires deceptively burning, march around Cornwallis' rear and fall at daybreak upon the three British regiments at Princeton. These were put to flight with a loss of 500 men, and Washington escaped with more captured munitions to a strong position at Morristown, New Jersey. The effect of these victories heartened all Americans, brought recruits flocking to camp in the spring, and encouraged foreign sympathizers with the American cause.

Thus far the important successes had been won by Washington; then they fell to others, while he was left to face popular apathy, military cabals, and the disaffection of Congress. The year 1777 was marked by the British capture of Philadelphia and the surrender of Gen. John Burgoyne's invading army to Gen. Horatio Gates at Saratoga followed by intrigues to displace Washington from his command. Howe's main British army of 18,000 left New York by sea on July 23, 1777, and landed on August 25 in Maryland, not far below Philadelphia. Washington, despite his inferiority of force, for he had only 11,000 men, mostly militia and in Lafayette's words "badly armed and worse clothed," risked a pitched battle on September 11 at the fords of Brandywine Creek, about 13 miles north of Wilmington. While part of the British force held the Americans engaged, Gen. Charles Cornwallis, with the rest, made a secret 17-mile detour and fell with crushing effect on the American right and rear, the result being a complete defeat from which Washington was fortunate to extricate his army in fairly good order. For a time he hoped to hold the Schuylkill Fords, but the British passed them and on September 26 triumphantly marched into Philadelphia. Congress fled to the interior of Pennsylvania, and Washington, after an unsuccessful effort to repeat his stroke at Trenton against the British troops posted at Germantown, had to take up winter quarters at Valley Forge. His army, twice beaten, ill housed, and ill fed, with thousands of men "barefoot and otherwise naked," was at the point of exhaustion; it could not keep the field, for inside of a month it would have disappeared. Under these circumstances, there is nothing that better proves the true fibre of Washington's character and the courage of his soul than the unyielding persistence with which he held his strong position at Valley Forge through a winter of semi-starvation, of justified grumbling by his men, of harsh public criticism, and of captious meddling by a Congress that was too weak to help him.

Valley Forge

Washington's enemies seized the moment of his greatest weakness to give vent to an antagonism that had been

Plot to displace Washington

nourished by sectional jealousies of north against south, by the ambition of small rivals, and by baseless accusations that he showed favouritism to such foreigners as Lafayette. The intrigues of Thomas Conway, an Irish adventurer who had served in the French army and had become American inspector general, enlisted Thomas Mifflin, Charles Lee, Benjamin Rush, and others in an attempt to displace Washington. General Gates appears to have been a tool of rather than a party to the plot, expecting that the chief command would devolve upon himself. A faction of Congress sympathized with the movement and attempted to paralyze Washington by reorganizing the board of war, a body vested with the general superintendence of operations, of which Gates became president; his chief of staff, James Wilkinson, the secretary; and Mifflin and Timothy Pickering, members. Washington was well aware of the hostility in congress, of the slanders spread by Benjamin Rush and James Lovell of Massachusetts, and of the effect of forgeries published in the American press by adroit British agents. He realized the intense jealousy of many New Englanders, which made even John Adams write his wife that he was thankful Burgoyne had not been captured by Washington, who would then "have been deified. It is bad enough as it is." But Washington decisively crushed the cabal when, the loose tongue of Wilkinson having disclosed Conway's treachery, he sent the latter officer on November 9, 1777, proof of his knowledge of the whole affair.

With the conclusion of the French alliance in the spring of 1778, the aspect of the war was radically altered; and the British army in Philadelphia, fearing that a French fleet would blockade the Delaware while the militia of New Jersey and Pennsylvania invested the city, hastily retreated upon New York City. Washington hoped to cut off part of the enemy and by a hurried march with six brigades interposed himself at the end of June between Sir Henry Clinton (who had succeeded Howe) and the Jersey coast. The result was the Battle of Monmouth on June 28, where a shrewd strategic plan and vigorous assault were brought to naught by the treachery of Charles Lee. When Lee ruined the attack by a sudden order to retreat, Washington hurried forward, fiercely denounced him, and restored the line, but the golden opportunity had been lost. The British made good their march to Sandy Hook, and Washington took up his quarters at New Brunswick. Lee was arrested, court-martialled, and convicted on all three of the charges made against him; but instead of being shot, as he deserved, he was sentenced to a suspension from command for one year. The arrival of the French fleet under Adm. Charles-Hector Estaing on July 1778 completed the isolation of the British, and Clinton was thenceforth held to New York City and the surrounding area. Washington made his headquarters in the highlands of the Hudson and distributed his troops in cantonments around the city and in New Jersey.

Yorktown and the end of the war

The final decisive stroke of the war, the capture of Cornwallis at Yorktown, is to be credited chiefly to Washington's vision. With the domestic situation intensely gloomy early in 1781, he was hampered by the feebleness of Congress, the popular discouragement, and the lack of prompt and strong support by the French fleet. A French army under Comte de Rochambeau having arrived to reinforce him in 1780, he pressed Admiral de Grasse to assist in an attack upon either Cornwallis in the south or Clinton in New York. In August the French admiral sent definite word that he preferred the Chesapeake, with its large area and deep water, as the scene of his operations; and within a week, on August 19, 1781, Washington marched south with his army, leaving Gen. William Heath with 4,000 men to hold West Point. He hurried his troops through New Jersey, embarked them on transports in Delaware Bay, and landed them at Williamsburg, Virginia, where he had arrived on September 14. Cornwallis had retreated to Yorktown and entrenched his army of 7,000 British regulars. Their works were completely invested before the end of the month; the siege was pressed with vigour by the allied armies under Washington, consisting of 5,500 Continentals, 3,500 Virginia militia, and 5,000 French regulars; and on October 19 Cornwallis surrendered. By this campaign, probably the finest single display of Washington's generalship, the war was brought to a virtual close.

Washington remained during the winter of 1781–82 with the Continental Congress in Philadelphia, exhorting it to maintain its exertions for liberty and to settle the army's claims for pay. He continued these exhortations after he joined his command at Newburgh on the Hudson in April 1782. He was astounded and angered when some loose camp suggestions found expression in a letter from Col. Lewis Nicola offering a plan by which he should use the army to make himself king. He blasted the proposal with fierce condemnation. When the discontent of his unpaid men came to a head in the circulation of the "Newburgh Address" early in 1783, he issued a general order censuring the paper and at a meeting of officers on March 15 read a speech admonishing the army to obey Congress and promising his best efforts for a redress of grievances. He was present at the entrance of the American army into New York on the day of Clinton's evacuation, November 25, 1783, and on December 4 took leave of his closest officers in an affecting scene at Fraunces' tavern. Travelling south, on December 23, in a solemn ceremonial immortalized by the pen of William Makepeace Thackeray, he resigned his commission to the Continental Congress in the state senate chamber of Maryland in Annapolis and received the thanks of the nation. His accounts of personal expenditures during his service, kept with minute exactness in his own handwriting and totalling £24,700, without charge for salary, had been given the controller of the treasury to be discharged. Washington left Annapolis at sunrise of December 24 and before nightfall was at home in Mount Vernon.

Return to Mount Vernon

In the next four years Washington found sufficient occupation in his estates, wishing to close his days as a gentleman-farmer and giving to agriculture as much energy and thought as to the army. He enlarged the Mount Vernon house; he laid out the grounds anew, with sunken walls, or ha-has; and he embarked on experiments with mahogany, palmetto, pepper, and other foreign trees, English grasses and grains. His farm manager during the Revolution, a distant relative named Lund Washington, retired in 1785 and was succeeded by a nephew, Maj. George Augustine Washington, who resided at Mount Vernon until his death, in 1792. Washington's losses during the war had been heavy, caused by neglect of his lands, stoppage of exportation, and a depreciation of paper money, which cost him hardly less than $30,000. He then attempted successfully to repair his fortunes, his annual receipts from all his estates being from $10,000 to $15,000 a year. In 1784 he made a tour of nearly 700 miles to view the wild lands he owned to the westward, Congress having made him a generous grant. As a national figure, he was constrained to offer hospitality to old army friends, visitors from other states and nations, diplomats, and Indian delegations, and he and his household seldom sat down to dinner alone.

PRESIDENCY

Postrevolutionary politics. Viewing the chaotic political condition of the United States after 1783 with frank pessimism and declaring (May 18, 1786) that "something must be done, or the fabric must fall, for it is certainly tottering," Washington repeatedly wrote his friends urging steps toward "an indissoluble union." At first he believed that the Articles of Confederation might be amended. Later, especially after the shock of Shays's rebellion, he took the view that a more radical reform was necessary but doubted as late as the end of 1786 that the time was ripe. His progress toward adoption of the idea of a federal convention was, in fact, puzzlingly slow. Though John Jay assured him in March 1786 that breakup of the nation seemed near and opinion for the convention was crystallizing, Washington remained noncommittal. But despite long hesitations, he earnestly supported the proposal for a federal impost, warning the states that their policy must decide "whether the Revolution must ultimately be considered a blessing or a curse." And his numerous letters to the leading men of the country assisted greatly to form a sentiment favourable to a more perfect union.

Some understanding being necessary between Virginia and Maryland regarding the navigation of the Potomac, commissioners from the two states met at Mount Vernon in the spring of 1785; from this seed sprang the federal convention. Washington approved in advance the call for a gathering of all the states to meet in Philadelphia in May 1787 to "render the Constitution of the Federal Government adequate to the exigencies of the Union." But he was again hesitant about attending, partly because he felt tired and infirm, partly because of doubts about the outcome. Although he hoped to the last to be excused, he was chosen one of Virginia's five delegates.

Washington arrived in Philadelphia on May 13, the day before the opening of the Convention, and as soon as a quorum was obtained he was unanimously chosen its president. For four months he presided over the Constitutional Convention, breaking his silence only once upon a minor question of congressional apportionment. Though he said little in debate, no one did more outside the hall to insist on stern measures. "My wish is," he wrote, "that the convention may adopt no temporizing expedients, but probe the defects of the Constitution to the bottom, and provide a radical cure." His weight of character did more than any other single force to bring the convention to an agreement and obtain ratification of the instrument afterward. He did not believe it perfect, though his precise criticisms of it are unknown. But his support gave it victory in Virginia, where he sent copies to Patrick Henry and other leaders with a hint that the alternative to adoption was anarchy, declaring that "it or dis-union is before us to chuse from," told powerfully in Massachusetts. He received and personally circulated copies of *The Federalist.* When once ratification was obtained, he wrote leaders in the various states urging that men staunchly favourable to it be elected to Congress. For a time he sincerely believed that, the new framework completed, he would be allowed to retire again to privacy. But all eyes immediately turned to him for the first president. He alone commanded the respect of both the parties engendered by the struggle over ratification, and he alone would be able to give prestige to the republic throughout Europe. In no state was any other name considered. The electors chosen in the first days of 1789 cast a unanimous vote for him, and reluctantly—for his love of peace, his distrust of his own abilities, and his fear that his motives in advocating the new government might be misconstrued all made him unwilling—he accepted.

Election as President of the United States

On April 16, after receiving congressional notification of the honour, he set out from Mount Vernon, reaching New York in time to be inaugurated on April 30. The ceremony was performed in Wall Street, near the spot now marked by Ward's statue of Washington; and a great crowd broke into cheers as, standing on the balcony of Federal Hall, he took the oath administered by Chancellor Robert Livingston and retired indoors to read Congress his inaugural address.

The Washington administration. Washington's administration of the government in the next eight years was marked by the caution, the methodical precision, and the sober judgment that had always characterized him. He regarded himself as standing aloof from party divisions and emphasized his position as president of the whole country by a tour first through the Northern states and later through the Southern. A painstaking inquiry into all the problems confronting the new nation laid the basis for a series of judicious recommendations to Congress in his first message. In selecting the four members of his first cabinet, Thomas Jefferson as secretary of state, Alexander Hamilton as secretary of treasury, Henry Knox as secretary of war, and Edmund Randolph as attorney general, Washington balanced the two parties evenly. But he leaned with especial weight upon Hamilton, supporting his scheme for the assumption of state debts, taking his view that the bill establishing the Bank of the United States was constitutional, and in general strengthening the authority of the federal government.

Distressed when the inevitable clash between Jefferson and Hamilton arose, he tried to keep harmony, writing frankly to each and refusing to accept their resignations. But when war was declared between France and England in 1793, he again took Hamilton's view that the United States should completely disregard the treaty of alliance with France and pursue a course of strict neutrality, while he acted decisively to stop the improper operations of the French minister, Edmund-Charles Genet. He had a firm belief that the United States must insist on its national identity, strength, and dignity. His object, he wrote, was to keep the country "free from political connections with *every* other country, to see them independent of all, and under the influence of none. In a word, I want an *American* character that the powers of Europe may be convinced that we act for *ourselves*, and not for others." The sequel was the resignation of Jefferson at the close of 1793, the two men parting on good terms and Washington praising Jefferson's "integrity and talents." The suppression of the Whisky Insurrection in 1794 by federal troops whom Hamilton led in person and the dispatch of John Jay to conclude a treaty of commerce with Great Britain tended further to align Washington with the Federalist Party. Though the general voice of the people compelled him to acquiesce reluctantly to a second term in 1792 and his election that year was again unanimous, during his last four years in office he suffered from a fierce personal and partisan animosity. This culminated when the publication of the terms of the Jay Treaty, which Washington signed on June 25, 1794, provoked a bitter discussion, and the house of representatives called upon the president for the instructions and correspondence relating to the treaty. These Washington, who had already clashed with the Senate on foreign affairs, refused to deliver; and in the face of an acrimonious debate, he firmly maintained his position.

Second term

Early in his first term, Washington, who by education and natural inclination was minutely careful of the proprieties of life, established the rules of a virtual republican court. In both New York and Philadelphia he rented the best houses procurable, refusing to accept the hospitality of George Clinton, for he believed the head of the nation should be no man's guest. He returned no calls and shook hands with no one, acknowledging salutations by a formal bow. He drove in a coach drawn by four or six smart horses, with outriders and lackeys in rich livery. He attended receptions dressed in a black velvet suit with gold buckles, with yellow gloves, powdered hair, a cocked hat with an ostrich plume in one hand and a sword in a white leather scabbard. After being overwhelmed by callers, he announced that except for a weekly levee open to all, persons desiring to see him must make previous engagements. On Friday afternoons the First Lady held informal receptions, at which the President appeared. Though the presidents of the Continental Congress had made their tables partly public, Washington, who entertained largely, inviting members of Congress in rotation, insisted that his hospitality be private. He served good wines and the menus were elaborate, but such visitors as Senator Maclay complained that the atmosphere was too "solemn." Indeed, his simple ceremony offended many of the more radical anti-Federalists, who did not share his sense of its fitness and accused the president of conducting himself as a king. But his cold and reserved manner was caused by native diffidence rather than any excessive sense of dignity.

Retirement. Earnestly desiring leisure, feeling a decline of his physical powers, and wincing under opposition abuse, Washington refused to yield to the general pressure for a third term. This refusal was blended with a testament of sagacious advice to his country in the *Farewell Address* of September 19, 1796, written largely by Hamilton but remolded by Washington and expressing his ideas. Retiring in March 1797 to Mount Vernon, he devoted himself for the last two and a half years of his life to his family, farm operations, and care of his slaves. In 1798 his seclusion was briefly interrupted when the prospect of war with France caused his appointment as commander in chief of the provisional army, and he was much worried by the political quarrels over high commissions; but the war cloud passed away. On December 12, 1799, he exposed himself on horseback for several hours to cold and snow and, returning home exhausted, was attacked late next day with quinsy or acute laryngitis. He was bled heavily four

times and given gargles of "molasses, vinegar and butter," and a blister of cantharides (a preparation of dried beetles) was placed on his throat, his strength meanwhile rapidly sinking. He faced the end with characteristic serenity, saying, "I die hard, but I am not afraid to go," and later: "I feel myself going. I thank you for your attentions; but I pray you to take no more trouble about me. Let me go off quietly. I cannot last long." After giving instructions to his secretary, Tobias Lear, about his burial, he died at 10:00 PM on December 14. The news of his death placed the entire country in mourning, and the sentiment of the country endorsed the famous words of Henry Lee, embodied in resolutions which John Marshall introduced in the house of representatives, that he was "first in war, first in peace, and first in the hearts of his countrymen." When the news reached Europe, the British channel fleet and the armies of Napoleon paid tribute to his memory, and many of the leaders of the time joined in according him a preeminent place among the heroes of history. (Al.N.)

BIBLIOGRAPHY. The earliest known portrait of Washington is that by Charles Willson Peale, painted in 1772. A long line of painters and sculptors followed, and their work may be found criticized in JUSTIN WINSOR (ed.), *Narrative and Critical History of America,* vol. 7 (1888, reprinted 1967). Washington himself thought highly of the likeness by Joseph Wright, painted in 1782. According to Winsor, the favourite profile is Houdon's, while Gilbert Stuart's canvas had been popularly preferred for the full face and John Trumbull's florid paintings for the whole figure. Stuart's pictures are somewhat idealized, while all the later portraits suffer from the fact that the artificial teeth worn by Washington in later years altered the expression of his face. Houdon's statue hardly does justice to Washington's imposing stature, dignified carriage, and great poise. But the Houdon bust, modelled from life, is excellent. For the iconography, see GUSTAV EISEN, *Portraits of Washington,* 3 vol. (1932).

A full compilation of Washington's *Writings* is the United States bicentennial commission edition, ed. by JOHN C. FITZPATRICK, 39 vol. (1931–44). This displaces WORTHINGTON C. FORD, *The Writings of George Washington,* 14 vol. (1889–93); and the edition by JARED SPARKS in 12 vol. (1834–37). The *Diaries* were separately edited by JOHN C. FITZPATRICK in 4 vol. (1925), and later by DONALD JACKSON and DOROTHY TWOHIG in 6 vol. (1976–80). *The Journal of the Proceedings of the President: 1793–97,* ed. by DOROTHY TWOHIG (1981), is an annotated executive daybook covering years for most of which the diaries have not been found. Much the fullest and best biography is that by DOUGLAS SOUTHALL FREEMAN, *George Washington,* 6 vol. (1948–54), which is supplemented by a seventh volume, *First in Peace* (1957), written by his associates JOHN A. CARROLL and MARY W. ASHWORTH. Among good earlier biographies now corrected by Freeman are those by JOHN MARSHALL, *The Life of George Washington* (including plagiarized material) in 5 vol. (1804–07, reprinted 1969); WASHINGTON IRVING, *Life of Washington,* also in 5 vol. (1855–59; reissued in 3 vol., 1982); EDWARD EVERETT HALE, *The Life of George Washington* (1888); HENRY CABOT LODGE, *George Washington,* 2 vol. (1889, reprinted 1972); WOODROW WILSON, *George Washington* (1896, reprinted 1969); WORTHINGTON C. FORD, *George Washington* (1899); and NORMAN HAPGOOD, *George Washington* (1901). A new edition of the early 19th-century biography by MASON L. WEEMS, *The Life of Washington,* was issued in 1977. RUPERT HUGHES published an incomplete three-volume biography in 1926–29. Among the best modern works are JOHN C. FITZPATRICK, *George Washington Himself* (1933, reprinted 1975); SHELBY LITTLE, *George Washington* (1929); MARCUS CUNLIFFE, *George Washington: Man and Monument,* rev. ed. (1982); and JAMES T. FLEXNER, *George Washington: The Forge of Experience, 1732–1775* (1965), *George Washington in the American Revolution, 1775–1783* (1968), *George Washington and the New Nation, 1783–1793* (1970), *George Washington: Anguish and Farewell, 1793–1799* (1972), and *George Washington: The Indispensable Man* (1975). CHARLES C. WALL, *George Washington* (1980), emphasizes the personal life.

Valuable studies on special aspects of Washington's career include: CHARLES H. AMBLER, *George Washington and the West* (1936, reprinted 1971); HUGH CLELAND, *George Washington in the Ohio Valley* (1955); JOHN CORBIN, *The Unknown Washington* (1930, reprinted 1972), largely on his constitutional ideas; PAUL LELAND HAWORTH, *George Washington, Country Gentleman: Being an Account of His Home Life and Agricultural Activities* (1925); EUGENE E. PRUSSING, *The Estate of George Washington, Deceased* (1927); BURKE DAVIS, *George Washington and the American Revolution* (1975), an account of his role as military commander; FORREST MCDONALD, *The Presidency of George Washington* (1974), a study of the politics and economics; THOMAS G. FROTHINGHAM, *Washington, Commander in Chief* (1930); HALSTED L. RITTER, *Washington as a Business Man* (1931); CHARLES W. STETSON, *Washington and His Neighbors* (1956); and GEORGE A. BILLIAS (ed.), *George Washington's Generals* (1964, reprinted 1980). PAUL LEICESTER FORD, *The True George Washington* (1896, reprinted 1971; republished as *George Washington,* 1970), is a classic examination of all sides of Washington's career and personality; while CHARLES MOORE has examined *The Family Life of George Washington* (1926). An excellent work on Martha Washington is ANNE HOLLINGSWORTH WHARTON, *Martha Washington* (1897, reprinted 1967), supplemented by PAUL WILSTACH, *Mount Vernon: Washington's Home and the Nation's Shrine* (1916).

(Al.N./Ed.)

Washington, D.C.

The capital of the United States of America, the city of Washington is coextensive with the District of Columbia, whose site was agreed upon by Congress in 1790 as the permanent seat of government for the new nation. It is located at the head of navigation of the Potomac River, which separates it from Virginia to the southwest, while its other boundaries make it essentially a 68-square-mile (177-square-kilometre) enclave carved from Maryland. The Washington Standard Metropolitan Statistical Area (SMSA), however, encompasses 2,809 square miles, including two counties in Maryland and the cities of Fairfax, Falls Church, and Alexandria and four counties in Virginia.

This article is divided into the following sections:

Physical and human geography 707
- Character of the city 707
  - Impediments to municipal development
  - The incomplete city
- The landscape 707
  - The city layout
  - Modern features
- The people 712
- The economy 713
  - Government
  - Tourism
  - Transportation
- Administrative and social conditions 713
  - Government
  - Public services
  - Employment
  - Housing
  - Education
- Cultural life 715
  - The arts
  - National collections
  - Festivals
  - Recreation

History 716
- The early period 716
  - Conception, siting, and design
  - The young city
- Evolution of the modern city 717

Bibliography 717

Physical and human geography

CHARACTER OF THE CITY

Washington is one of the few capital cities of the world founded expressly as a seat of government and as a centre for international representation. The expansive designs for the city were to symbolize the ideals of the freedom so recently achieved yet still so tenuously held by the citizenry of the nation. It was to be a vital city, the proper seat for the federal government.

The modern city also holds the nation's most sacred monuments and the most meaningful artifacts of its history, the embassies of foreign nations, and an impressive collection of the national art treasures. Nearly every significant national organization has its headquarters or a major branch in the District, often for the purpose of lobbying in Congress or within federal agencies. The city was meant to be, and has remained, the focal point of the nation for sightseers and for seekers after the spirit of the American past and present.

Impediments to municipal development. For many reasons, Washington has had a development that is unique among the world's major cities. An inherent tendency to distinguish between the District of Columbia as the capital, with its paraphernalia of federal facilities, and Washington as the city, with its complexity of social, economic, and political problems, has produced a somewhat schizophrenic metropolis as notorious for its ugliness and crime as it is famous for its diverse and truly awesome beauties. The city is located near the end of the sprawling urbanized agglomeration that spreads southward from Boston along the Eastern Seaboard, and it has most of the same problems that face other great core metropolises of the region—Boston, New York City, Philadelphia, and Baltimore—and of the nation as a whole.

Another set of disparities exists between the city of Washington and the Washington metropolitan area. More than one-fourth of the people over 26 years of age in the metropolitan area hold college degrees, the highest percentage among the 10 largest such areas in the country, and the area's population has one of the highest annual per capita incomes. On the other hand, a high proportion of the population within the city is made up of families with low incomes or of handicapped or aged persons, many requiring governmental assistance.

These factors provide insight into why Washington—the city as distinct from the capital and the metropolitan area—has developed neither the social stability nor the continuity that provides the lifeblood of most other large cities. Both the living and working populations of Washington are among the most transient in the nation. Only a small percentage of residents have longtime roots in Washington, while a high proportion of government and service workers commute into the city from suburban homes. This more affluent Washington worker typically spends most of his income and pays most of his taxes in adjacent counties and states, leaving the needs and destiny of the city to the dictates of an essentially alien body of lawmakers and administrators.

The incomplete city. With its history and character so profoundly affected by the varying and nonlocal interests and values of Congress, however, concern for Washington's social and environmental quality has been limited. The prominence of the park and highway systems in the city's landscape reflects the interests of Congress and various federal agencies, but the historic lack of concern for the city as a whole has left the grandeur of its vistas, monuments, and governmental facilities in startling contrast to the large areas of the city that have struck many observers as being in advanced stages of physical and spiritual decay.

This may be epitomized in the contrasts of weekday and weekend Washington. Perhaps no other American city except Los Angeles depends so heavily upon private automobiles for transportation. During rush hours, cars and buses choke the streets and avenues, a twice-daily scene interspersed with weekends of deserted streets and empty lots. The lonely grandeur of acres of empty, gleaming public buildings and their attending fountains and statues is a depressingly typical Sunday scene.

The many factors producing this periodic entrance and exit of citizenry may vary little from those in other major cities. The emotional impact in Washington, however, is especially poignant in view of the very different vision of the city's original planners.

THE LANDSCAPE

The city layout. Two elements of the original plan for Washington, completed in 1791 by the French engineer Pierre-Charles L'Enfant, are of special significance. First, his ideas for a capital city were based not on 13 colonies with 3,000,000 inhabitants but on a republic ultimately having 50 states and 500,000,000 citizens. He envisioned a city of 800,000 inhabitants at a time when the entire United States had less than four times that number. Second, L'Enfant had reached artistic maturity in Paris and Versailles, where he was influenced strongly by Baroque landscape architecture, which was at its zenith in the

Concepts underlying L'Enfant's designs

Capitol Hill, with the Capitol in the foreground. The Senate occupies the left wing and the House of Representatives the right. The Supreme Court is at upper left and buildings of the Library of Congress at upper right.

H. Armstrong Roberts

late decades of the 18th century. His original designs for Washington reflected the grandeur both of his projections and of the Baroque.

L'Enfant's basic plan. Perhaps the dominant element in L'Enfant's designs is the complex revolving about the Capitol, the Mall, and the executive mansion, which came to be known as the White House. Both buildings, incorporating elements of Baroque design, were placed to form the background, or terminating vista, of long straight pathways, or malls. Radiating from the buildings were two series of broad avenues converging into circular intersections, the effect of which was to create, in L'Enfant's phrase, "a reciprocity of view," a means of terminating long vistas that would give direction and character to the city and would create throughout it a series of subcentres within view of one another. Most of these subcentres—now circles and squares with small green parks—were carefully located on natural rises in the terrain, as were the Capitol and the White House.

The Mall, which extends from the Capitol to the Lincoln Memorial, was intended to be one broad, tree-lined avenue, in the manner of the Champs-Élysées in Paris, rather than the green lawn, with occasional buildings and a crisscross of roads, that it has become. At the point where the north–south axis of the White House meets the east–west axis of the Capitol, an equestrian statue of George Washington was planned. This spot, slightly relocated because of soft subsoil and subterranean streams, is now the site of the Washington Monument.

Avenues, streets, and quadrants

The pattern of radiating avenues was to be joined and filled by a gridiron matrix of streets. With the Capitol as the axis, the streets were lettered to the north and south, numbered to the east and west, and the avenues were named for the states. The entire city was divided into northwest (NW), northeast (NE), southeast (SE), and southwest (SW) quadrants, also centred on the Capitol. L'Enfant's plan ended to the north at what became Florida Avenue, where a steep bluff was to provide the approaching traveller with the impressive expanse of the city spread out at his feet.

Subsequent development. Even before the Capitol's cornerstone was laid, L'Enfant had been dismissed reluctantly after disagreements with the commissioners in charge of raising the public buildings. During the 19th century and later, the original harmony of his design was eroded with the emergence of new municipal and governmental functions. Awkward relations developed between the grid pattern of cross streets and the diagonal avenues, especially at intersections, while the circular intersections, though providing pleasant oases for pedestrians, were fed by too many streets in the unforeseen age of the automobile and consequently became impediments along the city's thoroughfares. The original plan was effectively lost until 1887, when its recovery revealed that a number of 19th-century buildings along Pennsylvania Avenue, around the White House, and on Capitol Hill had altered the symmetries of the original conception.

The addition of trees along most major avenues and streets, though now a feature of the city, destroyed in great part the concept of a city of magnificent distances. The location of other notable monuments, especially the Lincoln and Jefferson memorials and the John F. Kennedy Center for the Performing Arts, modified the original plan further, as did the placement of various museums and temporary buildings on the Mall. Whatever the impact of later alterations, however, only a few efforts have been made since L'Enfant to redesign the capital totally.

Continuing affective elements of the capital. To the visitor, nothing is more expressive of Washington's character than the dozens of buildings and hundreds of monuments that relate to its functions as the nation's capital. The style of federal architecture has changed several times over the years, and L'Enfant's plans for the location of the Capitol and the White House remain as focal points for a complexity of monuments and buildings that retain a kind of basic coherence to the image of Washington as a capital city.

The Capitol, the White House, and adjacent buildings

Dominating the scene, the Capitol offers an impressive silhouette terminating the Mall. Criticized by many as a "potpourri of many chefs," the Capitol may lack the authenticity of the massive Renaissance domes that inspired it, but to many others it offers a feeling of emotional assurance that representative government rests on a broad and solid foundation. The other original building that

The South Portico of the White House.
Authenticated News International

underscores the purposes of L'Enfant is the charming, essentially modest White House, which has been judged by many as among the finest and most appropriate residences of state in the world.

The largest complex of public buildings that joins the Capitol to the White House is the Federal Triangle, an imposing facade of buildings that fronts on Constitution Avenue along the Mall and houses various federal departments and other activities of national importance. Other structures located along the Mall include the National Gallery of Art, the National Archives, the National Museum of Natural History, the National Museum of American History, and the National Air and Space Museum. Around the Capitol on Capitol Hill are the buildings of the Supreme Court, Library of Congress, and the various House and Senate office buildings.

The increasing number of governmental departments and agencies has resulted in an explosion of construction, particularly along Independence Avenue and extending into southwestern Washington's Foggy Bottom area. These buildings have been designed without particular regard for their individuality: a series of shafts or cubes faced with granite or marble. To many people they seem to represent, in contrast to the earlier forms of governmental architecture, an honest but ugly reminder of the anonymous bureaucratic routines upon which so much of modern organizational life depends.

Statuary and memorial structures

Ever since L'Enfant urged that Washington be lavishly equipped with "statues, columns or other ornaments" to honour the nation's great, a continuous effort has been made to assure that no open space in the city lacks its representative monument. Within the District of Columbia alone, more than 300 memorials and statues of varying size, purpose, and aesthetic merit have been raised—from the elegant and inspiring Lincoln and Jefferson memorials and the pure grandeur of the Washington Monument, to the statues of major and minor Civil War heroes, to memorial benches, Doric temples, and Japanese pagodas.

A growing concern that Washington would soon become a veritable quarry of stone and metal monuments dedicated to minor individuals and causes led to a movement toward "living" or "functional" memorials that attempt to translate the personality of the person being honoured. Within this context, the monument dedicated to Pres. John F. Kennedy was designed as a cultural centre for the performing arts. In addition, landscaping changes in the Mall area have removed most of the temporary buildings and returned the area to a purer interpretation of the L'Enfant design.

Attempts to bring Washington's waterways into an overall national capital plan have been relative failures. Important as were the Potomac and Anacostia rivers in the decision about the capital's location, neither has lived up to its anticipated potentiality in the commercial or cultural life of the city. Technology rendered them obsolete for transportation, while pollution limited their usefulness for recreation. Some modest progress, however, has been made to enhance their aesthetic and recreational possibilities.

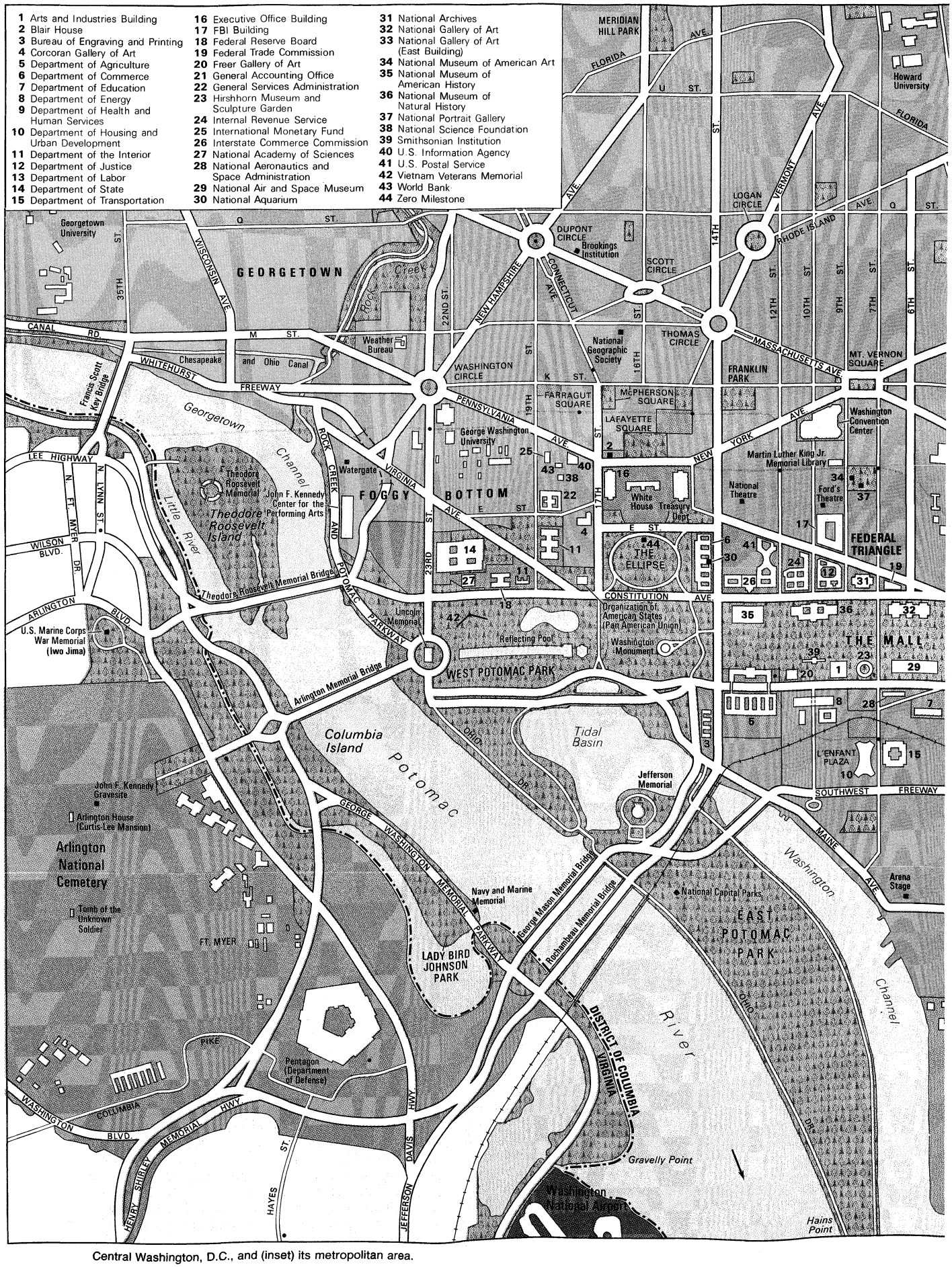
Modern features. Washington as a city has expanded its influence and functions well beyond the 1790 boundaries established for the District of Columbia. Like the other large cities and metropolitan areas of the East Coast, it is beset with extreme social and economic problems. Unlike others, however, its surrounding suburban sprawl is part of other sovereign states, and the District can exercise little political control beyond its borders.

The strain of this growing area, which in large part reflects the increasing size and influence of the federal government, is playing an enormous part in the changing physical character of the central city. Plans to further develop Washington include greater dispersal of federal activities and installations to suburban locations to alleviate transportation problems. Such a movement has been in progress for some time, along with the decentralization of many business and commercial functions. The result of these movements has been to concentrate within the physical boundaries of the city only those monumental, cultural, and governmental structures that logically and historically can be located nowhere else.

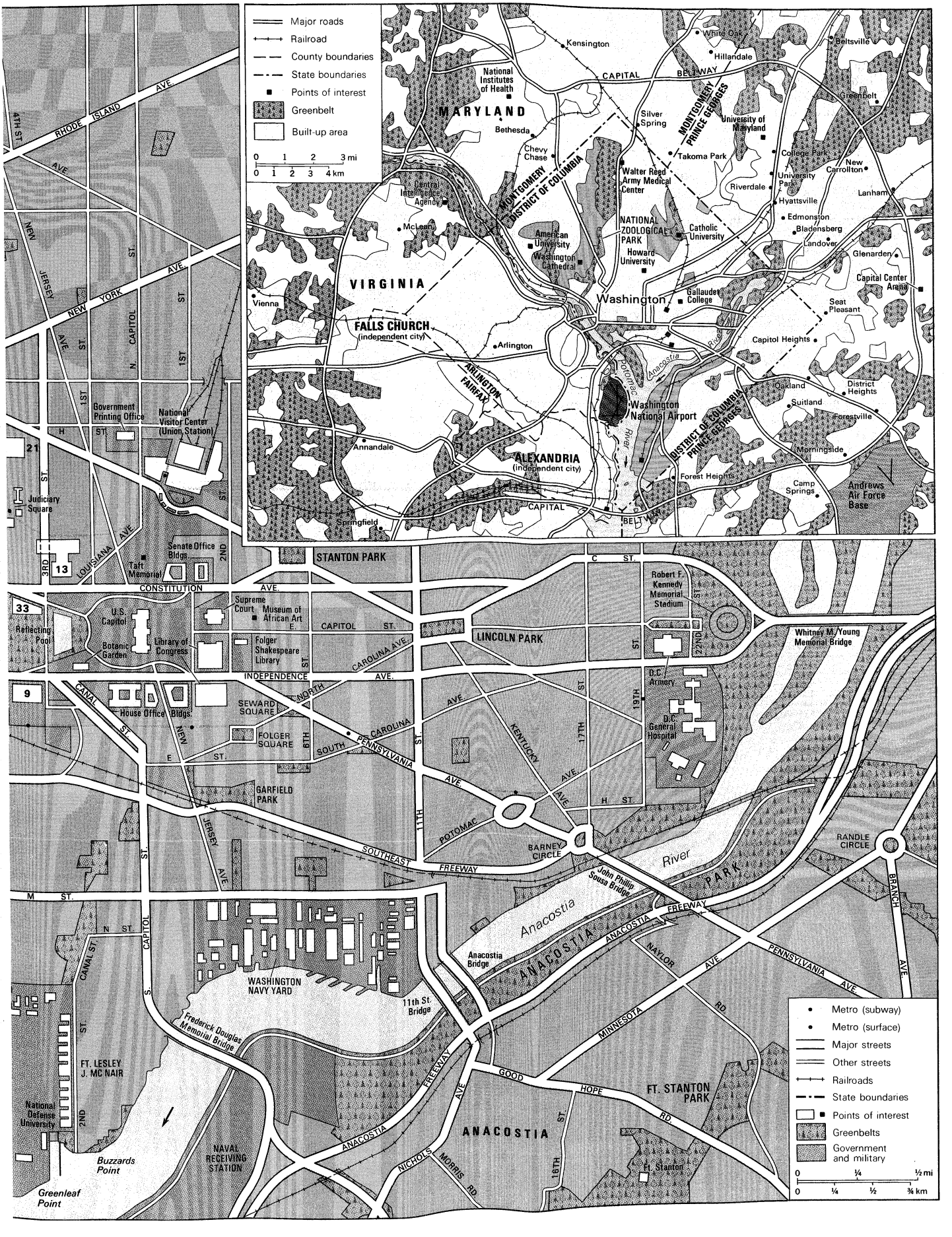
Changes in the face of Washington

The trends in planning that are changing the physical character of contemporary Washington are of two basic types. First, there has been an increasing effort to coordinate and enhance those monumental aspects of the city that are the outstanding physical features and most meaningful to the nation itself—notably in the Mall and Capitol areas. With this trend is the design of new public buildings to operate efficiently and yet blend harmoniously with the older federal architecture. Examples of such newer structures include the East Building of the National Gallery of Art, which employs the so-called L'Enfant Angle in its architectural design, and the congressional office buildings on Capitol Hill.

Second, there have been efforts to make downtown com-



Central Washington, D.C., and (inset) its metropolitan area.


Major roads
Railroad
County boundaries
State boundaries
Points of interest
Greenbelt
Built-up area
0 1 2 3 mi
0 1 2 3 4 km
MARYLAND
VIRGINIA
Washington
Kensington
White Oak
Hillandale
Beltsville
National Institutes of Health
CAPITAL BELTWAY
Bethesda
Silver Spring
MONTGOMERY
PRINCE GEORGES
University of Maryland
Greenbelt
Chevy Chase
Takoma Park
College Park
New Carrollton
University Park
Walter Reed Army Medical Center
Riverdale
Hyattsville
Lanham
Central Intelligence Agency
MONTGOMERY
DISTRICT OF COLUMBIA
Edmonston
NATIONAL ZOOLOGICAL PARK
Catholic University
Bladensberg
McLean
American University
Howard University
Washington Cathedral
Landover
Glenarden
Capital Center Arena
Vienna
Gallaudet College
Seat Pleasant
FALLS CHURCH
(independent city)
Arlington
Capitol Heights
Potomac
Anacostia
Park
ARLINGTON
FAIRFAX
Oakland
District Heights
Suitland
Washington National Airport
Forestville
DISTRICT OF COLUMBIA
PRINCE GEORGES
Annandale
ALEXANDRIA
(independent city)
Morningside
Forest Heights
Camp Springs
Andrews Air Force Base
CAPITAL BELTWAY
Springfield
Government Printing Office
National Visitor Center (Union Station)
Judiciary Square
Senate Office Bldgs.
Taft Memorial
CONSTITUTION AVE.
STANTON PARK
C ST.
Robert F. Kennedy Memorial Stadium
U.S. Capitol
Reflecting Pool
Supreme Court
Museum of African Art
Botanic Garden
Library of Congress
Folger Shakespeare Library
CAPITOL ST.
LINCOLN PARK
Whitney M. Young Memorial Bridge
INDEPENDENCE AVE.
CAROLINA AVE.
D.C. Armory
House Office Bldgs.
SEWARD SQUARE
FOLGER SQUARE
SOUTH CAROLINA AVE.
PENNSYLVANIA AVE.
KENTUCKY AVE.
D.C. General Hospital
GARFIELD PARK
POTOMAC
BARNEY CIRCLE
SOUTHEAST FREEWAY
John Philip Sousa Bridge
RANDLE CIRCLE
PARK
River
M ST.
Anacostia
ANACOSTIA FREEWAY
Anacostia Bridge
NAYLOR
PENNSYLVANIA AVE.
BRANCH
WASHINGTON NAVY YARD
11th St. Bridge
Frederick Douglass Memorial Bridge
MINNESOTA AVE.
GOOD HOPE RD
FT. LESLEY J. MC NAIR
FT. STANTON PARK
National Defense University
ANACOSTIA
NAVAL RECEIVING STATION
Buzzards Point
Greenleaf Point
NICHOLS
MORRIS RD
Ft. Stanton
Metro (subway)
Metro (surface)
Major streets
Other streets
Railroads
State boundaries
Points of interest
Greenbelts
Government and military
0 ¼ ½ mi
0 ¼ ½ ¾ km

mercial activities more attractive, as well as more accessible. Several pedestrian malls have been constructed in the major downtown shopping area, and a new business and commercial complex, L'Enfant Plaza, has been established within walking distance of a number of the newest public buildings. A large convention centre was built, and a subway system, called the Metro, was inaugurated.

Other new features since the city began more rapid changes during the 1970s include the renovation of Pennsylvania Avenue, which was designated as a historic site in 1965; the renovation of Union Station as the National Visitor Center; and the redesigning of Lafayette Square facing the White House, a move that was hotly debated because it involved the razing of a number of fine old buildings. The Vietnam Veterans Memorial, in West Potomac Park, was dedicated on November 11, 1982.

THE PEOPLE

A large majority of the city's population is black; the large increase among blacks since the 1960s is attributed to both natural increase and immigration. The suburban areas, however, remain predominantly white.

In terms of white ethnic composition, Washingtonians tend to reflect the nation as a whole. The city was never a magnet like Boston, New York City, or Chicago for large groups of immigrants from Europe. If any section of the country is overrepresented among Washingtonians, it is the Southeast and the nearby border states. Life in the city often has been described as more Southern and leisurely than Northern and fast-paced—a comparison that sometimes leads to conjectures about how the choice of Washington's site may have affected the functions of the national government throughout its history.

Changes in national administration bring about few real changes in the city's makeup. The personnel of foreign embassies and missions living temporarily or semipermanently in Washington are large in number but relatively small in the total population, and their participation in the city's life generally is fairly circumscribed. If the city has any true social aristocracy (aside from the upper echelons of the administration in power, members of Congress, and other high-ranking government officials), it is composed of a small cadre of well-to-do individuals or families who once held some office, found the capital's life congenial, and remained as permanent residents.

Partially because of its racial and socioeconomic makeup, the District of Columbia and its surrounding metropolitan counties present a distinctive pattern of population distribution and thus a variability in the city's neighbourhood characteristics.

Generally, within the northwestern area of the city and the suburban triangle of Maryland—including such unincorporated communities as Chevy Chase, Bethesda, and Silver Spring—that flanks this part of the city are located the majority of white middle-class and upper middle-class residents. In the west, overlooking the Potomac, is the oldest neighbourhood, Georgetown, which remains Washington's most exclusive and picturesque residential area. Within these neighbourhoods, extending eastward in the District to Connecticut Avenue and 16th Street, are located nearly all of the foreign embassies, the great churches, private country clubs, and housing that ranges from the moderately affluent to the luxurious.

East of 16th Street and extending north into the Maryland suburbs is the heart of the Washington black community. The area illustrates the considerable diversity in quality of living available to Washington's black residents: from tree-shaded, upper middle-class streets in the vicinity of Howard University to rows of deteriorating houses and tenements. Along 14th Street are the principal areas of vice and crime in the city and the scene of the "April Riots" of 1968.

The southern sections of the city form rather distinct areas, determined partly by geographical features and partly by the racial and socioeconomic character of the residents. The first area comprises the southwestern bank of the Potomac, which is crossed by six bridges joining the District and Virginia. In general this area is nonresidential, but it includes some of the main features of the city's life—the Pentagon, Washington National Airport, Arlington National Cemetery, and George Washington Memorial Parkway. The second area is an irregular triangle formed by the Potomac and Anacostia rivers, which converge at the apex of the triangle. Within this section are located a concentration of modern governmental buildings—Ft. Lesley J. McNair and the National Defense University—and the large recreational complex of East Potomac Park (known popularly as Hains Point). Urban renewal projects in this section have changed the residency pattern from predominantly black slums to white upper middle-class apartment and townhouse buildings. Many of the black residents displaced by the southwestern renewal plans

Ward Howe-FPG

Georgetown, a residential section of Northwest Washington, whose 18th- and early 19th-century architectural design has been preserved by Congressional legislation.

have shifted to already overcrowded northeastern parts of the District or have concentrated in a third major black population area, on the southeastern bank of the Anacostia River. This triangular-shaped, low-income area makes up the southeastern edge of the District and extends into Prince Georges County, Maryland.

THE ECONOMY

Few cities in the United States are so dominated by the nature of their economic base as is Washington. Only two major economic activities provide virtually all of the income to the city and its residents. The federal civil service is by far the largest single employer in the metropolitan area. Tourism, which includes its retail trade and related services, is second in economic importance. Manufacturing and other commercial activities occupy only a minor place in the economic structure.

Economic dominance of government and tourism

Government. The federal government's dominance in establishing the character of economic and social life in Washington can scarcely be overestimated. Although the federal government itself employs a large number of residents of the metropolitan area, the vast majority of employed people are engaged in activities that support or depend upon governmental programs or organizations of national or international scope.

Since the mid-20th century Washington has been transformed from a "federal town" to an information and communication centre that is competitive with the largest urban areas in the United States. In addition, it has become the coordinating centre for major foreign activities. As federal involvement has proliferated in both private and public sectors and both at home and abroad, there has been a steady trend toward relocation of the offices of national associations from cities such as New York and Chicago to the Washington metropolitan area. There are more headquarters of national trade and professional organizations and associations in Washington (about 300) than in any other area of the country. The presence of international organizations such as the World Bank and the International Monetary Fund has increased Washington's importance as one of the principal centres for coordinating aid, trade, and finance on the international level.

The District of Columbia, however, has a large population living at considerably lower socioeconomic levels. These residents, mostly black, also depend upon the opportunities offered for federal employment but generally lack the educational qualifications for jobs at higher incomes. This factor, plus the minimal local control over its revenue system, the lack of significant taxable industry, and the tax-exempt status of the great majority of its real estate, forces the city to depend heavily upon intergovernmental revenue and aids that, in effect, constitute payment by the federal government in lieu of the taxes it would pay if it were a private industry.

Tourism. Because it is the seat of the national government and the site of many of the nation's most significant monuments, Washington's second largest source of income derives from the millions of tourists who visit the city each year. The peak of the tourist influx occurs in the spring and summer months, but the capital city is basically a year-round attraction. In addition, Washington has become increasingly popular as a site for the annual conventions of national organizations and professional associations, most of which, as noted above, maintain a headquarters or branch office in the city. The business community, through such groups as the Metropolitan Washington Board of Trade and with the assistance of the federal government, has organized an increasing number of activities and facilities to make Washington more attractive to its visitors.

Transportation. Prior to the 1970s, various characteristics of the Washington area affected the nature and quality of its transportation networks. First, the L'Enfant plan covered only a small part of the present District of Columbia, excluding completely what has become the suburban area. Although many major radial avenues were extended to and beyond the District's borders, the street patterns outside the original city of 1791 were irregular at best. Second, Washington was a funnel for north–south traffic along the East Coast until the mid-20th century. In addition, the Potomac and Anacostia rivers always formed barriers to traffic flow between the city and points south. Finally, although it has long been a city of white-collar commuters whose work draws them into the city, the District had barely minimal mass-transit facilities from suburban areas, necessitating nearly total reliance on the automobile.

Since the 1970s modern beltways and freeways have been built to facilitate the flow of traffic around and into the city, and a number of bridges now span the Potomac and Anacostia rivers. A major stimulus to the city's development was the opening of a subway system. The subway and extended rail and bus service are part of the city's modern mass-transit system, operated by the Washington Metropolitan Area Transit Authority. The authority also serves the neighbouring suburbs in Maryland and Virginia.

ADMINISTRATIVE AND SOCIAL CONDITIONS

Government. The city of Washington, as the site of the nation's capital, has evolved a governmental structure that is unique among U.S. cities.

Changing municipal forms. The first government of the city of Washington, established in 1802, comprised a mayor appointed by the president of the United States and a city council elected by the people. The city's character was amended in 1812 to provide for an elected board of aldermen, which, along with the council, elected the mayor. In 1820 Congress permitted the residents to elect both mayor and council. Since Article 1 of the U.S. Constitution empowers Congress to exercise exclusive legislation over the seat of government, however, the powers of the mayor and the council were limited, and their administration of the city was generally ineffectual.

Evolution of District government

In 1871 Congress created a territorial form of government for the District. The officials, all appointed by the president, included a governor, a board of public works, and a legislative assembly comprising an 11-member Council and a 22-member House of Delegates. In addition, the District was permitted a popularly elected, nonvoting delegate to the House of Representatives. This arrangement was abandoned after only three years following a series of financial crises that aroused opposition among both politicians and taxpayers. Congress resumed direct control of the city, providing administration by three commissioners appointed by the president. No provision was made for the franchise under the commissioner form of government, and residents of the District were denied all rights to vote until 1961. The 23rd Amendment to the Constitution then allowed qualified voters to vote in presidential elections but failed to permit participation in elections for local officials, all of whom continued to be appointed.

The issue of home rule for the residents of the District became increasingly prominent in the 1960s, and it was not unrelated to the general struggle for civil rights that characterized the nation as a whole. The most serious criticism of the commissioner form of government was that all legislation affecting it had to be passed by Congress: the House District of Columbia Committee and the Senate Governmental Affairs Committee were required to initiate all legislation pertaining to the District. Since the members of these committees were not permanent residents of Washington and represented constituencies that had little or no interest in the problems of the city, the responsiveness of Congress was felt by many to be slow or entirely lacking. Efforts on the part of various local groups over the years to achieve some degree of home rule were consistently blocked by the House committee, although the Senate committee passed five such bills between 1951 and 1963. It was often pointed out that the committees tended to be dominated by Southern congressmen who resisted efforts to give the franchise and other powers to the District because of its increasing black majority.

In 1967 Congress reorganized the District's government, abolishing the three-commissioner system and creating in its place a single commissioner (who assumed the title of mayor), an assistant commissioner, and a nine-member city council, all appointed by the president. The city council was given authority to exercise certain legislative and

regulatory powers once vested in the three commissioners, but such actions were subject to the veto of the mayor. In 1970 Congress created again the position of a nonvoting delegate to the House of Representatives, elected by residents of the District.

Movement toward home rule has continued. In 1973, under President Nixon, the limited Home Rule Act of 1964 was amended, providing for the popular election every four years of the mayor and city council members. In addition, the city council was expanded to 13 members. The mayor was given broader reorganizational and appointive authority. The city council was empowered to establish and set tax rates and fees, make changes in the budget, and organize or abolish any agency of government of the District. Congress, in turn, reserved the power to veto any actions of the District government that threaten the "federal interest." Thus, while the District has a recognizable municipal form of government, Congress continues to treat it in some respects as a branch of the federal government. The city's "district attorney" is the U.S. attorney for the District of Columbia, appointed by the president. The budget, passed by the city council and approved by the mayor, is reviewed and enacted by Congress. Moreover, Congress retains the right to enact legislation on any subject for the District, whether within or outside of the scope of power delegated to the city council.

Administration of municipal services. As under previous forms of government, municipal functions remain in control of a combination of local and federal committees. School-board members, formerly appointed by the U.S. District Court for the District of Columbia, became popularly elected in 1968. Public utilities are under a Public Service Commission appointed by the president. The zoning of private property is handled by the Zoning Commission, consisting of the mayor, the chairman of the city council, the Architect of the Capitol, and the director of the National Park Serivce. The water supply is under the jurisdiction of an Army engineer, given the title of District Engineer, but its distibution is controlled by the mayor. The National Park Service supervises the public parks of the city.

Overlapping federal and District responsibilities

Public security and law enforcement are handled by four separate law-enforcement agencies, each with its own jurisdictional area. Under the mayor is the Metropolitan Police Force, which has the responsibility for enforcing the laws and ordinances of the municipal government. The Capitol Police are repsonsible for the security of the Capitol building and its grounds. The White House Security Guard protects the White House and the president, while the National Park Police are responsible for all public parks and many recreational facilities.

Court system. The unwieldiness of Washington's governmental apparatus has long been most apparent in the operation of its courts. Until the early 1970s legal jurisdiction over District matters was shared by two federal courts and three local courts, appeals from which were directed to separate appeals courts. The Court Reorganization Plan, was implemented in 1970 to reduce the confusion and inefficiency of this judiciary system.

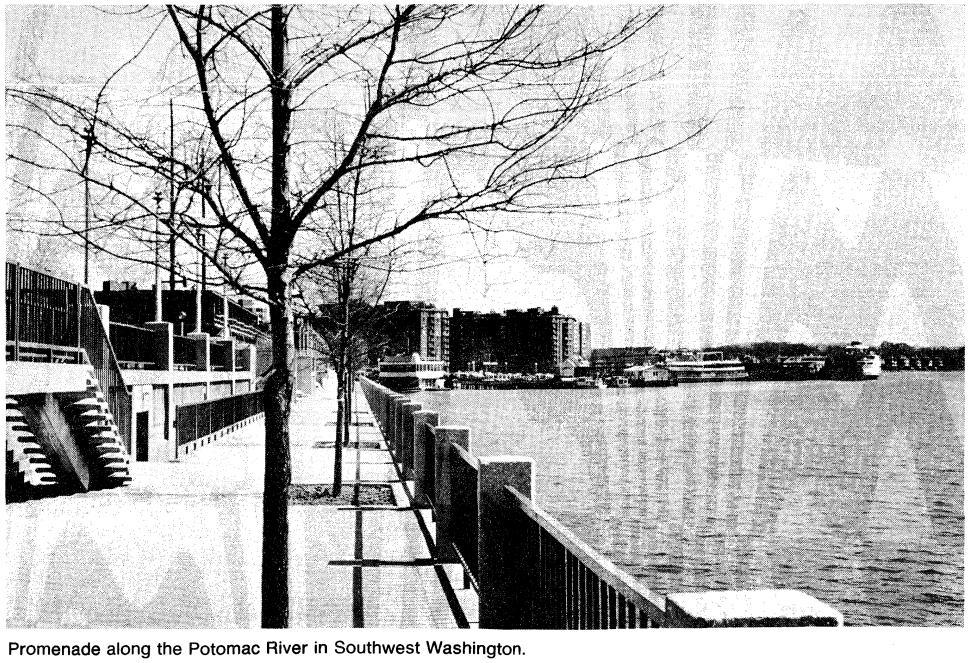
Under the plan a single trial court, the Superior Court of the District of Columbia, was established to assume the functions of all the former federal and local courts. A single appeals court, the U.S. Court of Appeals for the District of Columbia, was established to function in a manner similar to a state supreme court. For the first time in its history, Washington had an integrated court system similar to the systems in all of the states.

Public services. Washington's multitude of intertwined and seemingly insoluble social problems—such as race, poverty, crime, civil disorder, institutionalized inequality in education and other fields, and environmental pollution—is by no means unique. Like other cities, it finds social needs increasing as its human and financial resources decline. The District government can accomplish little in the way of service to its citizens without the support of the federal government. Some authorities have seen the suburban areas, with the high educational levels of their population and without high proportions of blue-collar workers, as being more willing than the federal government to assist the plight of the core city.

Employment. Two-thirds of Washington's employed live outside the District of Columbia and represent work force that is highly specialized, professionally skilled, and well above the national averages in educational level and per capita income. Among the residents of the city of Washington, where a majority of the residents are black, the unemployment rate, especially among youth, ranges well above the national average. Educational level and per capita income are considerably below the averages found in the surrounding suburbs. Nothing expresses the dilemma of the "two" Washingtons more clearly than the characteristics and location of its labour force.

Housing. In spite of the fact that the Washington metropolitan area experienced a decline in its rate of population growth (with an absolute decline in the District of Columbia) in the 1970s, the demand for housing has remained high. This demand was generated by a high rate of household formation, primarily among young adults (mostly affluent, one-family or single householders) from the post-World War II "baby boom" generation. Competition for existing housing resources has intensified between

Photo Research International



Promenade along the Potomac River in Southwest Washington.

new householders and existing occupants. This has stimulated not only the conversion of rental apartments to condominiums but also the renovation of many single-family dwellings formerly occupied by lower income renters. While this "gentrification" process has resulted in the upgrading of a number of residential areas in the District, it has aggravated the problems of household displacement, primarily among lower income black families, and has increased the pressure on the older, inner suburbs of the metropolitan area.

Fair housing at the federal, state, and local levels has reduced the more blatant forms of discrimination against racial minorities, but new construction of low-rental housing has been negligible. The consequences for this segment of the Washington population remain grave, and housing continues to be a major public-policy and local issue.

Education. More than any other service of the District government, the public school system must be understood in the context of the city's racial composition. Two public-opinion polls taken in Washington in 1954 clearly indicated that the majority of white residents disapproved of the U.S. Supreme Court decision declaring racial segregation of public schools to be unconstitutional. This decision probably contributed significantly to the subsequent large-scale emigration of white families from the District. A number of whites who remained in the District at that time sought transfers for their children to schools with low black enrollments, while many other children were transferred to private schools.

In 1956 the public school system initiated a new program, now familiarly known as the track system, or ability grouping, into four curricular tracks: honours, college preparatory, general, and basic. The objective of the program was to allow students of similar academic achievement to work together regardless of race. Opposition to this system grew in intensity throughout the 1960s. To many, the track system produced only a new form of segregation, since most of the black students gravitated to the general or basic curriculums. The system was abolished in 1966 through a court action brought by civil rights leaders against the District superintendent of schools.

Imbalances and restructurings of schools

By the early 1970s the public school system had become 94 percent black. A 15-month study of the situation after the court action of 1966 concluded that no school system with so heavy a racial imbalance could consider desegregation in any meaningful way, that desegregation methods that might work in other cities were largely irrelevant in Washington.

Aspects of black militancy. During the 1960s, and particularly after the fusion of the civil rights movement and the antipoverty programs, racial politics became of increasing importance in Washington, affecting the administration of all public services. After 1964 many new groups were formed that reflected the increasing militancy among the city's blacks who are the recipients or the victims of public services in the District.

In the framework of dissatisfaction and turmoil over the issue of desegregation and quality of education, Congress granted District citizens the right to elect their own Board of Education in 1968. The advent of this elected school board was viewed by many as the beginning of progress toward a public school system more responsive to the needs and aspirations of the predominantly black community.

On the other hand, Congress retained the ultimate right to review all budget requests upon which any new educational programs must rely. In addition, activity on behalf of the public schools tended to be divided between groups that focussed their activity on the school board and the District government and groups that tended to bypass District agencies and take their case directly to Congress.

Higher education. Several institutions of higher learning offering undergraduate and graduate programs are located in the metropolitan area. Georgetown University, founded in 1789 as a seminary, is the area's oldest; with the Catholic University of America (opened in 1889), it makes Washington a centre of Roman Catholic education. The George Washington University, chartered as Columbian College in 1821, and American University, chartered in 1893, have become large, diversified universities. Howard University, chartered in 1867 for the higher education of blacks, offers a wide variety of graduate and professional programs and is supported largely by federal appropriations. Federal City College, founded in 1969 and renamed the University of the District of Columbia in 1977, is a municipal institution. These six institutions within the District form the Consortium of Universities, which offers a combined facility for undergraduate and graduate education and a common pool of resources. Gallaudet College (founded in 1857 and an associate member of the Consortium), for the education of the deaf, receives both private and federal support. The University of Maryland, George Mason College, and the Northern Virginia Campus of the University of Virginia are located in suburban areas.

In addition, numerous federal and privately funded institutions and associations produce a variety of both statistical and scholarly materials. The Brookings Institution is a nationally recognized centre for research in politics and economics; and the jointly administered National Academy of Sciences, National Academy of Engineering, Institute of Medicine, and National Research Council provide public and private bodies with the research potentiality and publications of their memberships, comprising many of the most eminent scientists in the nation. The National Institutes of Health offer both a centre and funding for basic and applied research in wide areas of medicine and mental health.

CULTURAL LIFE

Until the end of World War II, Washington was literally outside the main centres of American arts and letters, and the original plan for the city to become a focus of national culture was far short of realization. Its institutions of higher education were uniformly small, poor, and struggling; it had no significant collections of art; its facilities for the performing arts were either too small, inconveniently located, or nonexistent; and the only truly national museum, the Smithsonian Institution, was laughingly characterized as "the nation's junk closet."

In early descriptive guides to the city, the section on cultural facilities was devoted largely to the number and variety of churches, which were distinguished mostly for the presidents of the United States who had attended them. Practically alone on the scene was the magnificent Library of Congress, with the largest collection of books, maps, newspapers, documents, and manuscripts in the world. Since World War II, however, the development of cultural and recreational facilities in Washington has been one of the city's—and the nation's—proudest achievements.

The arts. Since its opening in 1941, the National Gallery of Art has become recognized internationally as one of the world's major art collections. Together with the Corcoran Gallery of Art, the Phillips Collection, the Freer Gallery of Art, and the National Portrait Gallery and National Museum of American Art, the capital city has achieved a position of eminence in the art world.

Major art collections

Washington's Arena Stage has been recognized since its founding in 1950 as one of the leading regional theatre companies in the nation, both for the excellence of its productions and for its leadership in seeking out new playwrights. The National Symphony Orchestra has begun to achieve national acclaim. Aside from these features, the performing arts in the city comprised little more than road shows from New York City or pre-Broadway tryouts until the opening in 1971 of the John F. Kennedy Center for the Performing Arts, with its three theatres for music, opera, and drama. The renovation of historic Ford's Theatre, the scene of the assassination of Pres. Abraham Lincoln in 1865, has offered Washingtonians a unique opportunity to enjoy programs of cultural value in a setting of historical significance.

Architectural Washington holds interest outside its monumental areas. Prominent examples include the many fine 19th-century mansions, many of which are occupied by foreign embassies, along Massachusetts Avenue and in the area around Dupont Circle. Entire blocks of slum buildings of the Capitol Hill and other areas, deteriorated from a grander past, have undergone considerable renovation. Numerous homes in the Georgetown area, similarly re-

newed beginning in the 1930s, offer fine examples of the American Federal style of architecture dating from the early years of the nation. A stunning example of landscape architecture including formal gardens and a bird sanctuary, as well as a museum and research library of medieval art, is on display at Dumbarton Oaks.

Smithsonian Institution

National collections. The Smithsonian Institution has reorganized and more effectively displayed its enormous collections of natural and historical material. The National Museum of Natural History has been since 1951 the largest permanent exhibition in the world of anthropological, biological, and geological specimens relating to the United States and its possessions. The National Museum of History and Technology presents the history of social and technological achievements. Throughout the 1970s the growth of Smithsonian Institution facilities was dramatic. The addition of the East Building of the National Gallery of Art, designed to house a growing collection of 20th-century art and international exhibitions, has provided Washington with opportunities enjoyed by only a handful of U.S. cities. The National Air and Space Museum was constructed for the display and study of America's accomplishments in flight. The Hirshhorn Museum and Sculpture Garden houses a collection of 19th- and 20th-century art. Two new museums opened in 1987: the Arthur M. Sackler Gallery, devoted to Asian and Middle Eastern art, and the National Museum of African Art. Located together on the Mall between the Capitol and the Washington Monument, these various facilities of the Smithsonian Institution attract millions of visitors each year. The nearby National Archives supplements the riches of the Library of Congress with its thorough collection of documents from American history, headed by the original Declaration of Independence and Constitution.

Festivals. The traditional Cherry Blossom Festival held early each spring is the oldest of Washington's celebrations that combine the efforts of federal, civic, and commercial personnel. In spite of their attraction for tourists, the festival parade and crowning of a queen tend to be less important for Washingtonians than is the explosion of blossoms around the Tidal Basin and the Jefferson Memorial that the festival marks.

Also popular is Washington's annual festival of American folk arts and crafts, held for a week each summer on the Mall. This open-air display of the arts and crafts of many regional and ethnic subcultures has become an increasingly significant cultural event.

Recreation. Much of Washington's elaborate system of public parks and other open spaces was designed to heighten the visual impression of the federal buildings, but they have been applied increasingly to recreational purpose. The original L'Enfant plan set aside 17 areas for parks, including the Mall, the Washington Monument grounds, the Capitol Hill grounds, the White House grounds (including the present Lafayette Square), and the small plots of grass at the intersections of the major avenues. The preservation of these areas and the addition of others provides Washington with the several hundred parks and green spaces whose maintenance and development have been functions of the National Capital Parks, under the Department of the Interior, since 1791.

Rock Creek Park and the out-of-doors

In 1890 the city acquired Rock Creek Park, one of the largest natural parks within the boundaries of any city in the world. It comprises 1,754 acres (710 hectares) of virgin woodland forming the mile-wide valley of Rock Creek, which traverses the northwestern quarter of the city from suburban Silver Spring to the Potomac and is extensively developed for recreational uses. Within its boundaries is the National Zoological Park, occupying 176 acres of picturesque wooded hills and grassy meadows. Also within Rock Creek Park is the Carter Barron Amphitheater, where musical programs are presented in a wooded setting during the summer. Nature walks, horse trails, picnic areas, and sports facilities are spread throughout the park.

For 15 miles, from Georgetown to the Great Falls of the Potomac River (the shores of which have been protected from uncontrolled development on both the Virginia and District sides), stretches the historic Chesapeake and Ohio Canal, the original project of George Washington, which has been restored and declared a national monument. The towpath beside the canal is now available for hiking and bicycling, and regularly scheduled trips are enjoyed by visitors in reconstructed, mule-drawn canal boats. On the Virginia shore is the scenic George Washington Memorial Parkway, from which motorists may enjoy a nearly uninterrupted view of the Potomac from above the northwestern boundary of the District to Mount Vernon, the home of Washington, below Alexandria. Between Georgetown and the Virginia shore lies Theodore Roosevelt Island, once a private plantation but now a nature sanctuary. The island is used as a wilderness strolling place for Washingtonians seeking relief from the city and offers one of the area's most tranquil and beautiful natural settings.

Rock Creek Park.
Carl Purcell

History

THE EARLY PERIOD

The idea of a national capital city seems to have originated at a meeting of the Congress in June 1783 in the Old City Hall in Philadelphia. The War of Independence had but recently been concluded, the treasury was empty, the nation had no credit, and it was heavily in debt to its soldiers for back pay. There was no president, and, though the 13 colonies were free, they remained a gathering of semi-independent sovereignties with divergent interests. On June 20 a large body of unpaid soldiers invaded Philadelphia to present their grievances to Congress.

Conception, siting, and design. No actual violence occurred, but a number of congressmen started a movement to establish a federal city where the lawmakers could conduct the business of government without fear of intimidation. Several locations were considered over the next six years, but Northern and Southern disagreements prevented decision until 1790.

Bases of site selection

Although the decision to locate the capital on the Potomac was largely a political compromise, selection of the exact site for the city was left to the newly elected president, George Washington. The chosen district, or territory

as it was first called, was 10 miles square. Georgetown to the north and Alexandria to the south were both in the original district, while a third village, Hamburg, lay by the riverfront swamps in a part of Washington known traditionally as Foggy Bottom.

Important to Washington in his selection was the site's commercial potential. The river was navigable to Georgetown, an important tobacco market. The construction of a canal from there across the Cumberland Gap to the "western frontier" would tap the produce of the vast country beyond that was being opened to settlement. The President had established a private canal construction company before the final decision had been reached, but he immediately relinquished his interests in it.

While in Philadelphia, Washington negotiated with Pierre-Charles L'Enfant to lay out a plan for the new city. A volunteer in the Revolution whose democratic idealism was unquestioned, a well-trained engineer, and an artist who had designed the setting for the President's inaugural ceremonies in New York City, L'Enfant was highly respected and admired. Apparently sensing the historic significance of his appointment, he conceived his plan on a grand scale.

The Capitol's cornerstone was laid by Washington in September 1793, and construction was begun on the White House, designed by an Irishman, James Hoban, and on a modest cluster of nearby office buildings to house governmental departments. In October 1800 the archives, general offices, and officials of the government were moved to Washington from Philadelphia. Pres. John Adams took up residence in the White House, and the Congress met for the first time in the newly completed Senate wing of the Capitol.

The young city. Descriptions of life in early Washington reveal many of the shortcomings resulting from establishment of a capital city by fiat amid what was essentially a wilderness. What was conceived as a "city of magnificent distances" or, in Washington's words, "the Emporium of the West" was referred to by various statesmen and congressmen as "wilderness city," "Capital of Miserable Huts," "A Mud-hole Equal to the Great Serbonian Bog," and similar epithets. By the close of Thomas Jefferson's term of office in 1808, the population was scarcely 5,000. Until the introduction of the steam engine and the telegraph, a more or less continuous agitation went on in Congress and in the national press to move the capital because of its remoteness and inaccessibility.

Achievement of identity as the national capital

In 1814 the capital was temporarily abandoned as the result of the invasion by a British force under Admiral Sir George Cockburn, who ordered the burning of the Capitol, the White House, and the Navy Arsenal. Although this action was rather inconsequential to the outcome of the War of 1812, it had the effect of solidifying Washington in the minds of many Americans as the national capital. Public indignation over destruction of the seat of government ended all significant movements to relocate the federal city, and Washington became the national capital in fact as well as in name. By the outbreak of the Civil War, the intensity of this image was firmly established. The course of that conflict was deeply affected by the actions of the federal government to defend Washington at all costs from nearby Confederate forces, who often threatened the city from several sides. If the Civil War was the final stage of the historical process that changed a loose confederation of states into a united republic, it was also effective in completing the identity of Washington as the capital of the United States.

EVOLUTION OF THE MODERN CITY

Originally, the city of Washington and the District of Columbia were not coextensive, either geographically or administratively. The 10-mile-square district was reduced by about one-third in 1847 by the return of the land south of the Potomac to the state of Virginia. Alexandria city resumed its former independent existence, while Arlington County was created from the remainder. Self-governing bodies within the district existed until 1895, when Georgetown was annexed by Washington.

During and after the Civil War the District's population more than doubled within a few years, suddenly including 40,000 freed slaves who set a pattern of racial diversity that was to have a major impact on the city's life. The following century was filled also with physical and demographic growth and change within the city, with numerous political modifications attempting to harmonize the District's needs with its inherent status in relation to Congress, and with continuing activities to refine the cultural and monumental image of the city.

By the latter third of the 19th century the city had constructed an impressive number of monuments, but then and later many slums began to intrude on the city's image. Transformation of this and other problems was cut short by World Wars I and II but carried on after their conclusions. During the 20th century the city increasingly became a Mecca not only for visitors seeking from its many shrines a sense of the American heritage but also for protest groups—from several groups of veterans seeking pensions to blacks seeking relief from poverty and repression and youths decrying the war in Vietnam.

BIBLIOGRAPHY. For a general description and guide to the city, see JUDY DUFFIELD, *Washington, D.C.: The Complete Guide* (1982); and E.J. APPLEWHITE, *Washington Itself* (1981). For early descriptions of the city and its life, see CHRISTEN HINES, *Early Recollections of Washington City* (1866); and GEORGE WATTERSON, *A Picture of Washington* (1840). A complete and reliable history is WILHELMUS B. BRYAN, *A History of the National Capital,* 2 vol. (1914–16). More recent histories include CONSTANCE GREEN, *Washington,* 2 vol. (1962–63, reissued 1976); and DAVID L. LEWIS, *District of Columbia: A Bicentennial History* (1976). Among excellent works on planning for the city are GEORGE W. GRIER, *Washington and its Suburbs* (1971); FREDERICK A. GUTHEIM, *Planning Washington 1924–1976* (1976); UNITED STATES NATIONAL CAPITAL PLANNING COMMISSION, *Worthy of the Nation: The History of Planning for the National Capital* (1977); ATLEE E. SHIDLER (ed.), *Greater Washington in 1980* (1980); GEORGE and EUNICE GRIER, *Washingtonians in the 1980's* (1981); NELSON F. RIMENSNYDER, *The Political Evolution of the District of Columbia: Current Status and Proposed Alternatives* (1975); LEAGUE OF WOMEN VOTERS OF THE DISTRICT OF COLUMBIA, *Know the District of Columbia* (1980); and DISTRICT OF COLUMBIA OFFICE OF PLANNING AND DEVELOPMENT, *Comprehensive Plan for the National Capital: District of Columbia Components* (1982).

(Ri.W.S.)

Wellington

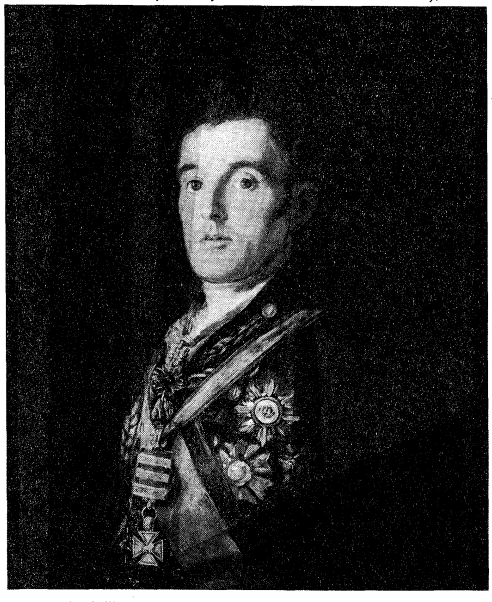
Arthur Wellesley, 1st duke of Wellington, twice reached the zenith of fame with a period of unexampled odium intervening. By defeating Napoleon at Waterloo he became the conqueror of the world's conqueror. After Waterloo he joined a repressive government, and later, as prime minister, he resisted pressure for constitutional reform. False pride, however, never prevented him from retreating either on the field or in Parliament, and for the country's sake he supported policies that he personally disapproved. In old age he was idolized as an incomparable public servant—the Great Duke. Reaction came after his death. He has been rated an overcautious general and, once, Britain's worst 19th-century prime minister. Today there is widespread appreciation of his military genius and of his character as an honest and selfless politician, uncorrupted by vast prestige.

Early life. Arthur Wesley (later, from 1798, Wellesley) was born on May 1, 1769, in Dublin, the fifth son of the 1st earl of Mornington. Too withdrawn to benefit from his Eton schooling, he was sent to a military academy in France, being, in his widowed mother's words, "food for powder and nothing more." At the age of 18 he was commissioned in the army and appointed aide-de-camp to the Irish viceroy. In 1790–97 he held the family seat of Trim in the Irish Parliament. At 24, though in debt, he proposed to Catherine (Kitty) Pakenham but was rejected. Arthur abandoned heavy gambling to concentrate on his profession. As lieutenant colonel of the 33rd Foot by purchase, he saw active service in Flanders (1794–95), learning from his superiors' blunders. After failing to obtain civil employment, he was glad to be posted to India in 1796.

Service in India

In India he adopted a regimen of abstemiousness and good humour. The arrival of his eldest brother, Richard, as viceroy enabled him to exploit his talents. He commanded a division against Tipu Sultan of Mysore and became governor of Mysore (1799) and commander in chief against the Marāthās. Victories, especially at Assaye (1803), resulted in a peace that he himself negotiated. All the successful qualities he later exhibited on European battlefields were developed in India: decision, common sense, and attention to detail; care of his soldiers and their supplies; and good relations with the civilian population.

By courtesy of the Trustees, The National Gallery, London



Duke of Wellington, portrait by Francisco de Goya, 1812. In the National Gallery, London.

Napoleon was unwise in later writing him off as a mere "Sepoy general." Wellesley returned to England in 1805 with a knighthood.

Wellesley's new assignments were disappointing: an abortive expedition to Hanover, followed by a brigade at Hastings. But he felt he must serve wherever duty required. One duty was to marry his faded Kitty in 1806; another was to enter Parliament in order to repel radical attacks on his brother's Indian record. He spent two years in Ireland as Tory chief secretary. On a brief military expedition in Copenhagen (1807), a welcome break, he defeated a small Danish force. When in 1808 the Portuguese rose against Napoleon, Wellesley was ordered to support them.

Victory in the Napoleonic Wars. Wellesley did not intend to be "half beaten before the battle began"—the usual effect on continental armies of Napoleon's supremacy. With "steady troops" he expected to master the French attack. His "thin red line" of British infantry did indeed defeat Gen. Andoche Junot's columns at Vimeiro (August 21), but the arrival of two superior British officers prevented a pursuit because they preferred to sign the unpopular convention of Sintra, whereby Junot's army was repatriated. Public outcry brought about the court-martial of Wellesley and his colleagues. Though acquitted, Wellesley returned to Ireland as chief secretary. After the British evacuated Spain, however, he persuaded the government to let him renew hostilities in 1809, arguing that Portugal could still be held, a decision that was crucial to Europe.

Campaign in Portugal and Spain

Landing at Lisbon, he surprised Marshal Nicolas-Jean de Dieu Soult, captured Oporto, and chased the French back into Spain; but a joint Anglo-Spanish advance on Madrid failed, despite a victory at Talavera (July 27–28). Though rewarded with a peerage for his offensive, Viscount Wellington retreated with his greatly outnumbered force to his Portuguese base, defeating Marshal André Masséna at Bussaco on the way (September 27, 1810). He had secretly fortified the famous "lines of Torres Vedras" across the Lisbon peninsula. Masséna's evacuation of Portugal in the spring of 1811 and the loss of Fuentes de Oñoro (May 3–5) triumphantly justified Wellington's defensive, scorched-earth policy and confirmed his soldiers' trust in him. He was nicknamed "nosey" by his men, and "the beau" by his officers, for his slim five feet nine inches, the perfectly cut civilian clothes he preferred to wear, his wavy brown hair, and brilliant blue eyes.

His slowly growing army was not strong enough to capture the Spanish fortresses of Ciudad-Rodrigo and Badajoz until 1812. Then, having defeated "40,000 Frenchmen in 40 minutes" at Salamanca (July 22), he entered Madrid (August 12). His siege of Burgos failed and his army retreated again to Portugal, from which it was launched for the last time into Spain in May 1813. After a dash across the peninsula, he brought the French to bay at Vitoria, routing them and capturing all their baggage (June 21). This glittering prize was too much for the victors, who let the French escape into the Pyrenees, while Wellington denounced his drunken troops as "the scum of the earth." The victory at Vitoria gave impetus to the European alliance against Napoleon, and Soult's initial success in the Pyrenees could not prevent Wellington from taking San Sebastián and Pamplona. Whey dry weather came, Wellington invaded France, crossing the river lines one after another until on April 10, 1814, he stormed into Toulouse, thus ending the Peninsular War. (Four days earlier Napoleon had abdicated.) Already marquess and field marshal, he was now created a duke, with the nation's gift of £500,000 and later of Stratfield Saye in Hampshire to keep up his position.

With Napoleon on Elba, Wellington was appointed ambassador to the restored Bourbon court of Louis XVIII. In February 1815 he took the place of Viscount Castlereagh,

the foreign secretary, at the Congress of Vienna, but, before delegates could finish their peacemaking, Napoleon had escaped, landing in France (March 1) to begin his Hundred Days. The victory of Wellington and the Prussian field marshal Gebhard Leberecht Blücher on June 18 at Waterloo established the Duke as Europe's most renowned—if not most jubilant—hero. "I hope to God that I have fought my last battle," he said, weeping for the fallen. "It is a bad thing to be always fighting." His hope was fulfilled. As commander in chief during the occupation of France, he opposed a punitive peace, organized loans to rescue French finances, and advised withdrawal of the occupying troops after three years. For these policies he won the gratitude of the peace congress, returning home in 1818 with the batons (symbol of field marshal) of six foreign countries.

Defeat of Napoleon

Role in the cabinet. Wellington's experiences abroad prevented him from ever becoming a party politician. Though he joined the Earl of Liverpool's Tory cabinet as master general of the ordnance, he exempted himself from automatically opposing a subsequent Whig government: "a factious opposition," he argued, "is highly injurious to the interests of the country." His identification with the party of law and order, however, increased when postwar discontent boiled over in the Peterloo Massacre at a Manchester demonstration for parliamentary reform and the Cato Street Conspiracy, a plot to murder the Cabinet. The popular George Canning succeeded Viscount Castlereagh as foreign secretary in 1822. Despite Canning's antipathy to the congress system, Wellington himself overbore George IV's personal objections to him, believing that the system was by now unshakably established. When Canning extricated Britain from its European commitments, Wellington was left to bitter self-reproach. His own diplomatic failures at the Congress of Verona (1822), at which he vainly sought to heal dissension among the European allies, and in Russia (1826) increased his chagrin. Straightforward to a fault, Wellington was unsuited to carrying out Canning's subtle policies, but he gained respect abroad as an honest man.

In 1825 Wellington turned to Ireland's problem, formulating it as a basic dilemma: political violence would end only after the Catholics' claim to sit in Parliament, known as Catholic Emancipation, had been granted; yet the Protestant establishment, or ascendancy, must be preserved. He worked privately at a solution, by which a papal concordat to ensure at least minimum control of Catholic clergy would be the precondition of Emancipation. When Canning, an unqualified Emancipator, became prime minister in April 1827, however, Wellington felt that Protestant ascendancy was in jeopardy. He and Robert Peel headed a mass exodus from the government, Wellington also resigning his command of the army. This action was interpreted as pique at the King's choosing his rival for prime minister. In denying the allegation, Wellington rashly asserted that he, a soldier, would be "worse than mad" to consider himself fit for the premiership. After Canning's death that August, he resumed his army command. Within five months Canning's successor, Viscount Goderich, had given up the task, and on January 9, 1828, the King summoned the Duke of Wellington.

Years as prime minister. The Duke's aim was to achieve a strong and balanced government by reuniting the Tory Party. Having reluctantly resigned again as commander in chief, he invited the Canningites, headed by William Huskisson, to serve, while dropping the ultra-Tories as incompatible with his policy of moderation. With the right wing thus alienated, a chasm began to open on the left. The opposition's demand for extensive reforms met with sympathy from Huskisson's group. Wisely, the Duke retreated, first on a church issue, himself reforming the Test and Corporation Acts that penalized Nonconformists, and again on a Corn Law (prohibiting importation of cheaper foreign grains) question, introducing a more liberal reform than he and the agricultural interest desired. Shortly afterward, however, he collided head-on with the Huskissonites on parliamentary reform; the whole group resigned in May. A further crisis immediately arose during the by-election in Clare, Ireland, where William Vesey-Fitzgerald, Huskisson's ministerial successor, defending his seat, was defeated by Daniel O'Connell, the Irish Catholic leader. The defeat of Vesey-Fitzgerald, a popular pro-Catholic, carried an alarming moral for the Duke: until Emancipation was granted, no Tory would win in southern Ireland. There might well be civil war. In August 1828 Wellington therefore undertook the most exacting political duty of his career—the conversion of George IV, Peel, who was now leader of the Commons, and a majority of Tories to Catholic Emancipation, a reform that they had hitherto regarded as anathema. It took six months of indefatigable persuasion behind closed doors to win over the King. Peel's position was equally problematic—as a publicly declared Protestant, he clung to the idea of supporting Emancipation only from the back benches; but finally Wellington's patience and Peel's generosity prevailed, and he agreed to continue leading the Commons. A number of ultra-Tories defied to the last Wellington's order to "right-about face," but the majority of the party obeyed. So in April 1829, though the Tories were split, Catholic Emancipation became law, the Duke's greatest political victory, with melodrama being added by his fighting a duel with an abusive ultra-Tory, the Earl of Winchilsea.

Crisis over Emancipation

Wellington has sometimes been criticized for inconsistency. It now appears that he was merely secretive in not taking the public into his confidence much earlier. His willingness for some form of Emancipation by 1825 might with advantage have been disclosed.

A demand for further changes, already stimulated by Wellington's own achievements, was powerfully reinforced by countrywide hardship during 1829–30 and canalized by Charles Grey, the 2nd Earl Grey, the Whig leader, into fresh moves for parliamentary reform that would allow industrial towns like Birmingham to have a voice in Parliament, in place of pocket boroughs owned by the nobility and gentry. Expression of dissatisfaction with Wellington's fatalistic attitude toward poverty and unemployment was made possible when the accession of William IV in 1830, following George IV's death, provided a general election. France's bourgeois revolution that same year—the July Revolution—greatly encouraged British reformers. Though Wellington's ministry survived, it was weakened, and Huskisson's sudden death frustrated tentative plans for reconciliation. Wellington saw parliamentary reform not as a panacea but as constitutional suicide. A fortnight before the opening of Parliament he wrote a letter to a friend denouncing reform as ruinous and disclosing his unalterable decision to oppose it. He staggered Parliament on November 2 with an uncompromising declaration against any reform whatever. A combination of reformers and vengeful ultra-Tories defeated him on the 15th. Peel made him resign the next day. He was succeeded by Grey.

Declaration against reform

As a soldier Wellington had shown uncanny ability in guessing what lay "on the other side of the hill." Through lack of political imagination, however, he saw revolution beyond the hill of reform—"revolution by due course of law." For this delusion he was deservedly called reactionary.

Last years. In opposition, the Duke proceeded to thwart Grey's attempts to get a reform bill through the Lords. Wellington's windows were twice smashed by radical mobs, and his iron shutters helped form the image of an iron duke. The titanic struggle culminated in the crisis of May 1832, which promised to end like the July Revolution of France. The King refused to create enough new peers to overwhelm the hostile Lords, Grey resigned, and Wellington failed to recruit an alternative government. Faced by tumultuous deadlock, Wellington, still opposing reform, then retreated for the sake of the country, persuading his followers to join him in absenting himself from Parliament until the Reform Bill became law in June. He was mobbed nonetheless by an angry crowd on Waterloo Day. "An odd day to choose," was his only comment.

The Duke's abstention had saved the Lords, and, as long as he led the Tory peers, he continued to steer them away from fatal clashes with the Commons. Whenever possible he supported the King's government. In 1834 William IV dismissed the Whigs by a political coup, summoning the Duke to form a ministry; but the 65-year-old duke

replied that Peel must be prime minister. This abnegation, most rare in a politician, did not go unappreciated. He served under Peel as foreign secretary (1834–35) and as minister without portfolio (1841–46). He also served as chancellor of Oxford, constable of the Tower, lord-lieutenant of Hampshire, and elder brother and later master of Trinity House, not to mention Queen Victoria's father figure. He made a mistake in holding the chief command of the army throughout his last 10 years, because he was past initiating the reforms that were later sorely needed. Nevertheless, he showed a touch of his old genius in 1848, when his calm handling of a threatened Chartist rising prevented any violence. Thanks to his again ordering the peers to "right-about face," this time over the Corn Laws, he enabled Peel to abolish them. Wellington retired from public life after 1846, though he was still consulted by all parties. Apsley House, his town residence at Hyde Park Corner, was known as No. 1 London. As lord warden of the Cinque Ports, he died at Walmer Castle, his favourite residence, from a stroke on September 14, 1852. He was given a monumental state funeral, the last heraldic one in Great Britain, and was buried in St. Paul's Cathedral.

Personal life. The phrase "retained servant of King and people" and variants of it were used repeatedly by the Duke of himself and aptly suggest the self-dedication for which he is chiefly honoured. Many amusing personal peculiarities in clothes and correspondence, together with a gift for repartee, made him a "character" as well as a hero. "Publish and be damned!" was his famous retort to a blackmailer. His marriage was not happy: Kitty both feared him and worshipped him to excess. She died on April 24, 1831. Of his two sons, the elder edited his latest *Despatches* and the younger produced the grandchildren to whom he was devoted, as he was to all children. His intense friendships with Harriet (the wife of Charles) Arbuthnot, Angela Georgina Burdett-Coutts, and others showed that he could have been happy with a clever woman; perhaps he was happiest of all, however, in the camaraderie of his staff—his military family. Some modern historians have objected to the posthumous title Iron Duke on the reasonable grounds that he was neither cold nor hardhearted. Yet he himself often boasted of his iron hand in maintaining discipline. His engaging simplicity and extraordinary lack of vanity were expressed in a favourite saying, "I am but a man." (E.Lo.)

BIBLIOGRAPHY. The *Wellington Dispatches,* published in three series, are essential source material: *Dispatches 1799–1818,* ed. by GURWOOD, 12 vol. (1834–38; Index, 1839); *Supplementary Despatches . . . 1794–1805,* 15 vol. (1858–72); and *Despatches, Correspondence, and Memoranda . . . ,* new series, ed. by the 2nd DUKE OF WELLINGTON, 8 vol. (1867–80). See also *Wellington: Political Correspondence,* ed. by JOHN BROOKE and JULIA GANDY (begun in 1976, covering the years 1833–46). Two outstanding biographies are SIR HERBERT MAXWELL, *The Life of Wellington,* 2 vol. (1899–1900); and PHILIP GUEDALLA, *The Duke* (1931, reissued 1976), though they did not have the advantage of material later collected and edited by the 7th DUKE OF WELLINGTON: *The Journal of Mrs. Arbuthnot, 1820–1832,* 2 vol. (1950); *Wellington and His Friends* (1965), being selected letters to Mr. and Mrs. Arbuthnot and three other women; *A Selection from the Private Correspondence . . .* (1952); and *Conversations of the First Duke of Wellington with George William Chad* (1956), a slight but useful diary. GEORGE R. GLEIG, *The Life of Arthur Duke of Wellington* (1862), and *Personal Reminiscences* (1904), are comprehensive but inaccurate; MURIEL WELLESLEY, a relative, pays tribute in *The Man Wellington Through the Eyes of Those Who Knew Him* (1937) and in *Wellington in Civil Life* (1939). CHARLES CRUTTWELL, *Wellington* (1936); and FRANCIS J. HUDLESTON, *Warriors in Undress* (1925, reissued 1969), give unsympathetic pictures. Military studies and biographies include S.G.P. WARD, *Wellington's Headquarters* (1957); MICHAEL GLOVER, *Wellington As Military Commander* (1968), and *Wellington's Army in the Peninsula: 1808–1814* (1977; U.S. title, *The Army of the Duke of Wellington in the Peninsula,* 1977); JAC WELLER, *Wellington in the Peninsula, 1808–1814* (1962), *Wellington at Waterloo* (1967), and *Wellington in India* (1972); and SIR ARTHUR BRYANT, *The Great Duke* (1971). Large-scale biography is found in ELIZABETH LONGFORD, *Wellington: The Years of the Sword* (1969), and *The Pillar of State* (1972, both reissued in 1982). Of especial interest are the perceptive and scholarly essays in *Wellington Studies,* ed. by MICHAEL E. HOWARD (1959). Valuable reminiscences by men who knew and admired Wellington are PHILIP STANHOPE, *Notes of Conversations with the Duke of Wellington 1831–1851* (1938); and ELLESMERE, *Personal Reminiscences of the Duke of Wellington* (1903). Collections of anecdotes proliferated after his death, the best being SIR WILLIAM FRASER, *Words on Wellington* (1899).

The West Indies

The West Indies region includes all of the islands that extend through the Caribbean Sea from the tip of the Florida Peninsula to the northern coast of South America. They include 23 political entities, some of them quite small and relatively unknown to the outside world. The West Indies derives its coherence and distinctiveness from a combination of four factors, one geographic, the other three historical. The geographic feature is insularity, and the three historical themes are colonialism, the sugar plantation, and slavery.

The West Indies range in size from Cuba, with an area of 42,804 square miles (110,861 square kilometres) and more than 10 million inhabitants, down to the tiny, uninhabited rocky islets of the Grenadines and the Virgin Islands. Most of the islands are smaller than Barbados, which has an area of 166 square miles and a quarter of a million inhabitants. Their physical fragmentation and the small size of most islands prepared the West Indies for a long history of colonialism and external domination and dependency. Territorial partition of the West Indies among the Spanish, French, British, Dutch, and Danish brought heavy recurrent fighting to the region in the 17th and 18th centuries; colonial possessions were captured in time of war only to be traded through peace treaties in furtherance of the geopolitical ambitions of the imperial powers.

Colonial partition of the West Indies has had a number of important consequences. Interisland connections have remained weak, because each colonizing power tended to dominate relationships with its own colonies. Where interisland linkage has occurred, it has been confined largely to islands belonging to the same European language group (each of these usually being of the same colonial affiliation). Moreover, the method of decolonization has generally followed these same political-linguistic lines. A further differentiating factor stems from this linguistic fragmentation. The Hispanic islands have a double identity—West Indian (or Caribbean) but also Latin-American; this is lacking in the English-, French-, and Dutch-speaking territories, where identities are above all insular. (The term Latin America is, however, often used geographically to encompass all of the territory south of the United States to Antarctica.)

If colonialism has been effaced over time by decolonization—and some parts of the West Indies experienced almost 500 years of European imperialism—the dominating economic entity of sugar monoculture, too, has diminished in all but Cuba and Saint Kitts. The rise of West Indian peasantries after the 19th-century slave emancipation, and economic diversification into industry, tourism, mining, and oil and natural gas drilling in the 20th, have transformed the economic base in most societies, especially since World War II. Plantations continue to occupy vast tracts of land, but their social significance has declined as people have left the countryside, either moving to the towns or emigrating to other countries. Still, dependence on one or a few products is the major economic problem in most of the states.

The most blatant forms of racial discrimination built into West Indian systems of slavery began to disappear with emancipation more than a century and a half ago, and the process has been hastened by decolonization. Nevertheless, most West Indian societies remain stratified by colour, culture, and class, and even in postrevolutionary Cuba, which is predominantly a white society, blacks are notably absent from the revolutionary leadership, despite the official elimination of class and colour lines. In the English-, French-, and Dutch-speaking territories, decolonization has created a mulatto and black political elite, although its position has been challenged in Trinidad and Tobago, where the large East Indian population has for long been in opposition, having been included in a coalition government since 1986. Generally the black political elites, supported by white and mulatto commercial interests, have promoted a multiracial ethic, which in most instances accepts the social order. This has been less true of Jamaica in particular, where white emigration, a powerful and creative arts movement, and the entrenchment of nationalistic governments since 1972 have led to the elaboration of a strong, black national identity.

Despite the region's common historical background and insularity, little cohesiveness or unity has developed among the West Indian states. Divided linguistically and culturally, the West Indies can for practical purposes be categorized into four distinct groupings according to common historical and political backgrounds: a Hispanic group plus Haiti, with long histories of political dictatorship and recurrent U.S. intervention; the British West Indies—now usually called the Commonwealth Caribbean (which includes mainland Belize and Guyana) and united through various regional associations such as the Caribbean Community and Common Market (Caricom) and the Organization of Eastern Caribbean States (OECS); the French Antilles (Guadeloupe and Martinique), now politically incorporated with their dependencies into France proper as overseas *départements;* and the Netherlands Antilles, from which Aruba was constitutionally separated in 1986 pending a decision on independence. Haiti and the Dominican Republic have observer status at Caricom meetings, but Cuba's membership in the socialist, mostly eastern European economic group Council for Mutual Economic Assistance (Comecon) and Puerto Rico's quasi-colonial ties to the United States make Pan-West Indian economic and political cooperation a difficult realization. The one institution of which virtually all West Indian states are members is the regional meteorologic service: hurricanes ignore all distinctions of West Indian people.

This article is divided into the following sections:

The region 722
- Physical and human geography 722
- History 729

The Greater Antilles 732
- Cuba 732
- Dominican Republic 740
- Haiti 746
- Jamaica 750
- Puerto Rico 755

The Lesser Antilles 762
- Anguilla 762
- Antigua and Barbuda 762
- Barbados 763
- Dominica 765
- Grenada 766
- Guadeloupe 768
- Martinique 769
- Montserrat 771
- Saint Kitts and Nevis 772
- Saint Lucia 772
- Saint Vincent and the Grenadines 774
- Virgin Islands 775

Other islands 778
- Aruba 778
- The Bahamas 778
- Cayman Islands 781
- Netherlands Antilles 782
- Trinidad and Tobago 783
- Turks and Caicos Islands 786

Bibliography 786

THE REGION

Physical and human geography

THE LAND

Relief, drainage, and soils. The West Indian archipelago stretches for more than 2,000 miles (3,200 kilometres) from Cuba almost to the north coast of South America. This island chain, divided principally into the Greater and Lesser Antilles, separates the Caribbean Sea from the Atlantic Ocean. Other islands of the region are isolated groups on the continental fringe of the Antilles, including The Bahamas, Trinidad and Tobago, and the Netherlands Antilles. (Bermuda, although sometimes discussed with the region, is not physically related to the other islands.)

The shape and alignment of the Greater Antilles—Cuba, Jamaica, Hispaniola (Haiti and the Dominican Republic), and Puerto Rico (to which the Virgin Islands are structurally related)—are determined by an ancient chain of folded and faulted mountains that in Cretaceous times extended from Central America through the Caribbean. Remnants of this system occur in the Blue Mountains of Jamaica and in the Sierra de los Órganos and the Sierra Maestra in Cuba. Duarte Peak in the Dominican Republic, another component of this range, rises to some 10,400 feet (3,170 metres) and is the highest point in the Caribbean.

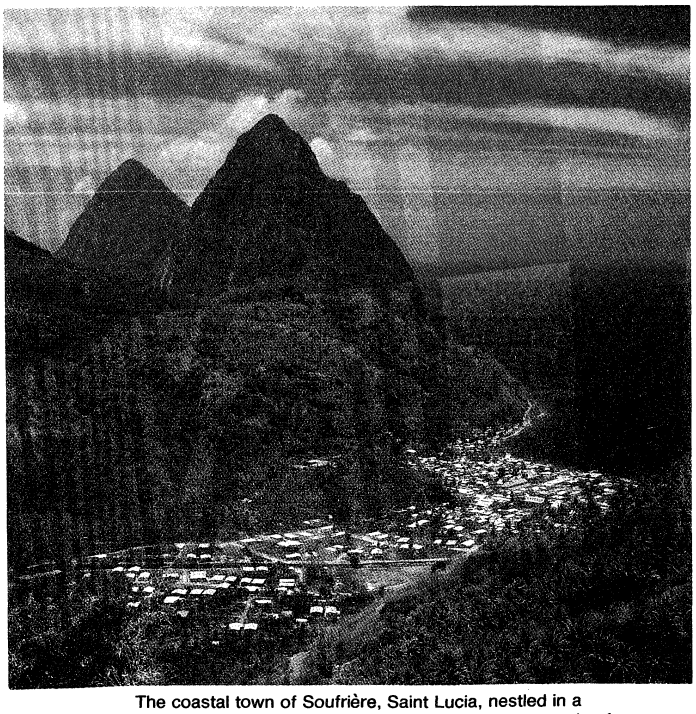
Geologic formation

Each Greater Antillean island has an encircling coastal plain, backed on the north coast of Cuba, Jamaica, and Hispaniola by Pleistocene-raised shorelines that reach heights of 1,000 feet. Structural depressions separate the original mountain ranges from interior valleys, notably in the Enriquillo basin in the Dominican Republic, which is below sea level, and in the Cayman Trench between Cuba and Jamaica, where the depth exceeds 25,200 feet, the deepest point in the Caribbean Sea. The Caribbean is a physically volatile zone characterized by frequent seismic disturbances, some of them disastrous, as the devastating Jamaican earthquakes at Port Royal (1692) and Kingston (1907) and the eruption of Mount Peleé on Martinique will recall.

The Lesser Antilles begin east and south of the Virgin Islands and are composed of a double arc of small islands. (For practical purposes the Virgin Islands are included in this discussion of the Lesser Antilles.) Stretching from Saint Kitts to Grenada, the mountainous inner arc consists of volcanic cones, some still active. The outer arc—running from Anguilla to Barbados—is made up of low, flat islands, whose limestone surface overlies older volcanic or crystalline rocks. The Lesser Antilles are also often divided into a Leeward (northern) group and a Windward (southern) group.

On the continental fringes of the Greater and Lesser Antilles are located The Bahamas, Trinidad and Tobago, and numerous islands off the Venezuelan coast, including the southern Netherlands Antilles. The Bahamas are flat patches of porous limestone and coral; coral reefs also fringe many of the Greater Antilles. Trinidad's Northern Range is a continuation of the Andean system, as are the planated (flattened) islands north of Venezuela, but the southern plains of Trinidad are formed from deposition from the Orinoco River in Venezuela.

Robert Harding Picture Library



The coastal town of Soufrière, Saint Lucia, nestled in a valley near the Gros and Petit Pitons, two huge mounds of weathered volcanic rock. Volcanic activity is common to this part of the Caribbean.

Rivers

The drainage of the West Indies is characterized by short rivers, whose steep mountain courses, often across crystalline rocks, give way to sluggish channels on the coastal plains. The extensive dry season, which causes some of these rivers to dry up for part of the year, is notably intense on the southern, leeward coasts of the Greater Antilles. Irrigation must be used or deep wells dug in some of the drier areas. Also in the Greater Antilles the extensive limestone outcrops have given rise to a tropical karst landscape, which is characterized by the porosity of the rock and systems of underground drainage; the condition has created rounded, conical, or elongated hummocks (knolls) separated by depressions.

Topography and geology have been crucial to soil formation in the West Indies. In the mountains, especially in the Greater Antilles, the original forest cover supplied a high organic content to the surface layer. But deforestation has led to erosion and to the exposure of an infertile mountain soil, especially in Haiti and Jamaica. Soils of the limestone areas fall into two groups: terra rossa is a red bauxitic soil with substantial organic matter; rendzina suffers impeded drainage. Lowland soils occur on the coastal plains and in interior basins; they are the most fertile soils and, together with the volcanic soils of the Lesser Antilles, have been cultivated in sugarcane for more than 300 years, in many instances without loss of quality.

Climate. Most of the West Indies lies within the tropical zone astride the track of the northeast trade winds. The tropical maritime climate exhibits no summer or winter but rather a change from a wet to a somewhat drier season. Temperatures remain fairly constant throughout the year. At weather stations anywhere in the small islands and near the coasts of the larger ones, temperatures vary within a mean daily range of 10° to 15° F (6° to 8° C), fluctuating between the low 70s and mid-80s F (low 20s and upper 20s C) from December to April and between the mid-70s and upper 80s F from May to November. Inland in the Greater Antilles the cooling influence of the sea, and the sea breeze, is lost. Frost is unusual, even in the highest mountains, and extreme low temperatures rarely drop below 55° F (13° C). Relative humidity is high throughout the year, usually ranging between 65 and 85 percent.

Differences in size, shape, topography, and location in relation to the northeast trade winds influence the amount and seasonality of rainfall. Many stations in the Greater Antilles and Leeward Islands record 40 to 65 inches (1,016 to 1,651 millimetres) per year, with more than 200 inches (5,000 millimetres) on the highest peaks. There is, however, a marked rain-shadow effect on the Greater Antilles' southern coasts, which are distinctly arid, and Saint Martin and Anguilla have salt ponds and water cisterns. The entire Windward group is well watered; but to the south the rainfall diminishes, and Aruba and the Netherlands Antilles (Bonaire and Curaçao) at the southwestern extreme are desertic. Bonaire has one short wet season; Trinidad and Dominica have one wet and one dry season; the Greater Antilles have two wet and two dry seasons.

Climatic influences originating outside the West Indies bring extreme conditions. Cold fronts extend south from the North American high-pressure anticyclone system from December to April, but these winds, which produce a sudden drop in temperature, overcast skies, and light rain, seldom blow south of Antigua. Severe tropical storms originating over the Atlantic occasionally become hurricanes. Skirting north of Trinidad and Tobago (although these have been hit occasionally), they usually sweep through the Lesser Antilles before swinging west across the Greater Antilles. Hurricanes are a source of great concern because of the damage they cause to houses, crops, animals, and human life, and the economy of a hard-hit island can require a year or more to recover.

Hurricanes

Plant and animal life. Relief and climate have a direct bearing on the natural vegetation of the West Indies, but the influence of human beings is also manifest. Clearing and burning the vegetation have transformed the plant geographies of all but the most isolated and inhospitable regions.

The original vegetation of the plains was a sparse forest of lignum vitae and ceiba (cotton) trees, but most of the lowland is now cultivated and supports imported food trees, such as breadfruit and mango, and floral species of great brilliance—royal poinciana (flamboyant tree), frangipani, bougainvillea, and hibiscus. Swamps, mangroves, and marsh woodlands occur along the coasts, but arid environments are characterized by thorn scrub, columnar cactus, or sparse, low forest. At heights above 1,000 feet the trees form a closed canopy and are covered with epiphytes, bromeliads, and tree ferns; in some places above 5,000 feet this mist forest gives way to elfin woodland and alpine grass.

West Indian fauna is an impoverished version of the fauna in the place from which much of it was derived—South America. Only Trinidad, adjacent to Venezuela, has a wide range of mammals—sloths, anteaters, raccoons, howler monkeys, and bats. Reptiles are more diverse because of their ability to cross water, and the islands generally abound with lizards, caimans, crocodiles, and snakes. Above all, the West Indies display a rich variety of native birds—hummingbirds, pigeons, doves, and parrots being ubiquitous, with flamingos, ibis, and mockingbirds having a more restricted distribution.

In prehistoric times the most abundant faunal habitats were the forest canopies—with varieties of birds, bats, and insects—and the sea margins, which were rich in turtles, clams, oysters, caimans, manatees, dolphins, red snapper, bonito, and flying fish. These are still the main habitats of the species, but environmental destruction by humans or human-introduced predators, such as the African green monkey, the mongoose, and some rat species, has led to rapid species extinction, especially in the 20th century. Some nonextant species include the herbivorous tortoise, the rock iguana, and the giant barn owl.

Settlement patterns. West Indian islands have been intensively settled since the 17th century, when sugar plantations were introduced to the British and French colonies. The 19th century witnessed slave emancipation and the creation of free peasantries, while the 20th century has been characterized by shifts in population to the cities and towns. The capitals of the Greater and Lesser Antilles are all major ports and the sites of much of the urban migration that has developed since the mid-20th century. Havana, Cuba, and Santo Domingo, Dom. Rep., are cities of more than two million and 1.8 million, respectively. San Juan, P.R., and Port-au-Prince, Haiti, have more than 400,000 inhabitants each. Other relatively large cities of the region include Willemstad, Netherlands Antilles, Kingston, Jam., and Fort-de-France, Martinique.

Urban centres

THE PEOPLE

Ethnic composition. The numbers of white and black West Indians are about equal, and only a slightly smaller percentage is mulatto, who form a distinct colour category. Many West Indian countries have large black majorities, but whites account for some 70 percent of the population of Cuba and Puerto Rico, which between them have more than 40 percent of the total West Indian population. Elsewhere, except for the Dominican Republic, where mulattoes form about three-fourths of the population, and some small dependencies, blacks are the largest single element but not necessarily the majority. Haiti and Jamaica have the largest numbers of black population.

Race, colour, culture, ethnicity, and class are intertwined in complex ways in Caribbean societies, but they may be categorized into four different social types: (1) folk societies that are characteristic of small (less than 20,000 population) islands or combinations of islands, lacking class stratification and ethnic diversity, and containing one or more colour groupings, (2) stratified societies that exist where there is a close association between colour, class, and culture, as in Jamaica, Haiti, and the Windward and Leeward Islands, all of which were slave societies in the 17th and 18th centuries, (3) societies in which race salience is of prime importance, as in Trinidad and Tobago, where the black, brown, and white stratification has been confronted by a large, segmented East Indian population, and (4) class-stratified societies that have non-black majorities—such as predominantly white Cuba and Puerto Rico and the predominantly mulatto Dominican Republic, which developed as sugar producers in the 19th century and where blacks and black culture are now confined largely to enclaves in the stratification.

Social categories

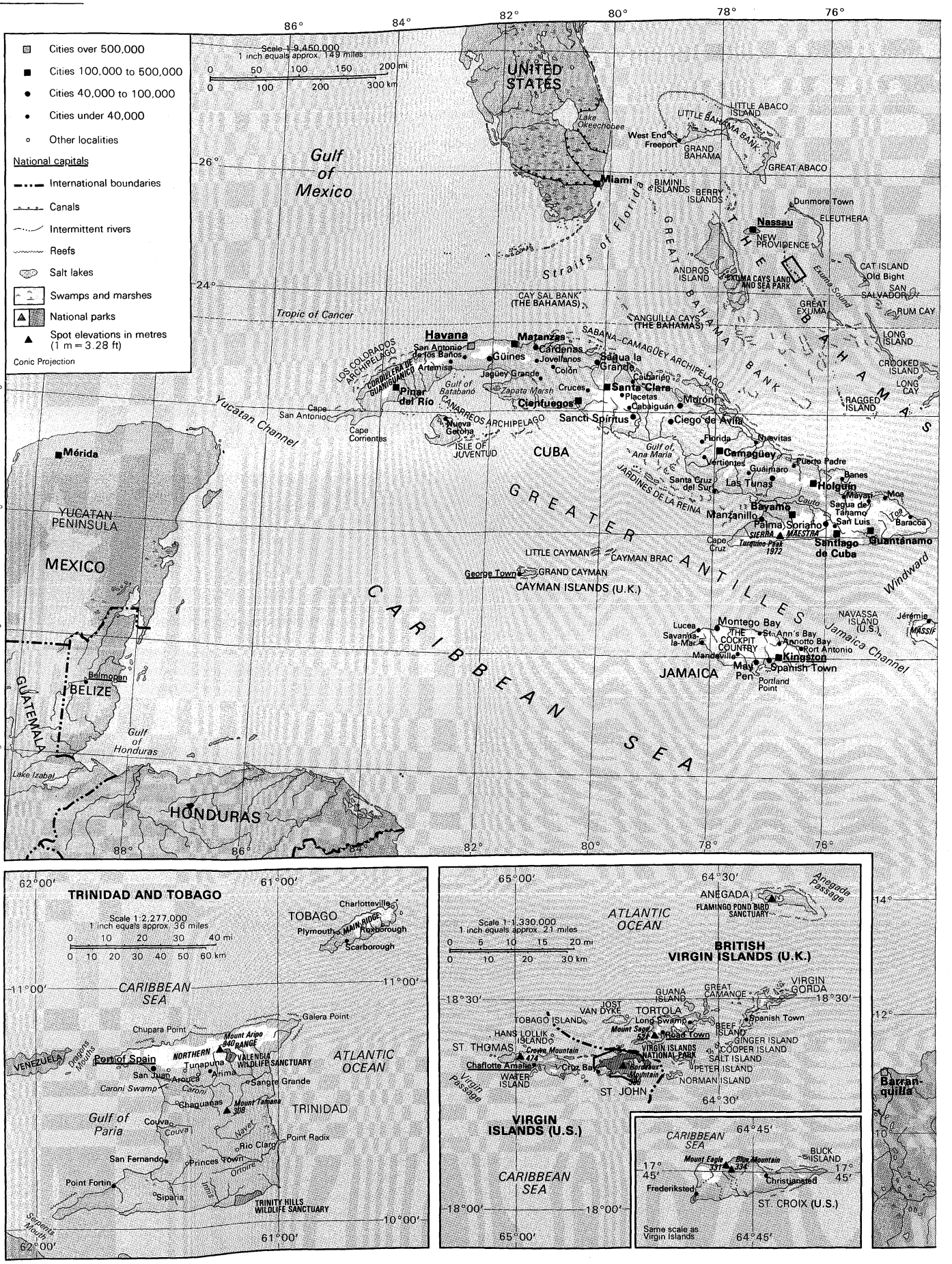
Linguistic composition. The West Indies are linguistically diverse. In addition to the languages of the colonial powers—English, Spanish, French, and Dutch—there are a number of tongues spoken by non-European immigrants—Hindī, Urdū, and Chinese. Also, there are pidgin variants of the European languages known as creoles, which have become the common languages of many of the people. English and French creoles are the most widespread and are best developed in Jamaica and Haiti, respectively. Spanish creoles did not evolve in the major Spanish-language communities—Cuba, Puerto Rico, and the Dominican Republic. A variant called Papiamento, a Spanish-Dutch-Portuguese-English creole, is widely spoken in Aruba and the Netherlands Antilles. There is a marked correlation between whiteness and the use of standard language forms, though with a distinctive Caribbean accent, and blackness and creole usage: the association holds island by island and regionally.

Hindī, Urdū, and Chinese are languages of the home, and the younger generation are often no longer fluent in their use. In contrast, English creole has increased in significance in the Commonwealth Caribbean since decolonization; it is used extensively on radio broadcasts in Jamaica, and modern Jamaican novels, poems, and plays are often solely in English creole or move between it and standard English as the mood demands. A similar situation has existed for decades in Haiti, where the vast majority use French creole and neither read, write, nor understand standard French. The political incorporation of the French Antilles (Guadeloupe and Martinique) as overseas *départements* of France has resulted in the proliferation of standard French in those communities, while in Puerto Rico, Spanish is influenced by American English and by the so-called Spanglish (American English-Spanish) of returning migrants.

English and French creole

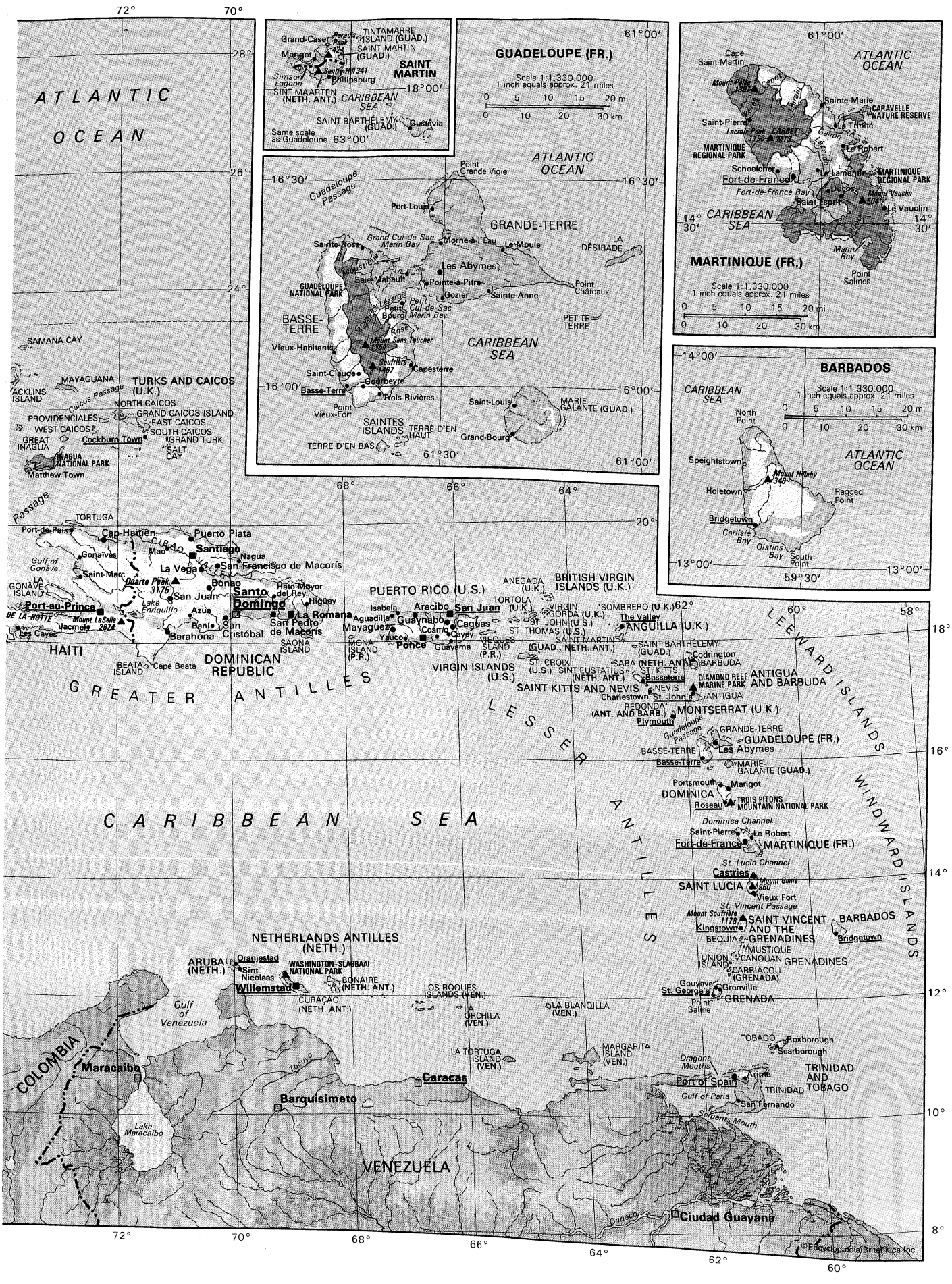
Religion. Religious affiliation in the West Indies follows a pattern similar to that of language. Roman Catholicism is the predominant faith in the Spanish- and French-speaking islands, while Protestantism is the norm in the Commonwealth Caribbean and the Dutch territories. Among the masses, however, syncretic cults, a mixture of either Catholicism or Protestantism and African elements, prevail, notably in Haiti, Jamaica, and Trinidad and Tobago, though elements of the Afro-Christian tradition also occur in Cuba, Puerto Rico, and the Lesser Antilles.

In Trinidad and Tobago about three-fourths of the East Indian group, which accounts for some 40 percent of the population, has retained Hinduism and Islām, and, while the first half of the 20th century was characterized by Indian conversion to Christianity, the second half has witnessed strong revivals in Asian religions. In part this has been due to the political development of the Indian population; but it has also been the consequence of improved education and the emergence of sophisticated


Cities over 500,000
Cities 100,000 to 500,000
Cities 40,000 to 100,000
Cities under 40,000
Other localities
National capitals
International boundaries
Canals
Intermittent rivers
Reefs
Salt lakes
Swamps and marshes
National parks
Spot elevations in metres (1 m = 3.28 ft)
Conic Projection
Scale 1:9,450,000
1 inch equals approx. 149 miles
UNITED STATES
Gulf of Mexico
Lake Okeechobee
Miami
Straits of Florida
LITTLE ABACO ISLAND
LITTLE BAHAMA BANK
West End
Freeport
GRAND BAHAMA
GREAT ABACO
BIMINI ISLANDS
BERRY ISLANDS
Dunmore Town
ELEUTHERA
Nassau
NEW PROVIDENCE
THE BAHAMAS
GREAT BAHAMA BANK
ANDROS ISLAND
EXUMA CAYS LAND AND SEA PARK
Exuma Sound
CAT ISLAND
Old Bight
SAN SALVADOR
RUM CAY
LONG ISLAND
CROOKED ISLAND
LONG CAY
RAGGED ISLAND
CAY SAL BANK (THE BAHAMAS)
ANGUILLA CAYS (THE BAHAMAS)
GREAT EXUMA
Tropic of Cancer
SABANA-CAMAGÜEY ARCHIPELAGO
LOS COLORADOS ARCHIPELAGO
CORDILLERA DE GUANIGUANICO
Havana
San Antonio de los Baños
Artemisa
Matanzas
Cárdenas
Güines
Jovellanos
Jagüey Grande
Colón
Sagua la Grande
Caibarién
Gulf of Batabanó
Zapata Marsh
Cruces
Santa Clara
Placetas
Morón
Pinar del Río
Cienfuegos
Cabaiguán
Cape San Antonio
CANARREOS ARCHIPELAGO
Nueva Gerona
Cape Corrientes
ISLE OF JUVENTUD
Sancti Spíritus
Ciego de Avila
Florida
Nuevitas
Gulf of Ana María
CUBA
Camagüey
Vertientes
Guáimaro
Puerto Padre
Banes
Santa Cruz del Sur
Las Tunas
Holguín
Mayarí
Moa
Sagua de Tánamo
Baracoa
JARDINES DE LA REINA
Bayamo
Manzanillo
Cauto
Palma Soriano
San Luis
SIERRA MAESTRA
Cape Cruz
Turquino Peak 1972
Santiago de Cuba
Guantánamo
Yucatan Channel
Mérida
YUCATAN PENINSULA
MEXICO
GREATER ANTILLES
CARIBBEAN SEA
LITTLE CAYMAN
CAYMAN BRAC
George Town
GRAND CAYMAN
CAYMAN ISLANDS (U.K.)
Windward Passage
NAVASSA ISLAND (U.S.)
Jérémie
MASSIF
Montego Bay
Luces
Savanna-la-Mar
THE COCKPIT COUNTRY
St. Ann's Bay
Ocho Rios
Port Antonio
Mandeville
Kingston
Spanish Town
May Pen
Portland Point
JAMAICA
Jamaica Channel
Belmopan
BELIZE
GUATEMALA
Gulf of Honduras
Lake Izabal
HONDURAS
88°
86°
84°
82°
80°
78°
76°
26°
24°
22°
20°
18°
16°
14°
12°
10°

TRINIDAD AND TOBAGO
Scale 1:2,277,000
1 inch equals approx. 36 miles
TOBAGO
Charlotteville
Plymouth
MAIN RIDGE
Roxborough
Scarborough
CARIBBEAN SEA
ATLANTIC OCEAN
Chupara Point
Galera Point
Mount Aripo 940
NORTHERN RANGE
VENEZUELA
Dragons Mouths
Port of Spain
Tunapuna
VALENCIA WILDLIFE SANCTUARY
Arima
San Juan
Arouca
Sangre Grande
Caroni Swamp
Caroni
Chaguanas
Mount Tamana 308
TRINIDAD
Gulf of Paria
Couva
Rio Claro
Point Radix
San Fernando
Princes Town
Ortoire
Point Fortin
Siparia
Nariva
TRINITY HILLS WILDLIFE SANCTUARY
Serpents Mouths
62°00′
61°00′
11°00′
10°00′

Scale 1:1,330,000
1 inch equals approx. 21 miles
ATLANTIC OCEAN
ANEGADA
Anegada Passage
FLAMINGO POND BIRD SANCTUARY
BRITISH VIRGIN ISLANDS (U.K.)
GUANA ISLAND
GREAT CAMANOE
VIRGIN GORDA
JOST VAN DYKE
TORTOLA
Spanish Town
TOBAGO ISLANDS
Long Swamp
BEEF ISLAND
Mount Sage 542
Road Town
HANS LOLLIK ISLAND
ST THOMAS
Crown Mountain 474
GINGER ISLAND
VIRGIN ISLANDS NATIONAL PARK
COOPER ISLAND
SALT ISLAND
Charlotte Amalie
Cruz Bay
Mountain 398
PETER ISLAND
WATER ISLAND
NORMAN ISLAND
Virgin Passage
ST JOHN
VIRGIN ISLANDS (U.S.)
CARIBBEAN SEA
64°30′
65°00′
18°30′
18°00′
CARIBBEAN SEA
64°45′
Mount Eagle 331
Blue Mountain 334
BUCK ISLAND
Christiansted
Frederiksted
ST. CROIX (U.S.)
17°45′
Same scale as Virgin Islands
Barranquilla
14°
12°
10°



© Encyclopædia Britannica, Inc.

MAP INDEX

Political divisions

Anguilla (U.K.) . . . . 18 13 N 63 03 W
Antigua and Barbuda . . . . . 17 03 N 61 48 W
Aruba (Neth.) . . . . 12 30 N 69 58 W
Bahamas, The . . . . 24 15 N 76 00 W
Barbados . . . . . 13 10 N 59 32 W
British Virgin Islands (U.K.) . . . . 18 30 N 64 30 W
Cayman Islands (U.K.) . . . . . 19 30 N 80 40 W
Cuba . . . . . 21 30 N 80 00 W
Dominica . . . . . 15 30 N 61 20 W
Dominican Republic . . . . . 19 00 N 70 40 W
Grenada . . . . . 12 07 N 61 40 W
Guadeloupe (Fr.) . . 16 15 N 61 35 W
Haiti . . . . . 19 00 N 72 25 W
Jamaica . . . . . 18 15 N 77 30 W
Martinique (Fr.) . . . 14 40 N 61 00 W
Montserrat (U.K.) . . 16 45 N 62 12 W
Netherlands Antilles (Neth.) . . . 12 10 N 69 00 W
Puerto Rico (U.S.) . . . . . 18 15 N 66 30 W
Saint Kitts and Nevis . . . . . 17 20 N 62 45 W
Saint Lucia . . . . . 13 53 N 60 58 W
Saint Vincent and the Grenadines . . . . . 13 05 N 61 12 W
Trinidad and Tobago . . . . . 11 00 N 61 00 W
Turks and Caicos Islands (U.K.) . . . . 21 44 N 71 35 W
Virgin Islands (U.S.) . . . . . 18 20 N 64 50 W

Cities and towns

Aguadilla . . . . . 18 26 N 67 09 W
Annotto Bay . . . . . 18 16 N 76 46 W
Arecibo . . . . . 18 28 N 66 43 W
Arima . . . . . 10 38 N 61 17 W
Arouca . . . . . 10 38 N 61 20 W
Artemisa . . . . . 22 49 N 82 46 W
Aux Cayes, see Les Cayes
Azua . . . . . 18 27 N 70 44 W
Baie-Mahault . . . . . 16 16 N 61 35 W
Banes . . . . . 20 58 N 75 43 W
Baní . . . . . 18 17 N 70 20 W
Baracoa . . . . . 20 21 N 74 30 W
Barahona . . . . . 18 12 N 71 06 W
Basse-Terre . . . . . 16 00 N 61 44 W
Basseterre . . . . . 17 18 N 62 43 W
Bayamo . . . . . 20 23 N 76 39 W
Bonao . . . . . 18 56 N 70 25 W
Bridgetown . . . . . 13 06 N 59 37 W
Cabaiguán . . . . . 22 05 N 79 30 W
Caguas . . . . . 18 14 N 66 02 W
Caibarién . . . . . 22 31 N 79 28 W
Camagüey . . . . . 21 23 N 77 55 W
Cap-Haïtien . . . . . 19 45 N 72 12 W
Capesterre . . . . . 16 02 N 61 33 W
Cárdenas . . . . . 23 02 N 81 12 W
Castries . . . . . 14 01 N 61 00 W
Cayey . . . . . 18 07 N 66 10 W
Chaguanas . . . . . 10 31 N 61 25 W
Charlestown . . . . . 17 08 N 62 37 W
Charlotte Amalie . . 18 20 N 64 56 W
Charlotteville . . . . . 11 19 N 60 33 W
Christiansted . . . . . 17 45 N 64 42 W
Ciego de Ávila . . . 21 51 N 78 46 W
Cienfuegos . . . . . 22 09 N 80 27 W
Coamo . . . . . 18 05 N 66 22 W
Cockburn Town . . . 21 29 N 71 30 W
Codrington . . . . . 17 38 N 61 50 W
Colón . . . . . 22 43 N 80 54 W
Couva . . . . . 10 25 N 61 27 W
Cruces . . . . . 22 21 N 80 16 W
Cruz Bay . . . . . 18 20 N 64 48 W
Ducos . . . . . 14 34 N 60 58 W
Dunmore Town . . . 25 30 N 76 39 W
Florida . . . . . 21 32 N 78 14 W
Fort-de-France . . . 14 36 N 61 05 W
Frederiksted . . . . . 17 43 N 64 53 W
Freeport . . . . . 26 32 N 78 42 W
George Town . . . . 19 18 N 81 23 W
Gonaïves . . . . . 19 27 N 72 41 W
Gourbeyre . . . . . 16 00 N 61 42 W
Gouyave . . . . . 12 10 N 61 44 W
Gozier . . . . . 16 12 N 61 30 W
Grand-Bourg . . . . . 15 53 N 61 19 W
Grand-Case . . . . . 18 06 N 63 04 W
Grenville . . . . . 12 07 N 61 37 W
Guáimaro . . . . . 21 03 N 77 21 W
Guantánamo . . . . . 20 08 N 75 12 W
Guayama . . . . . 17 59 N 66 07 W
Guaynabo . . . . . 18 22 N 66 07 W
Güines . . . . . 22 50 N 82 02 W
Gustavia . . . . . 17 54 N 62 52 W
Hato Mayor del Rey . . . . . 18 46 N 69 15 W
Havana (La Habana) . . . . 23 08 N 82 22 W
Higüey . . . . . 18 37 N 68 42 W
Holetown . . . . . 13 11 N 59 39 W
Holguín . . . . . 20 53 N 76 15 W
Isabela . . . . . 18 30 N 67 01 W
Jacmel . . . . . 18 14 N 72 32 W
Jagüey Grande . . . 22 32 N 81 08 W
Jérémie . . . . . 18 39 N 74 07 W
Jovellanos . . . . . 22 48 N 81 12 W
Julia Molina, see Nagua
Kingston . . . . . 17 58 N 76 48 W
Kingstown . . . . . 13 09 N 61 14 W
La Habana, see Havana
La Romana . . . . . 18 25 N 68 58 W
La Trinité . . . . . 14 44 N 60 58 W
La Vega . . . . . 19 13 N 70 31 W
Las Tunas . . . . . 20 58 N 76 57 W
Le Lamentin . . . . . 14 37 N 61 01 W
Le Moule . . . . . 16 20 N 61 21 W
Le Robert . . . . . 14 41 N 60 57 W
Le Vauclin . . . . . 14 33 N 60 51 W
Les Abymes . . . . . 16 16 N 61 31 W
Les Cayes (Aux Cayes) . . . . . 18 12 N 73 45 W
Long Swamp . . . . . 18 25 N 65 37 W
Lucea . . . . . 18 27 N 78 10 W
Mandeville . . . . . 18 02 N 77 30 W
Manzanillo . . . . . 20 21 N 77 07 W
Mao . . . . . 19 34 N 71 05 W
Marigot . . . . . 15 32 N 61 18 W
Marigot . . . . . 18 04 N 63 06 W
Matanzas . . . . . 23 03 N 81 35 W
Matthew Town . . . 20 57 N 73 40 W
May Pen . . . . . 17 58 N 77 14 W
Mayagüez . . . . . 18 12 N 67 09 W
Mayarí . . . . . 20 40 N 75 41 W
Moa . . . . . 20 40 N 74 56 W
Montego Bay . . . . 18 28 N 77 55 W
Morne-à-l'Eau . . . . 16 21 N 61 31 W
Morón . . . . . 22 06 N 78 38 W
Nagua (Julia Molina) . . . . . 19 23 N 69 50 W
Nassau . . . . . 25 05 N 77 21 W
Nueva Gerona . . . . 21 53 N 82 48 W
Nuevitas . . . . . 21 33 N 77 16 W
Old Bight . . . . . 24 15 N 75 21 W
Oranjestad . . . . . 12 31 N 70 02 W
Palma Soriano . . . . 20 13 N 76 00 W
Petit-Bourg . . . . . 16 12 N 61 36 W
Philipsburg . . . . . 18 01 N 63 04 W
Pinar del Río . . . . . 22 25 N 83 42 W
Placetas . . . . . 22 19 N 79 40 W
Plymouth . . . . . 11 13 N 60 47 W
Plymouth . . . . . 16 42 N 62 13 W
Point Fortin . . . . . 10 11 N 61 41 W
Pointe-à-Pitre . . . . . 16 14 N 61 32 W
Ponce . . . . . 18 01 N 66 37 W
Port Antonio . . . . . 18 11 N 76 28 W
Port-au-Prince . . . . 18 32 N 72 20 W
Port-de-Paix . . . . . 19 57 N 72 50 W
Port-Louis . . . . . 16 25 N 61 32 W
Port of Spain . . . . . 10 39 N 61 31 W
Portsmouth . . . . . 15 35 N 61 28 W
Princes Town . . . . 10 16 N 61 23 W
Puerto Padre . . . . . 21 12 N 76 36 W
Puerto Plata . . . . . 19 48 N 70 41 W
Rio Claro . . . . . 10 18 N 61 11 W
Road Town . . . . . 18 27 N 64 37 W
Roseau . . . . . 15 18 N 61 24 W
Roxborough . . . . . 11 15 N 60 35 W
Sagua de Tánamo . . . . . 20 35 N 75 14 W
Sagua la Grande . . 22 49 N 80 05 W
St. Ann's Bay . . . . 18 26 N 77 08 W
Saint-Claude . . . . . 16 02 N 61 42 W
Saint-Esprit . . . . . 14 34 N 60 57 W
St. George's . . . . . 12 03 N 61 45 W
St. John's . . . . . 17 06 N 61 51 W
Saint-Louis . . . . . 15 57 N 61 19 W
Saint-Marc . . . . . 19 07 N 72 42 W
Saint-Pierre . . . . . 14 45 N 61 11 W
Sainte-Anne . . . . . 16 14 N 61 23 W
Sainte-Marie . . . . . 14 47 N 61 00 W
Sainte-Rose . . . . . 16 20 N 61 42 W
San Antonio de los Baños . . . . . 22 53 N 82 30 W
San Cristóbal . . . . 18 25 N 70 06 W
San Fernando . . . . 10 17 N 61 28 W
San Francisco de Macorís . . . . . 19 18 N 70 15 W
San Juan . . . . . 10 39 N 61 27 W
San Juan . . . . . 18 28 N 66 07 W
San Juan . . . . . 18 48 N 71 14 W
San Luis . . . . . 20 12 N 75 51 W
San Pedro de Macorís . . . . . 18 27 N 69 18 W
Sancti Spíritus . . . . 21 56 N 79 27 W
Sangre Grande . . . 10 35 N 61 07 W
Santa Clara . . . . . 22 24 N 79 58 W
Santa Cruz del Sur . . . . . 20 43 N 78 00 W
Santiago . . . . . 19 27 N 70 42 W
Santiago de Cuba . . . . . 20 01 N 75 49 W
Santo Domingo . . 18 28 N 69 54 W
Savanna-la-Mar . . . 18 13 N 78 08 W
Scarborough . . . . . 11 11 N 60 44 W
Schoelcher . . . . . 14 37 N 61 06 W
Sint Nicolaas . . . . . 12 26 N 69 55 W
Siparia . . . . . 10 08 N 61 30 W
Spanish Town . . . . 17 59 N 76 57 W
Spanish Town . . . . 18 27 N 64 26 W
Speightstown . . . . 13 15 N 59 39 W
Trois-Rivières . . . . 15 59 N 61 39 W
Tunapuna . . . . . 10 38 N 61 23 W
Valley, The . . . . . 18 03 N 63 04 W
Vertientes . . . . . 21 16 N 78 09 W
Vieux Fort . . . . . 13 44 N 60 57 W
Vieux-Habitants . . . 16 04 N 61 46 W
West End . . . . . 26 41 N 78 58 W
Willemstad . . . . . 12 07 N 68 57 W
Yauco . . . . . 18 02 N 66 51 W

Physical features and points of interest

Acklins Island . . . . 22 26 N 73 58 W
Ana María, Gulf of . . . . . 21 25 N 78 40 W
Andros Island . . . . 24 26 N 77 57 W
Anegada, *island* . . 18 45 N 64 20 W
Anegada Passage . . . . . 18 30 N 63 40 W
Anguilla Cays . . . . 23 31 N 79 33 W
Antigua, *island* . . . 17 03 N 61 48 W
Aripo, Mount . . . . . 10 43 N 61 15 W
Atlantic Ocean . . . 28 00 N 76 00 W
Barbuda, *island* . . . 17 38 N 61 48 W
Basse-Terre, *island* . . . . . 16 10 N 61 40 W
Batabanó, Gulf of . . . . . 22 15 N 82 30 W
Beata, Cape . . . . . 17 36 N 71 25 W
Beata Island . . . . . 17 35 N 71 31 W
Beef Island . . . . . 18 27 N 64 31 W
Bequia, *island* . . . . 13 01 N 61 13 W
Berry Islands . . . . . 25 34 N 77 45 W
Bimini Islands . . . . 25 44 N 79 15 W
Blue Mountain . . . . 17 45 N 64 48 W
Bonaire, *island* . . . 12 12 N 68 15 W
Bordeaux Mountain . . . . . 18 20 N 64 45 W
Buck Island . . . . . 17 47 N 64 37 W
Caicos Passage, *channel* . . . . . 22 00 N 72 30 W
Canarreos Archipelago . . . . . 21 50 N 82 30 W
Canouan, *island* . . 12 43 N 61 20 W
Capot, *river* . . . . . 14 51 N 61 05 W
Caravelle Nature Reserve . . . . . 14 45 N 60 55 W
Carbet Mountains . . . . . 14 42 N 61 07 W
Caribbean Sea . . . 15 00 N 76 00 W
Carlisle Bay . . . . . 13 05 N 59 37 W
Caroni, *river* . . . . . 10 37 N 61 30 W
Caroni Swamp . . . 10 36 N 61 27 W
Carriacou, *island* . . 12 30 N 61 27 W
Cat Island . . . . . 24 27 N 75 30 W
Cauto, *river* . . . . . 20 33 N 77 14 W
Cay Sal Bank, *reef* . . . . . 23 45 N 80 00 W
Cayman Brac, *island* . . . . . 19 43 N 79 49 W
Châteaux, Point . . 16 15 N 61 11 W
Chupara Point . . . . 10 48 N 61 22 W
Cibao Valley . . . . . 19 27 N 70 36 W
Cockpit Country, The, *region* . . . . . 18 18 N 77 43 W
Cooper Island . . . . 18 22 N 64 30 W
Corrientes, Cape . . 21 45 N 84 31 W
Couva, *river* . . . . . 10 25 N 61 30 W
Crooked Island . . . 22 45 N 74 13 W
Crown Mountain . . 18 21 N 64 58 W
Cruz, Cape . . . . . 19 51 N 77 44 W
Curaçao, *island* . . . 12 10 N 69 00 W
Diamond Reef Marine Park . . . . . 17 15 N 61 50 W
Dominica Channel . . . . . 15 10 N 61 15 W
Dragons Mouths, *channel* . . . . . 10 45 N 61 46 W
Duarte Peak . . . . . 19 02 N 70 59 W
Eagle, Mount . . . . . 17 44 N 64 48 W
East Caicos, *island* . . . . . 21 41 N 71 30 W
Eleuthera, *island* . . 25 10 N 76 14 W
Enriquillo, Lake . . . 18 27 N 71 39 W
Exuma Cays Land and Sea Park . . . . . 24 35 N 76 40 W
Exuma Sound . . . . 24 15 N 76 00 W
Flamingo Pond Bird Sanctuary . . 18 44 N 64 23 W
Florida, Straits of . . 25 00 N 79 45 W
Fort-de-France Bay . . . . . 14 34 N 61 04 W
Galera Point . . . . . 10 49 N 60 55 W
Galion, *river* . . . . . 14 44 N 60 57 W
Gimie, Mount . . . . 13 52 N 61 01 W
Ginger Island . . . . 18 24 N 64 28 W
Gonâve, Gulf of . . . 19 00 N 73 30 W
Goyaves, *river* . . . . 16 18 N 61 36 W
Grand Bahama, *island* . . . . . 26 38 N 78 25 W
Grand Caicos, *island* . . . . . 21 47 N 71 43 W
Grand Cayman, *island* . . . . . 19 20 N 81 15 W
Grand Cul-de-Sac Marin Bay . . . . . 16 20 N 61 35 W
Grand Turk, *island* . . . . . 21 29 N 71 07 W
Grande-Terre, *island* . . . . . 16 20 N 61 25 W
Grande Vigie, Point . . . . . 16 31 N 61 28 W
Great Abaco, *island* . . . . . 26 28 N 77 05 W
Great Bahama Bank . . . . . 23 15 N 78 00 W
Great Camanoe, *island* . . . . . 18 29 N 64 32 W
Great Exuma, *island* . . . . . 23 32 N 75 50 W
Great Inagua, *island* . . . . . 21 05 N 73 18 W
Great Tobago, see Tobago
Greater Antilles, *islands* . . . . . 20 00 N 74 00 W
Grenadines, *islands* . . . . . 12 40 N 61 15 W
Guadeloupe National Park . . . . 16 20 N 61 40 W
Guadeloupe Passage . . . . . 16 45 N 61 30 W
Guana Island . . . . . 18 29 N 64 34 W
Guaniguanico, Cordillera de . . . . 22 35 N 83 50 W
Hans Lollik Island . 18 24 N 64 54 W
Hillaby, Mount . . . . 13 12 N 59 35 W
Hotte (Sud), Massif de la . . . . 18 25 N 73 55 W
Inagua National Park . . . . . 21 05 N 73 20 W
Inniss, *river* . . . . . 10 05 N 61 15 W
Jamaica Channel . . 18 00 N 75 30 W
Jardines de la Reina, *islands* . . . 20 50 N 78 55 W
Jost Van Dyke, *island* . . . . . 18 28 N 64 45 W
Juventud (Pinos), Isle of . . . . . 21 40 N 82 50 W
La Désirade, *island* . . . . . 16 19 N 61 03 W
La Gonâve Island . . . . . 18 51 N 73 03 W
La Selle, Mount . . . 18 22 N 71 59 W

La Tortue, see Tortuga
Lacroix Peak ..... 14 42 N 61 07 W
Leeward Islands .. 17 00 N 63 00 W
Lesser Antilles, *islands* ..... 15 00 N 61 00 W
Lézarde, *river* .... 14 36 N 61 01 W
Lézarde, *river* .... 16 13 N 61 35 W
Little Abaco Island ..... 26 53 N 77 43 W
Little Bahama Bank, *reef* ..... 26 55 N 78 40 W
Little Cayman, *island* ..... 19 41 N 80 03 W
Long Cay, *island* ..... 22 37 N 74 20 W
Long Island ..... 23 15 N 75 07 W
Lorrain, *river* ..... 14 50 N 61 03 W
Los Colorados Archipelago ..... 22 36 N 84 20 W
Maestra, Sierra ... 20 00 N 76 45 W
Main Ridge ..... 11 16 N 60 38 W
Marie-Galante, *island* ..... 15 56 N 61 16 W
Marin Bay ..... 14 27 N 60 53 W
Martinique Regional Park ... 14 45 N 61 07 W
Mayaguana, *island* ..... 22 23 N 72 57 W
Mexico, Gulf of ... 26 00 N 86 00 W
Mona Island ..... 18 05 N 67 54 W
Moustique, *river* .. 16 19 N 61 41 W
Mustique, *island* .. 12 53 N 61 11 W
Navassa Island ... 18 24 N 75 01 W
Navet, *river* ..... 10 24 N 61 05 W
Nevis, *island* ..... 17 10 N 62 34 W
New Providence, *island* ..... 25 02 N 77 24 W
Norman Island .... 18 20 N 64 37 W
North Caicos, *island* ..... 21 56 N 71 59 W
North Point ..... 13 20 N 59 37 W
Northern Range ... 10 44 N 61 15 W
Oistins Bay ..... 13 03 N 59 33 W
Ortoire, *river* ..... 10 20 N 61 00 W
Paradis Peak ..... 18 04 N 63 04 W
Paria, Gulf of ..... 10 20 N 62 00 W
Pelée, Mount ..... 14 48 N 61 10 W
Pères, *river* ..... 14 45 N 61 11 W
Peter Island ..... 18 22 N 64 35 W
Petit Cul-de-Sac Marin Bay ..... 16 12 N 61 33 W
Petite Terre, *islands* ..... 16 10 N 61 07 W
Pilote, *river* ..... 14 28 N 60 55 W
Pinos, see Juventud, Isle of
Point Radix, *peninsula* ..... 10 20 N 60 59 W
Portland Point .... 17 42 N 77 11 W
Providenciales, *island* ..... 21 47 N 72 17 W
Ragged Island .... 22 12 N 75 44 W
Ragged Point .... 13 10 N 59 25 W
Redonda, *island* ..... 16 55 N 62 19 W
Rose, *river* ..... 16 09 N 61 35 W
Rum Cay, *island* .. 23 40 N 74 53 W
Saba, *island* ..... 17 38 N 63 14 W
Sabana-Camagüey, Archipelago .... 22 30 N 79 00 W
Sage, Mount ..... 18 25 N 64 39 W
Saint-Barthélemy, *island* ..... 17 54 N 62 50 W
St. Christopher, see St. Kitts
St. Croix, *island* .. 17 44 N 64 44 W
Saint Eustatius, see Sint Eustatius
St. John, *island* ... 18 20 N 64 45 W
St. Kitts (St. Christopher), *island* ..... 17 20 N 62 45 W
St. Lucia Channel ..... 14 09 N 60 57 W
Saint Martin (Saint-Martin, Sint Maarten), *island* ..... 18 04 N 63 04 W
Saint-Martin, Cape ..... 14 52 N 61 13 W
St. Thomas, *island* ..... 18 21 N 64 56 W
St. Vincent Passage, *channel* ..... 13 30 N 61 00 W
Saintes Islands ... 15 52 N 61 37 W
Saline, Point ..... 12 00 N 61 48 W
Salines, Point .... 14 24 N 60 53 W
Salt Cay, *island* ... 21 20 N 71 11 W
Salt Island ..... 18 23 N 64 31 W
Samana Cay ..... 23 06 N 73 42 W
San Antonio, Cape ..... 21 52 N 84 57 W
San Salvador (Watling), *island* ..... 24 02 N 74 28 W
Sans Toucher, Mount ..... 16 06 N 61 41 W
Saona Island ..... 18 09 N 68 40 W
Sentry Hill ..... 18 03 N 63 05 W
Serpents Mouth, *channel* ..... 10 00 N 62 00 W
Simson (Simson Bay) Lagoon .... 18 03 N 63 07 W
Sint (Saint) Eustatius, *island* ..... 17 30 N 62 58 W
Sint Maarten, see Saint Martin
Sombrero, *island* .. 18 36 N 63 26 W
Soufrière, *volcano* ..... 16 03 N 61 40 W
Soufrière, Mount .. 13 20 N 61 11 W
South Caicos, *island* ..... 21 31 N 71 30 W
South Point ..... 13 02 N 59 31 W
Sud, see Hotte, Massif de la
Tamana, Mount ... 10 28 N 61 12 W
Terre d'en Bas, *island* ..... 15 52 N 61 38 W
Terre d'en Haut, *island* ..... 15 52 N 61 35 W
Tintamarre Island ..... 18 07 N 62 59 W
Toa, *river* ..... 20 23 N 74 32 W
Tobago, *island* ... 11 15 N 60 40 W
Tobago (Great Tobago) Island .. 18 27 N 64 48 W
Tortola, *island* .... 18 27 N 64 36 W
Tortuga (La Tortue), *island* ... 20 04 N 72 49 W
Trinidad, *island* ... 10 30 N 61 15 W
Trinity Hills Wildlife Sanctuary ..... 10 05 N 61 07 W
Trois Pitons Mountain National Park .... 15 19 N 61 18 W
Turquino Peak .... 19 59 N 76 50 W
Union Island ..... 12 36 N 61 26 W
Valencia Wildlife Sanctuary ..... 10 38 N 61 09 W
Vauclin, Mount ... 14 33 N 60 53 W
Vieques Island ... 18 08 N 65 25 W
Vieux-Fort, Point ..... 15 57 N 61 43 W
Virgin Gorda, *island* ..... 18 30 N 64 24 W
Virgin Islands National Park .... 18 20 N 64 45 W
Virgin Passage, *channel* ..... 18 20 N 65 10 W
Washington-Slagbaai National Park ... 12 17 N 68 23 W
Water Island ..... 18 19 N 64 57 W
Watling, see San Salvador
West Caicos, *island* ..... 21 39 N 72 28 W
Windward Islands ..... 13 00 N 61 00 W
Windward Passage, *channel* ..... 20 00 N 73 50 W
Yucatan Channel .. 21 45 N 85 45 W
Zapata Marsh .... 22 25 N 81 20 W

young Hindus and Muslims who have responded to the more philosophical aspects of their religions, as distinct from the folk tenets of their forebears.

Hinduism and Islām apart, religion tends to coincide closely with colour and class. Upper-class whites and browns (mulattoes) in Jamaica tend to be Anglican, lower-class blacks largely followers of Protestant faiths such as the Church of God and the Baptist church. Similarly, voodoo (voudou, or *vodun*) in Haiti is a folk religion, although its practitioners frequent the Roman Catholic church for the major rites of passage associated with baptism, marriage, and death. In the Commonwealth Caribbean, obia (obeah), which is similar to black magic, is still used against enemies or to enhance the user's position; other sects, such as believers in myalism, are devoted to removing spirits and healing the sick.

Demographic trends. The West Indies have experienced a common demographic history: decimation of the Amerindian population by the Spanish in the Greater Antilles, and by the British and French in the Lesser Antilles, led to the introduction of African slaves to cultivate the sugar and coffee plantations of the European settlers during the 17th and 18th centuries. After the abolition of slavery in the British, French, and Dutch territories in the 19th century, Asiatic indentured labour was introduced for plantation work—notably from northern India. Since World War I, however, unemployment has been growing in most of the islands, exacerbated by a high rate of natural increase.

Recent demographic trends in the region have followed closely, if not led, changes occurring in developing countries elsewhere in the world. Improvements in health care and hygiene since 1920 have rapidly reduced the mortality rate, and the fertility rate, although in a downward trend, continues to be relatively high. Infant mortality has been reduced to less than 20 per 1,000 live births in Cuba, Puerto Rico, Jamaica, and Trinidad and Tobago, and most of the other territories—Haiti being a notable exception—have achieved a similar reduction. Birth rates remain high in most of the states. The rate of natural increase, although still high, has shown a downward trend due in part to the availability of birth control—abortion (particularly in Cuba) and family planning. The total population growth, however, may be more seriously affected by opportunities for emigration. For example, between 20,000 and 25,000 Jamaicans a year have been settling in the United States and Canada since the 1970s. Much of this movement is legal and documented—educated West Indians have little difficulty in satisfying the skill criteria required by the U.S. and Canadian immigration authorities; and Puerto Ricans and French West Indians have virtually unrestricted access to the United States and France, respectively. In addition, there is a substantial undocumented migration, particularly from Haiti and the Dominican Republic, both within the West Indies and to external destinations, especially the United States.

Birth and death rates

The heavy emigration and high rates of natural increase notwithstanding, the dominance of large plantations and the limitation of small peasant plots, together with rural underemployment, have propelled West Indians into the towns, especially in the Greater Antilles and Trinidad, at a rapid rate. More than one-half of West Indians are now urban dwellers, but frequently under conditions of labour surplus and overcrowded housing that have led to the growth of shantytowns, an endemic condition throughout much of Latin America. Urbanization based on rapid natural increase and rural-to-urban migration has produced massive societally marginal populations, especially in the towns of Jamaica, the Dominican Republic, Haiti, and Puerto Rico.

THE ECONOMY

West Indian economies are largely underdeveloped and dependent upon foreign assistance, often from the former colonizing power. Although the sugar plantation remains dominant only in Cuba and Saint Kitts, its legacy is more widespread, as is reflected in the common heavy reliance on exports of primary agricultural and mined products of variable quality and in fluctuating quantities. The states are generally reliant upon sheltered markets in North

America and the European Economic Community to enhance their trade position, especially with regard to agricultural products. Economic diversification into industry, tourism, and mining, as a means of breaking from the single product economy, has produced only enclaves in most island nations, and modernization has been only partially achieved in most places and barely at all in others, such as Haiti. The reliance on one or a few products leaves most West Indian states highly vulnerable to fluctuations in market demand.

Government initiatives in some of the Greater Antilles and in such places as Trinidad and Tobago and Barbados have raised educational standards and improved health care services and communications, yet high unemployment and heavy reliance of urban labour on low-level jobs remain major economic problems. From time to time, governments have attempted to imitate Cuban socialism by nationalizing some sectors of the economy, as an alternative to U.S.-backed capitalism, operating in such places as Puerto Rico and the Dominican Republic. Despite such attempts, most West Indian states operate to one degree or another under a free enterprise system, although social programs are more heavily influenced by socialism. Cuba's centrally run economy is the main variant in the region.

Nationalization

Resources. While the West Indian islands were ideal for plantation agriculture during European colonial expansion in the 17th and 18th centuries, their subsequent development has been hampered by a lack of natural resources, especially the traditional mineral bases of heavy industry—coal and iron ore. During the 20th century, however, major petroleum, natural gas, and bauxite deposits have been identified and exploited. Petroleum and natural gas now provide the basis for Trinidad and Tobago's economy and are found in smaller but commercially viable quantities in Cuba and Barbados. Bauxite is mined by open-cast methods in Jamaica, Haiti, and the Dominican Republic; Jamaica was once the world's largest exporter and is still a major producer of both bauxite and refined alumina.

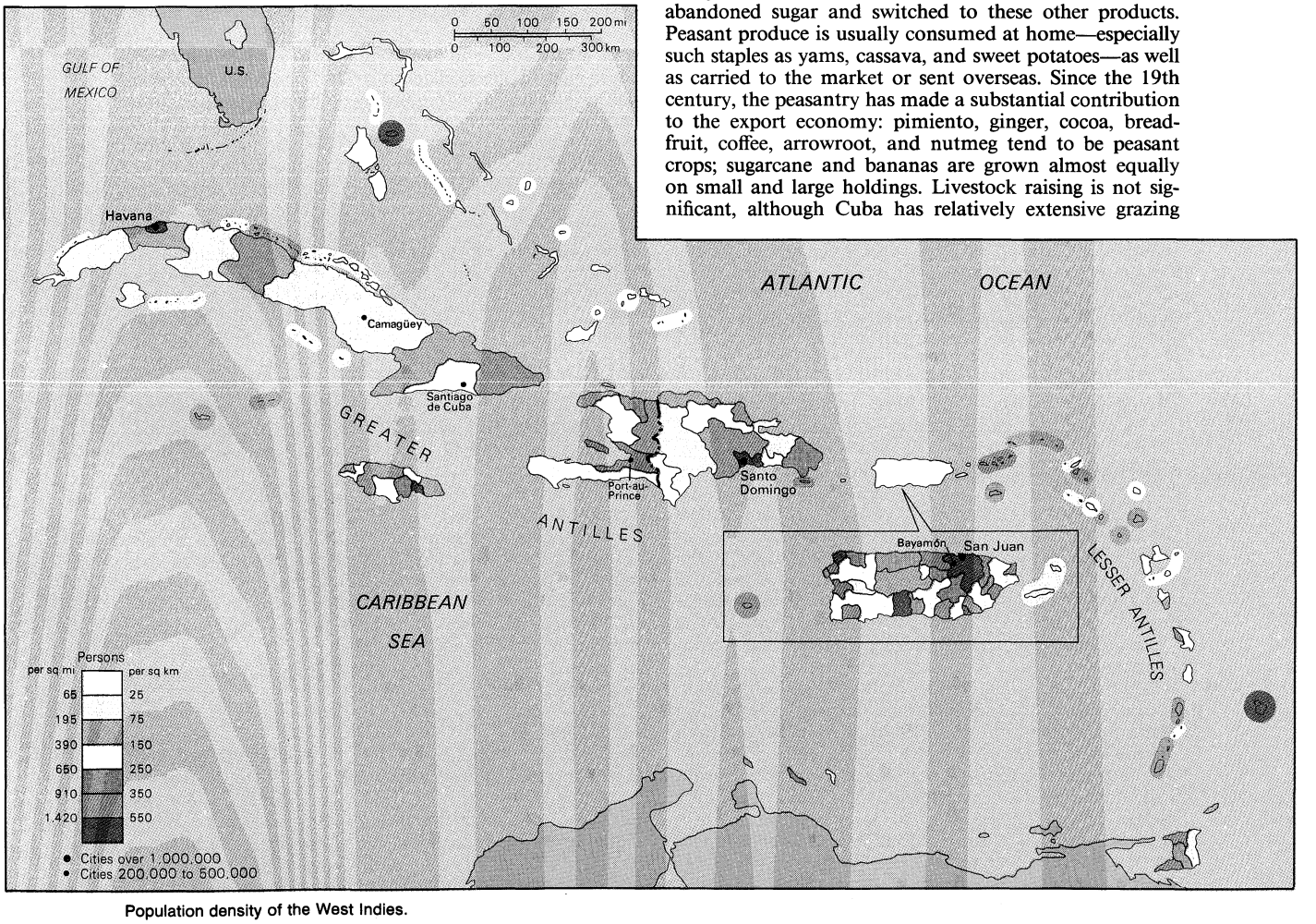
Other mineral resources are rather sparse, although Cuba is one of the main producers of nickel, and some asphalt is taken from Trinidad's Pitch Lake. Hydroelectric potential is slight, and timber and fishing resources are inadequate for large-scale commercial development except in Cuba. The climate, however, is a powerful resource that is increasingly exploited in the interests of tourism.

Agriculture. The process of modernization and industrialization has drawn West Indian populations away from agriculture, so that only in Haiti is most of the labour force devoted to farming. Nevertheless, agriculture remains dichotomized into peasant and plantation systems, located in the mountainous interior and on the fertile plains, respectively. Plantations are highly capitalized, employ labour, and are mechanized; many occupy land cultivated in sugarcane over the centuries, though now consolidated with central processing factories. Peasants, on the other hand, cultivate their own land, rely on hand labour, and make low capital investments. Often they possess no title to their land and engage in extended family ownership or sharecropping.

In the French- and English-speaking islands, where peasants are invariably black, they are the descendants of ex-slave ancestors who created the peasant farming system following emancipation in the first half of the 19th century. Throughout the West Indies, including Cuba, land distribution is still highly uneven. The majority of holdings are less than five acres (two hectares) and are too small and fragmented to be economically viable. A few large estates of more than 500 acres (about 200 hectares) monopolize the greater part of the cultivated area.

Uneven land distribution

Large estates are often monocultural enterprises producing such items as sugar, rum, and molasses, but many also raise cattle—notably in Cuba—and grow bananas, mangoes, and citrus fruits; indeed, some estates have abandoned sugar and switched to these other products. Peasant produce is usually consumed at home—especially such staples as yams, cassava, and sweet potatoes—as well as carried to the market or sent overseas. Since the 19th century, the peasantry has made a substantial contribution to the export economy: pimiento, ginger, cocoa, breadfruit, coffee, arrowroot, and nutmeg tend to be peasant crops; sugarcane and bananas are grown almost equally on small and large holdings. Livestock raising is not significant, although Cuba has relatively extensive grazing



Population density of the West Indies.

A scenic beach near Grande Anse de Salines, Martinique, a typical attraction of Caribbean resorts.

© J. Messerschmidt from TSW—CLICK/Chicago

land for cattle. The islands produce much of the meat that they consume.

Industry. Petroleum is the West Indies' major industrial resource and has given rise to an important refining industry in Trinidad and Tobago, based partly on imported crude oil. An oil boom between 1975 and 1982—stimulated by new petroleum finds and world price rises—enabled the Trinidad and Tobago government to invest in steel and petrochemical complexes, but they did not produce immediate profits. Several largely foreign-owned multinational companies have refining or storage facilities in the Antilles, but the oil refining on Aruba, based on Venezuelan exports, was shut down, and that on Curaçao survives only with Venezuelan support.

Since World War II Puerto Rico has attempted to diversify its economy by giving tax concessions to foreign investors to encourage them to set up factories that take advantage of a cheap labour supply; this policy of intentive industrialization has been copied—but with less success—in Jamaica, Trinidad and Tobago, Haiti, Barbados, and The Bahamas. Industries that have been attracted involve mostly light, capital-intensive operations producing plastic and metal products, garments, sportswear, and chemicals and pharmaceuticals, usually for the U.S. market. In Puerto Rico manufacturing is the largest contributor to the gross domestic product; elsewhere, that sector relies on the sugar industry. A major problem with assembly industries using imported components is that they create few jobs and are rarely linked to local agricultural resources. Similar criticism has been made of the free-port manufacturing zones in Jamaica and the Dominican Republic, set up to take advantage of U.S. concessions offered under the 1982 Caribbean Basin Initiative. These industrial developments, coupled to government commitments to raise standards of living, other than in Haiti, have strained energy resources, in terms of both installed energy capacity and the cost of oil imports.

The Caribbean Basin Initiative

Sun, sand, and sea are the major natural advantages of the West Indies' tourist industry, plus their location within easy flying and cruising time for the large pool of North American tourists. Since the 1950s the north coasts of Jamaica and Puerto Rico, the leeward coast of Barbados, and parts of Antigua have developed as major tourist resorts. Tourism is a major contribution to the economy of Jamaica and Puerto Rico and is dominant in Antigua, Barbados, and the Virgin Islands.

Trade. The West Indies remain trapped in a classic dependency position reflected by unfavourable terms of trade: exporting cheap primary products—agricultural crops and mineral ores and petroleum—and importing comparatively expensive manufactured goods, increasingly consumer durables (such as vehicles and appliances). Major markets are the European Economic Community—notably in the former colonial metropoles, the United Kingdom and France—and Canada and the United States. The latter is the West Indies' major trading partner, especially for imports, which include enormous quantities of food. Cuba, since 1959, has become as dependent on Soviet trade—and subsidies—as it previously was on the United States and that country's sugar quota.

Transportation. Highly dependent upon trade, the West Indies requires good seaport and airport facilities. St. John's in Antigua, Bridgetown in Barbados, and Montego Bay in Jamaica have developed deepwater landing facilities, while Kingston has become a container port (able to handle specially packed cargoes). Most of the islands have international airports that can handle jet passenger aircraft, and the major tourist centres can accommodate the jumbo jets of the North American and European airlines.

Railways are important only in the larger islands, especially Cuba and Jamaica, where passenger services are maintained by systems that largely serve the sugar and bauxite industries, respectively. Major roads tend to follow the coast. Road networks are dense in Jamaica, Puerto Rico, and the Lesser Antilles but are difficult to maintain against the rainy season and hurricane damage. Only Cuba has a major highway serving the nation, in this case dating to the 1930s, but Trinidad and Tobago, Jamaica, and the French Antilles have some modern motorway sections.

History

Hispanic control of the West Indies began in 1492 with Christopher Columbus' first landing in the New World and was followed by the partitioning of the region by the Spanish, French, British, Dutch, and Danish during the 17th and 18th centuries. U.S. intervention in the Greater Antilles started in the early 19th century and culminated in the occupation of Cuba and annexation of Puerto Rico in 1898 and the purchase of the Danish Virgin Islands in 1917. Hence, each unit's connections—political, economic, and cultural—have been forged almost exclusively with the countries of western Europe or the United States. Soviet collaboration with Cuba since Fidel Castro's revolution in 1959 has perpetuated the history of great-power involvement in the West Indies. Before the colonization of the West Indies, however, pre-Columbian peoples there had evolved important and distinctive cultures.

THE PRE-COLUMBIAN PERIOD

Contemporary historians estimate that the Amerindian inhabitants of the West Indies numbered approximately six million in 1492—roughly one-fifth of the present-day total population. Three distinct Indian groups occupied the islands: the Ciboney, concentrated in the western parts of what are now Cuba and Haiti; the Arawak, located in the Greater Antilles and Trinidad; and the Carib, who lived mostly in northern Trinidad and the Lesser Antilles. Apart from a small reservation of Caribs in Dominica with some 500 members, mostly miscegenated with blacks, and a few scattered populations with Amerindian physical features in Cuba and Puerto Rico, the pre-Columbian population completely disappeared under the impact of culture shock, slavery, and diseases introduced by the Europeans.

Aboriginal Indian groups

Archaeologists divide the pre-Columbian populations of the West Indies into three chronological groups: the Paleo-Indians (5000–2000 BC), who were hunter-gatherers on the littorals of Cuba, Hispaniola, and Trinidad, and who probably originated in Central America; Meso-Indians (1000–500 BC), who were also hunter-gatherers but with a more sophisticated material culture—that of pottery, toolmaking, etc.—and who spread from South America to Trinidad and the Greater Antilles, where their remnants have been labeled Ciboney; and the Neo-Indians: first the Arawak, who entered Trinidad from South America about

300 BC and spread rapidly to the Lesser and Greater Antilles, only to be displaced from the Lesser Antilles after AD 1000 by aggressive Caribs, who migrated from Venezuela.

Arawak groups in the Greater Antilles shared a common life-style, language, and social organization. Their social structure was stratified and dominated by hereditary rulers called caciques, who may have had matrilineal lines of inheritance, and shamans presided over the Arawak's complex religious activity. The Arawak settled in villages that were established inland in forest clearings, and each village had its own chief, also called a cacique. Houses with circular ground plans, timber walls, and palm thatch roofs were arranged around a central open space. Villages were particularly plentiful in Hispaniola, usually with populations of between 1,000 and 2,000. Dancing and ball games were popular forms of recreation.

Arawak groups also exhibited throughout the Greater Antilles a uniform development in technology and techniques of subsistence. They fished, hunted, collected wild plants, cultivated kitchen gardens, and developed a system of shifting cultivation known as *conuco* for growing starch- and sugar-rich foods. The Spanish were impressed not only by their agricultural techniques but also by their use of fibres and their manufacture of canoes, gold ornaments, and pottery.

Carib villages in the Lesser Antilles, usually located on the windward coasts, were protected from surprise attack. Their social relationships were probably more flexible than those of the Arawak, and they had no hereditary caciques.

Carib culture

Many similarities, however, existed between Carib and Arawak material culture, especially with regard to *conuco* cultivation. While Carib pottery was inferior to that of the Arawak, their canoes and woven cloth were superior. Their houses, constructed of pole frames covered with palm thatch, were oval or rectangular.

EUROPEAN EXPLORATION AND COLONIALISM, 1492–1800

Discovery. With Columbus' voyage of discovery in 1492, the Caribbean Sea was transformed into a Spanish lake. Settlement by the Spanish concentrated on the Greater Antilles and above all on the densely populated island of Hispaniola (today divided into Haiti and the Dominican Republic), where the first permanent Spanish settlement in the Americas was established at Santo Domingo. Prospecting for precious metals led only to modest discoveries, but Santo Domingo rapidly became the "mother of settlement" in Latin America; the momentous expeditions to Mexico under Hernán Cortés and to Peru under Francisco Pizarro began from there. Their success diverted Spanish attention to the mainland in the 1520s, and Santo Domingo was soon superseded in commercial if not administrative significance by Havana (Cuba) and San Juan (Puerto Rico), which provided staging posts for the fleets of galleons transporting cargoes of bullion from the "Spanish Main" (the mainland bordering the Caribbean) to the Iberian Peninsula.

The Spanish missionary Bartolomé de Las Casas' intervention to prevent the genocide of the Indian population came too late to save the Arawak, although it did lead to the introduction of black slaves from Africa in the early 16th century, a solution to the Spaniards' labour problem that Las Casas had suggested. Small sugar industries were set up on a plantation basis in Cuba, Santo Domingo, and Puerto Rico, but they remained of minor significance and died out at the end of the 16th century. It was left to Sephardic Jews to introduce the sugar plantation to the British West Indies from northeastern Brazil in the 1640s, by which time the English and French had made colonial inroads into the Caribbean.

Colonialism. England was by far the most successful of the northwestern European predators on the Spanish possessions.

English conquests

In 1623 the English occupied Saint Christopher (Saint Kitts) and in 1625 Barbados. By 1655, when Jamaica was captured from a small Spanish garrison, English colonies had been established in Nevis, Antigua, and Montserrat. France took control of Guadeloupe and Martinique in 1635 and in 1697 formally annexed Saint-Domingue (Haiti), the western third of Hispaniola, which for about half a century had been occupied by buccaneers and French settlers. Curaçao, Aruba, and Bonaire, off the coast of present-day Venezuela, and Sint Eustatius, Saba, and half of Saint Martin (Sint Maarten), in the northern group of the Lesser Antilles, became Dutch possessions in the 1630s, but more as part of the military strategy of the Dutch war of independence against Spain than in expectation of agricultural riches.

These major gains against the Spanish were concentrated in the Lesser Antilles, which were poorly defended and essentially under Carib control. Only Jamaica and part of Hispaniola were wrested from the Spanish empire in the Greater Antilles, and Havana and San Juan continued to play a crucial part in Latin-American–Iberian trade until Spain lost its mainland empire through independence in the 1820s. The French and British continued to dispute the Lesser Antilles throughout the 18th century, and by the early 19th century Dominica, Saint Lucia, Saint Vincent, Tobago, and Grenada were in British hands, while Trinidad was formally ceded to Britain by Spain in 1802.

Plantation slavery. During the second half of the 17th century colonialism was linked to mercantilism (based on establishing gold and silver reserves and a favourable trade balance) and, in the British and French possessions in particular, to sugar plantations using slave labour imported from West Africa. The object of each of these imperial systems was to extract profits from the systems of trade in sugar, slaves, and manufactured goods. Mercantilism was most fully expressed in what is referred to as the triangular and quadrilateral trades; in their most complicated form these linked Europe, West Africa, the West Indies, and the eastern seaboard of what is now the United States in reciprocal commerce largely for the benefit of the British, French, and Dutch. Mercantilism peaked in the 18th century, before being replaced by the industrial capitalism that it had nurtured.

A major feature of white settlement in the West Indies was its transitory nature. The object of white adventurers, especially the British, was not to stay permanently in the West Indian colonies but to return to Europe with their fortunes made. Absenteeism became well established during the early 18th century, when many successful planters retired to Britain leaving representatives in charge of their estates. These absentees were a crucial element in the West India Interest, a powerful lobby that brought together merchants from the major ports, planters, and parliamentarians. It was the West India Interest that engineered the Molasses and Sugar acts in the first half of the 18th century. These acts protected British West Indian sugar in the British market and increased the prosperity of the planters.

The absentee planter

The power of the plantation was not confined to the economic sphere. It created an inegalitarian social hierarchy based upon racial distinctions and law. During the 17th century the major strata of West Indian society were free whites and black slaves, and, when by the mid-18th century miscegenation had become more prevalent, an intermediate mulatto stratum was formed through the manumission of the illegitimate offspring of the white elite. By law, however, whites alone were full citizens; the emancipated mulatto population exercised only limited civil rights, as did the increasing body of free blacks. Slaves, whether black or the illegitimate mulatto offspring of poor whites, were nonpersons, chattels to be bought and sold. Status and power were inversely related to group size; in Jamaica in the early 19th century, 25,000 whites dominated a colony with 40,000 mulattoes and 340,000 blacks.

THE WEST INDIES SINCE THE 19TH CENTURY

Emancipation. West Indian Creole societies were shaken by the successful slave rebellion in Saint-Domingue in the 1790s that resulted in Haiti's independence in 1804, thus creating the first black republic in the Americas. In 1807 Britain abolished the slave trade, and in 1833 slavery itself was abolished in the British West Indies. The French enacted emancipation in 1848 and the Dutch in 1863. But while these changes were taking place in the British, French, and Dutch West Indies, Spanish Cuba was developing as a slave-plantation producer of sugar. Slave imports into Cuba, despite a British naval blockade,

turned the island into a predominantly black and mulatto society by the second half of the 19th century. Full emancipation was not enacted in Cuba until 1886, 13 years after it was accomplished in Puerto Rico, where tobacco was more important than sugar and where slaves made up less than 5 percent of the population. Subsequent free white immigration, especially for plantation work in the early 20th century, once again transformed Cuba into a white society with a Hispanic culture.

Plight of the ex-slaves

Nineteenth-century emancipation created fewer changes than the slave owners feared, largely because most of the best land was in plantations and because property-restricted franchises—especially in the British West Indies—favoured the established order. Nonetheless, emancipated slaves were free to sell their labour, migrate, squat, or purchase land. A "reconstituted" peasantry emerged in Haiti, Jamaica, and the Windward Islands. In the larger Leewards and Barbados the ex-slaves were unable to acquire land because of the lack of a mountainous interior, and the mountainous land on other islands was all that was available to the former slaves; thus they remained as plantation workers or emigrated to Central America or the United States. Cuba's emancipated blacks were soon caught up in the war of independence with Spain; their descendants were later drawn into the burgeoning sugar industry developed by U.S. capital, much as mulattoes had been in the Dominican Republic and as whites would be in Puerto Rico.

The persistence of the plantation system and of white elitism, bolstered by colonialism, shored up the structure of the grossly inegalitarian societies of the West Indies after emancipation. Colour-class and culture-class correlations persisted in a situation where—excepting the French West Indies from the late 19th century—democracy was systematically denied. The complexity of the white-brown-black social hierarchy was compounded in some islands by the arrival of other ethnic groups. Chinese indentured immigration to Cuba; East Indian indentured immigration to Trinidad and to a lesser extent to Jamaica, Martinique, and Guadeloupe; and free movement of Chinese, Portuguese, Syrians, and Lebanese to Trinidad and the Greater Antilles (mainly in the 20th century) produced minorities with the potential for social mobility. But, whereas the Portuguese, Syrians, and Lebanese (like the Jews before them) used trade to achieve mobility, the East Indians remained largely tied to rural enclaves, even in Trinidad, where they have come to account for more than 40 percent of the population.

Decolonization. Radical changes in the social position of nonwhites has depended less upon emancipation than on decolonization. Having liberated themselves in 1804, the Haitians in the early 1820s invaded Santo Domingo and incorporated the former, almost forgotten Spanish colony into a Hispaniola-wide Haiti. In 1844, Dominicans rejected Haitian hegemony and declared their sovereignty. Later they reverted briefly to the Spanish crown and achieved their final independence in 1865. The third independence from a European power in the West Indies was Cuba's, in 1898, and it involved not only two wars of independence with Spain but also U.S. intervention. Cuba achieved formal independence from the United States in 1902 but remained a fief of its northern neighbour until the Platt Amendment was abrogated in 1934. By this time labour disturbances were about to break out throughout the region, generated by an international economic depression and the lack of means for democratic dissent; popular demands for development and decolonialization were spreading from Jamaica to Trinidad.

Types of independence

Most of the West Indian societies were decolonized with imperial consent after World War II, either by a grant of full independence, as in the case of most of the British territories; by incorporation into the mother country, as in the French Antilles; or through their association with the mother country, as in the Netherlands Antilles and some of the British territories. All these methods of decolonization have been endorsed by the United Nations. Nevertheless, Britain still retains several small West Indian dependencies, most of which have shown little interest in independence: Montserrat, Anguilla, the British Virgin Islands, the Cayman Islands, and the Turks and Caicos Islands.

When World War II ended in 1945, only three West Indian states were independent—Haiti, the Dominican Republic, and Cuba—and all either were or were about to become dictatorships. When the British, French, and Dutch began to decolonize, either by liberating or incorporating their territories, a major preoccupation was the establishment of democracy. In pursuit of this objective the British tried to create an interisland federation that would incorporate more checks and balances than was feasible in a unitary state, as well as provide a vehicle for small-island decolonization. It was the failure of this policy with the breakup of the West Indies Federation (1958–62) that led to the first phase of independence (1962–66), the second stage of independence involving the Bahamas and the British associated states (1973–83), and the ultimate recognition by Britain of a residual and continuing category of dependencies.

Dependency. If most West Indian political units are no longer colonial, dependence remains the hallmark of Caribbean economies. For decades the terms of trade, for example, have operated against West Indian primary producers; except for oil-rich Trinidad, none of the West Indian islands produces enough of any one major commodity to have a decisive role in fixing prices. More damaging in many cases than the terms of trade has been the penetration of each island economy by foreign enterprise. Neocolonialism prevails in most islands, and West Indians are acutely aware of their dependence on overseas capital, decision makers, and technologies, particularly in the French Antilles.

The importance of multinational corporations in West Indian economies is reflected in the preeminent position of North American companies in the Jamaican, Haitian, and Dominican Republican bauxite industries and in Trinidad's petroleum economy; the role of the British-based company Tate and Lyle in West Indian sugar production and refining; and the monopolistic position of both British and North American companies in the marketing of bananas from Jamaica and the Windward Islands.

Notwithstanding these constraints on sovereignty, the 1970s and '80s witnessed substantial nationalization of foreign-owned enterprise in the West Indies because of pressure from governments and, in the Commonwealth Caribbean, as a result of the willingness of companies to surrender their least viable operations and to use the compensation to open up activities elsewhere. The West Indies Sugar Company, a subsidiary of Tate and Lyle, divested itself first of sugarcane land, then of factories in Jamaica and Trinidad, and Jamaicans and Trinidadians now hold an interest in foreign banks. Trinidad and Tobago has purchased significant North American petroleum assets, and Jamaica in the 1970s concluded participation agreements with several North American aluminum companies. Almost half of the hotel industry in Jamaica fell by default into government hands as a result of the catastrophic reduction in visitors from North America after 1976, following the negative reports about Prime Minister Michael Manley's socialist policies in the U.S. press. None of the other Greater Antilles has approached Cuba's post-1959 policy of wholesale nationalization within an ideological framework of Marxism; but, at the other extreme, the nations of the Lesser Antilles—except for Grenada during the People's Revolutionary Government (1979–83)—have not made any concessions to national involvement in, or ownership of, the leading economic sectors. In the 1950s it was expected that decolonization would deliver more than halting development of dependent economies under generally conservative elites. When compared to the example of oppression and impoverishment in Haiti after 30 years of Duvalierist and post-Duvalierist rule, however, the British, French, and Dutch former colonies can be said to have made considerable social and economic progress under generally democratic governments. Indeed, the process of democratization—one of the major goals of decolonization—is largely regarded to have moved successfully. However, governments remain unclear on how the benefits of self-determination can be extended into

Democratization

grass-roots economic development and social change, although some countries, Jamaica in particular, have made considerable progress in breaking the real and psychological ties of dependency. (C.G.C.)

For later developments in the history of the West Indies, see the *Britannica Book of the Year* section of the BRITANNICA WORLD DATA ANNUAL.

For coverage of related topics in the *Macropædia* and *Micropædia*, see the *Propædia*, sections 966 and 974, and the *Index*.

THE GREATER ANTILLES

Cuba

Cuban archipelago

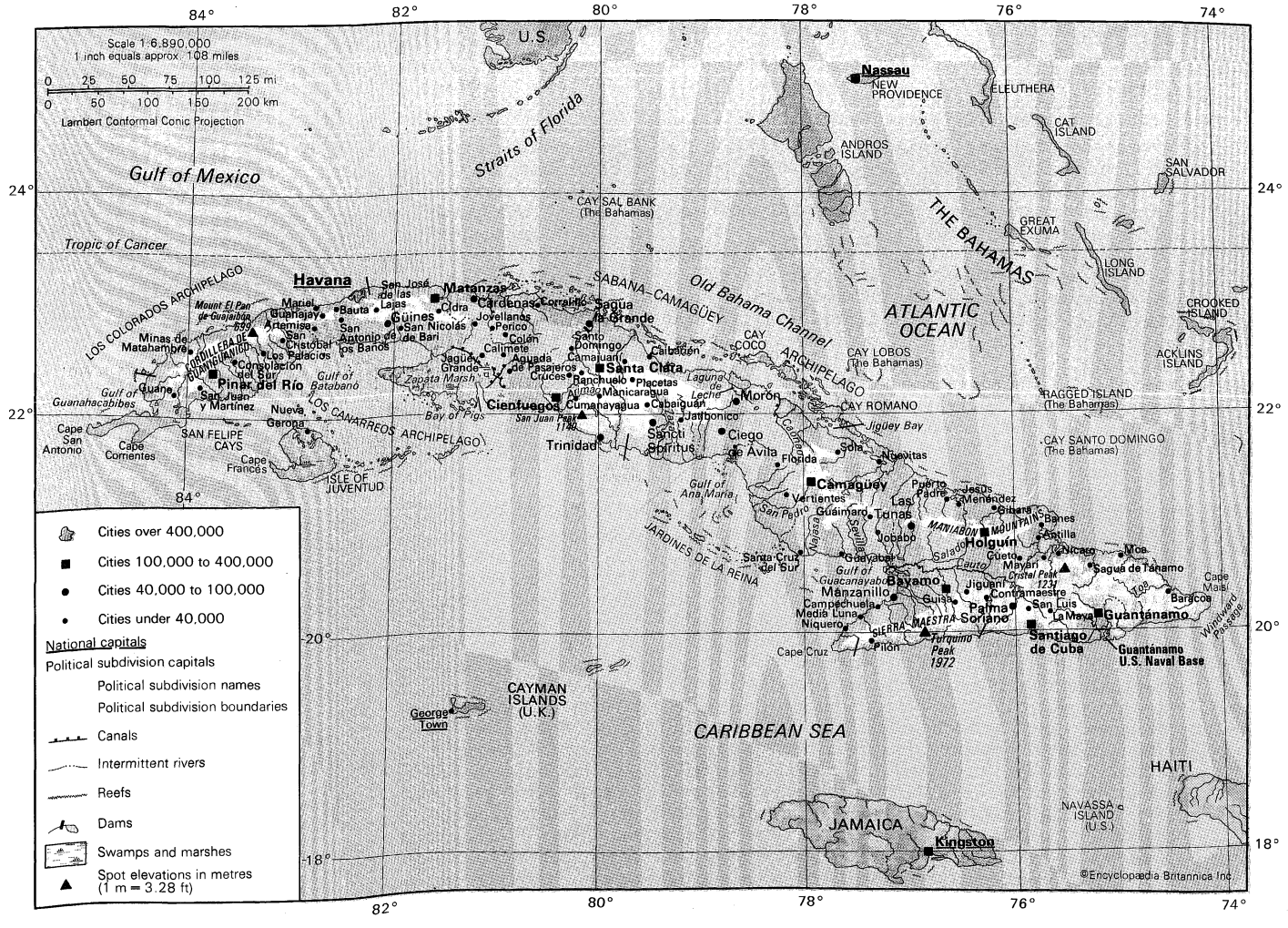
The Republic of Cuba comprises the Cuban archipelago, a formation of about 1,600 islands, islets, and cays with a combined area of 42,804 square miles (110,861 square kilometres). This archipelago is situated just south of the tropic of Cancer at the entrance to the Gulf of Mexico and forms an important segment of the Antilles island chain, which continues east and then south in a great arc enclosing the Caribbean Sea. The island of Cuba itself—the "Pearl of the Antilles"—is by far the largest island in the chain, covering 40,519 square miles (104,945 square kilometres). In general the island runs from northwest to southeast and is long and narrow—777 miles (1,250 kilometres) long but only 119 miles (191 kilometres) across at its widest and 19 miles (31 kilometres) at its narrowest point. The capital is Havana, on the northwestern coast. Haiti, the nearest neighbouring country, is 48 miles away across the Windward Passage; Jamaica, 87 miles to the south; and the United States, about 90 miles to the north across the Straits of Florida.

The Isle of Juventud (Isle of Youth; formerly called Isla de Pinos [Isle of Pines]), is the second largest in the archipelago (849 square miles); it rises to the southwest of Cuba itself. Other archipelagoes are the Los Colorados, to the northwest; Sabana-Camagüey, off the north central coast; the Jardines de la Reina (Queen's Gardens), near the south central coast; and the Canarreos Archipelago (technically including the Isle of Juventud), along the southwest coast.

Because of its geographic location and natural resources, Cuba was coveted by more than one foreign power over the centuries. A colony of Spain after its discovery by Columbus in 1492, it formally became a republic at the beginning of the 20th century, although with a high degree of political and economic dependence on the United States. On New Year's Day, 1959, revolutionary forces led by Fidel Castro won a civil war and took power. In 1961 the Marxist-Leninist nature of the revolution was proclaimed. Revolutionary Cuba developed close political and economic links to the Soviet Union and has become a leader of the nonaligned movement.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Mountains cover about a quarter of the total area of the island of Cuba. They are often interrupted by the plains that cover some two-thirds of the surface. The coastal basins of Santiago de Cuba and Guantánamo lie in the extreme east; a great central valley also begins in the east and then combines with a peneplain that continues westward across the island. It is these plains that have been hospitable to sugarcane and livestock raising.



The *alturas*—regions of moderate elevation—are in some cases residues of formerly higher surfaces. More rugged relief includes the Guaniguanico range in the west, comprising the Sierra de los Órganos and the Sierra del Rosario, which attains 2,270 feet (692 metres) at El Pan de Guajaibón; the Sierra de Trinidad in the central region, with the 3,793-foot (1,156-metre) San Juan Peak; and the Sierra Maestra, on the southeast coast, approximately 150 miles long and containing the island's highest peaks, with Real del Turquino Peak at 6,476 feet (1,974 metres) preeminent.

Mountain regions and the coastline

Cuba possesses an irregular 3,570-mile coastline, made picturesque by many bays, sandy beaches, mangrove plantations, swamps, coral reefs, and rugged cliffs. There are also some spectacular caverns in the interior, notably the Cave of Santo Tomás of the Quemado ridge region, which has a linear extension of 16 miles. The island of Cuba is surrounded by a submerged platform that is an additional 30,000 square miles (75,000 square kilometres) in area.

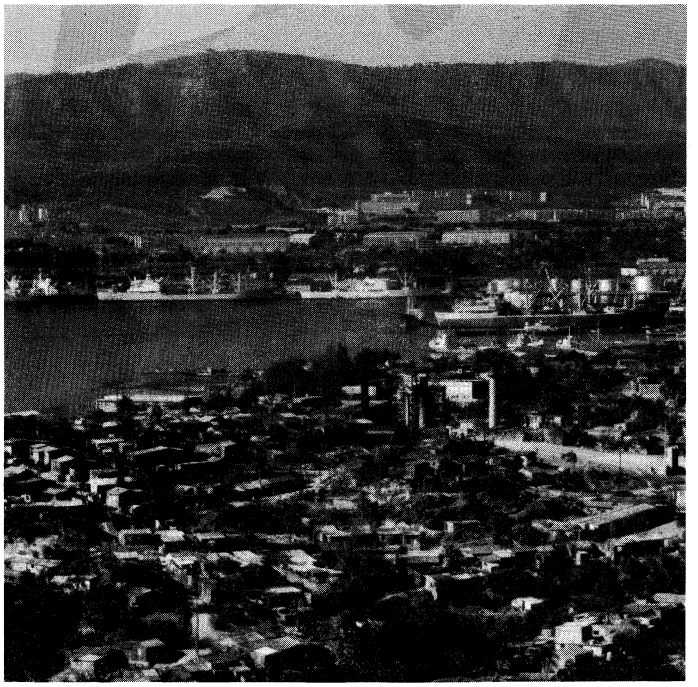
Drainage and soils. Cuba's supply of groundwater is utilized throughout the island but especially in La Habana province. The rivers are generally short, with very meagre flow; of the nearly 600 watercourses classified as rivers, two-fifths discharge to the north, the remainder to the south.

The island's heaviest rainfall and therefore its largest rivers are in the southeast, where the Cauto (at 230 miles the country's longest) and its tributaries, notably the Salado, drain the Sierra Maestra and the uplands to the north. Other major rivers in this region include the Guantánamo, Sagua de Tánamo, Toa, and Mayarí. To the west the most important southward-flowing rivers are the Sevilla, Najasa, San Pedro, Jatibonico del Sur, Zaza, Agabama, Arimao, Hondo, and Cuyaguateje; northward-flowing rivers include the Saramaguacán, Caonao, Sagua la Grande, and La Palma.

Cuban lakes are small and more properly classified as freshwater or saltwater lagoons. The latter include the 26-square-mile Leche Lagoon (Milky Lagoon), which is technically a sound, being connected to the sea by several natural channels. Sea movements generate disturbances in the calcium carbonate bottom deposits to produce the milky appearance.

The complicated Cuban topography and geology have produced soils of no fewer than 13 different major groups, the majority of which are fertile and amenable to year-round cultivation. Highly fertile red limestone soil extends from west of Havana to near Cienfuegos on the southern coast and lies in extensive patches in western Camagüey province, providing the basis for Cuba's main agricultural output. Another area of fertile soil is north of Cienfuegos between Sierra de Sancti Spiritus and the Caribbean coast. Camagüey province and the Guantánamo basin have some arable land, although of lower fertility. Areas of sandy soil in Pinar del Río, Villa Clara, and portions of Ciego de Ávila and Camagüey cannot hold moisture and are of low fertility, as are the soils of the mangrove-dotted coastal swamps and cays.

H. Armstrong Roberts



The Sierra Maestra mountains overlooking the port of Santiago de Cuba along the country's southeastern coast.

Climate. Cuba lies in the tropical zone, located on the southwestern periphery of the North Atlantic high atmospheric pressure zone and hence influenced by the northeast trade winds in winter and east–northeast winds in summer. The warm currents that form the Gulf Stream have a moderating influence along the coasts.

Annual mean temperature is 79° F (26° C), with little variation between January (at 73° F [23° C] the coolest month) and August (the warmest month, at 82° F [28° C]). The November–April dry season abruptly changes to the rainy May–October season. Annual rainfall averages 54 inches (1,380 millimetres). Between June and November, the country often is exposed to hurricanes, whose strong winds and heavy rains have occasioned great economic damage and human suffering.

Temperatures

Plant and animal life. Cuba's tropical plant life is very rich, with some 8,000 species represented, 7,000 of them flowering plants; half of these are endemic. Much of the original vegetation has been replaced by sugarcane, coffee, and rice plantations, made possible by enormous and indiscriminate destruction of forests. The revolutionary government introduced extensive reforestation for both economic and conservation reasons.

Cuban timber is of excellent quality, although the supply is limited. Pine is found in abundance, and the durable mahogany is of potential economic importance, while ebony (*Diospyros*) and granadilla (cocus, or West Indian ebony; *Brya ebenus*) are both beautiful and valuable. The abundant royal palm, reaching heights of 50 to 75 feet, is the national tree and a characteristic element of the rural landscape. The ceiba (kapok) tree plays a role in many legends, while the extremely rare cork palm (*Microcycas calocoma*) of the western regions is a "living fossil," thought to have survived for more than 100 million years.

Fruit trees include such citrus varieties as lemon, orange, and grapefruit; some species of the genus *Annona,* including the *guanábana* (soursop) and *anón* (sweetsop); and avocado and papaya. Banana plants are also common. The lower Cuban coasts and the shoals of the archipelago are given marked character by the mangrove, and the tobacco plant is commercially prized.

Cuban animal life is extraordinary in its abundance and variety, particularly the invertebrate species. The archipelago is the home of more than 7,000 insect species and 4,000 species of land, river, and sea mollusks. Commercially valuable sponges are found off the southwestern coast, and edible crustaceans abound. Similarly profuse, arachnids include the tarantula and scorpion. Fish, with more than 500 species (many edible), are economically the most important vertebrate group; there are about 35 species of shark. Freshwater fishes are considerably less abundant. Cuba is visited by many migratory birds. Only about a third of the 300 or so species found on the island are indigenous; these include the flamingo, royal thrush, and nightingale.

Invertebrates

Reptiles are distributed equally among sea, river, and dryland species. Marine species include the tortoise and hawkbill turtle; mud turtles inhabit the rivers; and the marshes contain two species of crocodile, once almost extinct but preserved through a repopulation program. Land reptiles include the iguana and the nonpoisonous *majá de Santa María* (*Epicrates angulifer*), Cuba's largest snake; none of the snake species are venomous. Amphibians are similarly varied, with 60 frog and toad species, the former including the plantain frog (*Hyla septentrionalis*) and bullfrog. The solenodon (*Atopogale cubana*), an almost extinct, ratlike insectivore, is found only in the remotest eastern regions. Other mammals include the hutia (an edible rodent) and

the manatee, or sea cow, which inhabits river mouths. A variety of bat species destroy mosquitoes as well as insects harmful to agriculture and in their roosting caves produce accumulations of guano that is valuable as fertilizer.

Settlement patterns. The various areas of Cuba do not differ greatly in customs and traditions, making a clear regional division on this basis difficult. Given this qualification, a division of Cuba into five geographic regions—Occidental (Western), Central, Camagüey–Maniabón, Oriental (Eastern), and the Isle of Juventud—is still possible on the basis of natural endowment and human exploitation.

The largest of Cuban regions, the Occidental region is 332 miles long and includes the provinces of Pinar del Río, La Habana, Ciudad de la Habana, Matanzas, and parts of Villa Clara and Cienfuegos. Its topography is notable for the *mogotes,* unusually shaped elevated hummocks, although it is primarily composed of an enormous plain that occupies its entire southern section. The 110-mile-long Cordillera de Guaniguanico and the serpentine highlands of northern and central La Habana and Matanzas provinces further characterize the region. The Occidental region contains about 40 percent of the country's population and includes Havana, the largest city and the major Cuban economic, cultural, and administrative centre. The regional economy is based on agriculture (some of the best tobacco in the world, citrus fruits, sugarcane, cattle, rice, and coffee), manufacturing, copper and iron mining, tourism, and fishing.

Cordillera de Guaniguanico

The Central region extends from the Manacas plain (west of the city of Santa Clara) to La Trocha plain in Ciego de Ávila province and includes the Trinidad and Sancti Spíritus highlands and lesser uplands, interspersed with plains and boggy regions in the province of Sancti Spíritus. It also includes the archipelago of Sabana–Camagüey. The major Cuban tobacco-growing area, the Central region also produces sugarcane, cattle, citrus fruits, and coffee as well as marble, copper, iron, and limestone. Industrial activity in the region is varied. Santa Clara is the region's main centre of population.

The Camagüey–Maniabón region, Cuba's main stock-raising area, centres on an ancient peneplain that extends over central Camagüey province and northwestern Las Tunas province between La Trocha plain and the Maniabón highlands. Other prominent economic activities include the cultivation of sugarcane, rice, and grains, mining of chromium, and processing of cement, fertilizer, and gypsum. Camagüey is Cuba's third largest city and the region's main population centre.

At the eastern tip of the main island is the Oriental region, including the province of Granma and parts of Holguín, Santiago de Cuba, and Guantánamo. There are the nation's highest mountains, fastest streams, richest mines, and most spectacular bays. The Sierra Maestra rises along the southern coastline, the Sierra del Cristal and other ranges on the northern, and between them lies the economically important central valley, the easternmost portion of which rises to the Sagua–Baracoa highlands, which are covered by the most extensive forests in Cuba. The region contains nickel reserves and produces chromium, iron, copper, and manganese as well as sugarcane, coffee, and other agricultural items; its industrial centres process minerals and food and generate electric power. It includes Cuba's second largest city, Santiago de Cuba.

The Isle of Juventud, dotted with groves of pine and palm, has sand and clay plains in the north and hills in both the northwest and southeast; much of the island, however, is taken up by the gravel bed in the south and by the bogs of the coasts and uninhabited interior. At one time sparsely populated, it became the object of government resettlement projects in the mid-1970s. The cultivation of citrus fruits was established as the basis for the island's economy. Nueva Gerona is the island's capital.

The people. *Ethnic distribution.* For more than four centuries, diverse ethnic groups have been settling in Cuba. Not only Spaniards and Africans (the predominant elements) but also Chinese, Jews, and Yucatecans (from the Yucatán in Mexico) have all superimposed their cultural and social characteristics on those of the earliest settlers. Contemporary Cuban society exhibits a remarkable diversity as a result.

Cuba's original inhabitants probably came to the island from South America. They were the Guanahatabey and the Ciboney, the former living in the extreme west of the island, the latter in various places in the island and particularly on the cays to the south. Both were hunter-gatherers. The Taino (Arawakan Indians), who arrived later and who spread over not only Cuba but also the rest of the Greater Antilles and the Bahamas, lived in villages and had rudimentary agriculture; they also made simple pottery. The Taino constituted 90 percent of the island's population at the time of the Spanish conquest.

Aboriginal peoples

At the beginning of the 16th century the total indigenous population was estimated at about 112,000, very unequally distributed, with density decreasing westward. Half a century after the Spanish invasion, however, only about 3,000 scattered individuals remained. Harsh treatment by the invaders, hard labour in the gold mines, hunger resulting from low agricultural productivity, suicides, and contagious diseases introduced by Europeans had all taken their toll. A few families with Taino physical characteristics living in the Sierra del Purial of easternmost Cuba are perhaps the only surviving descendants.

White Cubans, who constitute about two-thirds of the total population, are almost entirely of Spanish origin. Throughout Cuban history the dominant classes were primarily the Europeans and their descendants, white and mestizo (the latter being of particular importance in the 20th century), who monopolized not only the direction of the economy but also access to education and culture.

The Spaniards soon imported African slaves (approximately 760,000 in all, most of them for work on the sugar plantations) as a substitute for the rapidly disappearing Indians. The Africans came mainly from Senegal and the Guinea Coast, with diverse origins that included Yoruba and Bantu tribal backgrounds. Between 1919 and 1926 some 250,000 black Antillean labourers, 90 percent from Haiti and Jamaica, arrived under labour contract; nearly all remained. By the 1980s blacks and mulattoes made up one-third of the population. Their cultural influence has been considerable, especially in music and dance.

In order to supplement the interrupted slave trade, the Hispano-Cuban landholders imported 125,000 indentured Chinese labourers, nearly all of them Cantonese, between 1853 and 1874 to work under contract for eight years. Bad living conditions reduced their numbers to 14,000 by the census of 1899. In the 1920s an additional 30,000 Cantonese and small groups of Japanese also arrived; both immigrations were exclusively male, and there was rapid intermarriage with white, black, mulatto, and mestizo populations. In contemporary Cuba the Asian element makes up less than 1 percent of the total population.

Indentured Chinese

Linguistic composition. Spanish is the Cuban national language; there are no local dialects, although the diversity of ethnic origins has influenced speech patterns. Some words are of native Indian origin, and a few (such as *hamaca* ["hammock"]) have passed into other languages. Africans have also enriched the vocabulary and contributed the soft, somewhat nasal accent and the rhythmic intonation that distinguish contemporary Cuban speech.

Religions. Among the religions in Cuba are Roman Catholicism, Santería (a cult devoted to certain African divinities formally identified with Catholic saints), and a number of Protestant, Jewish, and other groups. Prior to the revolution Roman Catholicism was predominant, although permeated by Santería and rather weak in rural areas. In the early 1960s church and state confronted one another with open hostility, the church seeing the revolutionary government as antireligious (one aspect of this being the nationalization of all schools) and the government seeing the church (the largest mass organization in the country) as a repository of counterrevolution. In the first half of the 1960s about 70 percent of all Roman Catholic priests, 90 percent of the nuns, some Protestant clergy, and all rabbis left the country either voluntarily, under social pressure, or through deportation. By the late 1960s relations between Roman Catholic and Protestant churches and the state had begun to improve; the 1976

constitution proclaimed scientific materialism as the base of the state and education but recognized and guaranteed religious freedom, although with restrictions. By the 1980s both the church and the government had made some concessions and had entered a period of rapprochement.

Demographic trends. Prior to 1960 immigration was a contributing factor in the total population growth of Cuba, but since that year the number of persons leaving the country has outnumbered new arrivals and the growth of the population has been mostly a consequence of changing birth and death rates. The birth rate rose steadily from 1958 to 1963 due to the increased standards of living and expectations among low-income groups, increased sexual freedom among females, and larger numbers of women marrying at a younger age. On the other hand, there were increasing rates of death and infant mortality, caused by a mass exodus of physicians, a scarcity of medicines, and a rise in contagious diseases. In the mid-1960s the high birth rate stabilized and began to decline as increasing numbers of females entered the labour force, sex education improved, strict limitations on consumption were instituted, housing became scarce, and compulsory military service at age 16 was initiated. By 1978, the birth rate had declined by more than one-half of its 1960s peak of 35 births per thousand persons. In the early 1970s the mortality rate also dropped as more physicians were trained and became available, the supply of medicines increased, and contagious diseases were controlled through vaccinations. In succeeding years, however, a slightly increasing trend in the mortality rate developed because of population aging. The combined effect of the declining birth rate and heavy emigration has resulted in a dramatic decrease in the overall rate of population growth.

Trend of vital statistics

The economy. By the end of the 1950s Cuba had developed one of the leading economies among Latin-American nations. Nevertheless, the country was confronted by a number of major problems: a sugar monoculture (sugar accounted for four-fifths of total exports); a low rate of economic growth; a heavy dependence upon the United States for investment and trade; high rates of unemployment and underemployment; and significant inequalities between urban and rural areas and among the various ethnic divisions.

When the revolutionary government took over in 1959, it set out to correct these problems through various means, the most significant being collectivization of all means of production (except for 9 percent of agricultural land); establishment of a centrally planned economy; emphasis on industrialization and deemphasis of sugar production (both later reversed); formation of close economic ties with the Soviet Union; and strong development of social services, particularly in rural areas.

The measures taken achieved mixed results. The attempt to introduce central planning (following the Soviet model of the 1950s) in 1961–63 failed because of the lack of infrastructure and qualified personnel and because of overly ambitious goals. After a period of intense debate (1964–66), the role of the central plan was reduced and emphasis placed on moral incentives (nonmonetary awards such as medals and titles, labour mobilization, and the development of the "new man" [1966–70]). When this approach also failed to bring about the desired results, there was a return to Soviet-type central planning and the orthodox system of socialist incentives. In the mid-1980s, however, policies again reversed the economic direction; Cuban leaders rejected the policies of Soviet leader Mikhail Gorbachev (advocating the use of market mechanisms) as inadequate for their nation's needs. The policies of the government were more successful in reducing unemployment (to about 3 percent of the labour force), although at the cost of overstaffing state enterprises and lowering labour productivity, and sharply reducing income and urban-rural inequities. Economic growth rates have been erratic and relatively low; periodic experiments with moral incentives have provoked economic slowdowns and negative growth rates. Sugar continues to account for about three-fourths of exports, but there has been significant development of domestic industry. Dependency on the Soviet Union is greater than it once was with the United States. However, the U.S.S.R. does not have direct investment in Cuba but has provided substantial economic aid.

The directorate of the Communist Party of Cuba is the planning agent for all phases of the Cuban economy. The institutional economic structure consists of the Central Planning Board, headed by the economics minister; the ministries and national organizations that control the economic sectors and basic activities; the various state enterprises (*empresas*); and the provincial delegations that direct the work of the factories and related services. Wages and prices are rigidly controlled and quota systems strictly enforced.

The national planning mechanism

Resources. Cuban soil is fertile, allowing for up to two crops a year, but agriculture has been traditionally plagued by the unreliability of the annual rainfall. Subterranean waters are an important additional resource for the country.

Petroleum deposits supply only a small percentage of the nation's needs, and the rest is met by imports, mostly from the Soviet Union. Peat, concentrated in the Zapata Peninsula, is still the most extensive fuel reserve. Nickel, chromite, and copper are the most important minerals mined, and the laterite (iron ore) beds in Holguín province have considerable potential. Nickel ore, which also yields cobalt, is processed in several large plants, and Cuba is a world leader in nickel production. There are also major magnetite and manganese reserves and lesser amounts of lead, zinc, gold, silver, and tungsten. Abundant limestone, rock salt, gypsum, and dolomite reserves and large kaolin and marble beds are also found on the Isle of Juventud.

Agriculture and fisheries. The Cuban economy remains heavily dependent upon the sugarcane crop, as it has since the end of the 18th century. Vast areas have been leveled and brought under irrigation, thus greatly expanding the acreage in sugarcane, and yields per acre have been increased with the application of fertilizers. Sugar output, except in years of drought or sugarcane blight, increased after the introduction of mechanized harvesters in the early 1970s, although prices were affected by a sugar glut in the 1980s.

The number of cattle was increased in the 1960s by cessation of slaughter of reproducing cows, by irrigation projects that added to available pasture, by artificial insemination, and by expanded veterinary services. Brahman (or Zebu) cattle, the dominant breed (resistant to tropical climate but low in milk yield), were crossed with Holsteins (productive but prone to illness in the Cuban environment) and also with Brown Swiss in an unsuccessful attempt to produce acclimatized meat and milk producers. The number of cattle decreased sharply during the period 1967–86.

Improvements in stock breeding

Apart from sugarcane, the chief agricultural products are rice (the main source of calories in the traditional diet) and citrus fruits, the latter an important export crop. Tobacco, traditionally the country's second crop, is grown mainly in the Pinar del Río area in the west and also in the centre of the island; production has slowed, however. Other products include bananas, coffee, cocoa, pineapples, sweet potatoes, potatoes, corn (maize), cassava, and beans. Cuba has found it necessary to import large amounts of food (particularly rice), oilseeds, and cotton.

Fishing resources are significant. Fish and other seafood production increased overall sevenfold during the period 1959–79, although fluctuations were common. The increase occurred largely because of government investment in fishing vessels and processing plants; however, the institution of fishing zones by neighbouring countries since the 1980s adversely affected the industry.

Industry. Official sources are not clear on the relative value of industry in the Cuban economy since the Castro revolution, some showing a slight decline, others indicating an increase. The output of nickel, Cuba's most important metal, declined in the early 1960s, recuperated in the 1970s, and stagnated in the 1980s; construction of new nickel complexes began in the 1970s, but completion was delayed through the 1980s. Petroleum output has expanded dramatically but by the end of the 1980s met less than 10 percent of domestic needs. Nevertheless it is the main source of power. Production of hydroelectric en-

ergy is minimal, but completion of a Soviet-made nuclear power plant was projected to bring Cuba its first nuclear power. Food processing (including sugar), beverages, and tobacco combined account for approximately one-half of the value of total industrial output; construction materials, chemicals, electricity, and fuels for one-fourth; and such items as textiles, clothing, and steel the remainder. In the late 1950s close to 300,000 tourists, mostly from the United States, visited Cuba each year, but the number declined to less than 2,000 in 1971. A slow recuperation followed.

National banking system

Finance. The banking system has been operated by the state through the National Bank of Cuba since 1966. There are no stock exchanges. Prices are centrally fixed, and investment is allocated by the national economic plans. With its controlled economy, inflation has been negligible, although there was a 12-percent increase in the early 1980s. Under the guidelines of the revolution, foreign investment was prohibited until 1982, when a joint-venture law was enacted; this, however, has failed to attract substantial investment.

Trade. In the 1950s more than two-thirds of Cuban foreign trade was with the United States. By 1961 this was down to 4 percent, and it soon dropped to zero under U.S. government embargo policies. Trade shifted to the Soviet Union and other socialist countries, and in 1972 Cuba became a full member of the Eastern-bloc Council for Mutual Economic Assistance (Comecon). By the end of the 1980s more than 80 percent of Cuban trade was with socialist countries, the Soviet Union handling most of it. The composition of exports has changed little: sugar continues to dominate, amounting to some three-fourths of the total, followed by reexports of Soviet imported fuels and by nickel, fish products, citrus fruits, and tobacco. Import composition, on the other hand, has changed more dramatically in some areas: the share of machinery and transportation equipment has remained about the same (one-third), but the share of fuels and minerals has tripled (to one-third), while that of manufactured goods and food products have been cut in half.

Transportation. Cuba's merchant fleet has grown to keep pace with the increase in foreign commerce, but it nevertheless can handle only a small percentage of the country's shipping and must rely on foreign fleets for the bulk of the trade. The major ports are Havana (fuels, grains, and other commodities), Cienfuegos (sugar exports), Santiago de Cuba, Guayabal, Matanzas, Antilla, Nuevitas, Mariel, and the U.S. naval base at Guantánamo.

The railway constructed between Havana and Bejucal in 1837 was the first in the Americas after those of the United States. The railway system deteriorated in the first years after the revolution of 1959, but much of it has been restored and most of the system continues to serve the sugar industry.

The most important highway is the Central Highway, built in the 1920s; it runs along almost the entire length of the island. Other major routes are the Vía Blanca (linking Havana with the Playa Varadero) and the Vía Mulata (connecting Baracoa, at the east end of the island, with the rest of the country). There are state transport enterprises (mostly for passenger traffic), one of which is national, the rest being provincial in scope. Automobile transport has been substantially downgraded because of the sharp curtailment of auto imports since the 1960s.

The Cuban Aviation Enterprise (Empresa Cubana de Aviación), or Cubana, operates the former four private airlines. International airports operate at Havana, Santiago de Cuba, Camagüey, and Varadero.

Administration and social conditions. *Government.* Until the adoption of the constitution of Feb. 24, 1976, Cuba had for some 36 years been governed either by the constitution of 1940 or by the postrevolutionary Fundamental Law of Feb. 7, 1959, modeled upon the constitution but centralizing governmental power. The 1940 constitution had been suspended twice—from 1952 to 1955 by the dictatorship of Fulgencio Batista, and after 1959 when it was supplanted by the Fundamental Law and by legislation that included the Agrarian Reform Law (May 17, 1959), the second Urban Reform Law (Oct. 14, 1960), the Nationalization of Education Law (July 6, 1961), and the second Agrarian Reform Law (Aug. 3, 1963).

The 1976 constitution

Drafting of a new constitution to succeed that of 1940 began in 1965 and continued for the next 10 years; a preliminary draft was approved by the Politburo of the Central Committee of the Communist Party of Cuba in 1975, and the final version was approved by referendum on Feb. 15, 1976, entering into force on February 24.

In October 1976, for the first time in 17 years, representatives for 169 municipal assemblies "of the people's power" were elected to give the people a more effective role in the running of their urban centres. These 169 assemblies met subsequently in November to elect a 481-member (510 after 1986) National Assembly, as they had somewhat earlier elected delegates to the 14 provincial assemblies; each of the bodies in turn elected executive committees to carry on the day-to-day work of their respective administrative organs.

The National Assembly, at its inaugural session in December 1976, appointed a State Council of 31 members, headed by a president, and a Council of Ministers, also headed by the president. The posts of president as well as those of first secretary of the Communist Party and commander in chief of the armed forces were all held by Fidel Castro from 1976. The State Council is the executive body of the state, carrying on the daily administration of the country between the twice-yearly sessions of the National Assembly.

The Communist Party of Cuba

Following the revolution, political parties were dissolved, and a single party was created out of the participating revolutionary organizations: the 26th of July Movement, the Popular Socialist Party, and the 13th of March Revolutionary Directorate. In 1965 this single national party was officially designated the Communist Party of Cuba.

The mass organizations were created after the revolution to replace former social organizations and are under the supervision of the government. The Confederation of Cuban Workers, reconstituted in 1970, with stated objectives to support the government, is designed to improve managerial performance and labour discipline and to raise the political consciousness of workers. The National Association of Small Farmers is composed of independent farmers, outside the system of collectivized state farms, who own a fraction of the total cultivated land. In 1960 the Committees for the Defense of the Revolution, which now enroll most of the adult population, were created to maintain vigilance against "enemies of the revolution"; they are organized in every city, factory, and place of work and in many rural counties. The objective of the Federation of Cuban Women is "to raise the ideological, political, cultural, and scientific level of women in order to incorporate them into the tasks assigned by the revolution."

In 1973 Cuba's judicial system was reorganized. The People's Supreme Court, divided into five chambers, became the main body of the new structure. Its jurisdiction includes criminal offenses, civil and administrative offenses, crimes against state security, labour offenses, and military offenses. The Provincial Courts deal with cases that warrant sentences of up to six years' imprisonment. Below the provincial courts are municipal courts, usually the courts of first appeal.

Cuban defense is based on the revolutionary armed forces and is equipped with the most advanced weaponry in Latin America. The Ministry of the Interior is charged with the maintenance of public order and state security, rehabilitation of prisoners and prison management, and fire fighting.

The campaign against illiteracy

Education. The eradication of illiteracy was given high priority by the revolutionary government. Nationalization of all private schools was accomplished in 1961, and a fundamentally altered, state-directed education system was introduced. It includes general education, 12 or 13 grades (three or four years of intermediate education) preceded by a preschool stage; higher, or university, education; teacher-training education; adult education, directed toward the eradication of residual illiteracy and toward continued study by working people; technical education, parallel to secondary education; language instruction; and specialized

education. Education is free at all levels, with supplementary scholarships to cover living expenses and medical assistance. Education expenditures receive high priority, and the number of students enrolled has increased sharply from prerevolution days.

Health and welfare. Health care is state-operated through the Ministry of Public Health and is available free to the entire population. The effects of the exodus of physicians in the 1960s have been generally overcome. The ratio of hospital beds and physicians per thousand has greatly increased since then, and infant mortality and mortality rates have declined overall. Social security (old age, disability, and survivor pensions, and other monetary benefits) covers more than 90 percent of the labour force and is financed by an enterprise tax.

Private rental of urban real estate was prohibited under the Urban Reform Law of 1960, and it was made possible for families to own their homes by paying the current rental sum for not less than five or more than 20 years. Many families have acquired titles to houses and apartments in this fashion, and the rest pay a small percentage of their salary as rent to the state. A large number of rural families have achieved free use of formerly rented lands, and the traditional rural *bohío* ("hut") is being slowly replaced by more modern housing units. However, a decline in new dwelling construction from the mid-1960s to the early 1970s, combined with the destruction of old housing due to neglect, induced a severe housing shortage. The government later experimented with housing brigades, but the shortage has continued.

Homes for the aged are under the direction of the Ministry of Public Health, but the *círculos infantiles,* institutions for the day care of children under seven years of age, are run by the Federation of Cuban Women. The institutions are intended to free women to work. Physical education and sports, under a national body, are an integral part of Cuban education. The government fixes prices for, and distribution and sale of, foodstuffs and other goods. Rationing was introduced in the 1960s and expanded to cover most goods but then was cut back considerably.

(I.E./R.E.Cr./C.M.-L.)

Cultural life. The cultural life of the Cuban people has undergone a major transformation since the revolutionary government came to power. The government believes that mass culture is essential to the fulfillment of its economic and social aims and since 1959 has played a leading role in cultural life. Since the creation of the Ministry of Culture in 1976, this role has expanded to include a network of professional and amateur cultural organizations throughout the island. Cultural institutions before 1959 were generally limited to Havana (and, to a lesser extent, the provincial capitals) and were almost entirely privately endowed. Before 1959 Cuba had 100 libraries and six museums; today it has approximately 2,000 libraries and 250 museums. The Ministry of Culture directs a program of eduction in music, plastic arts, ballet, dramatic arts, and modern dance, culminating in the university-level Higher Institute of Art. More than 200 neighbourhood cultural centres (*casas de cultura*) offer workshops in all branches of the arts.

The Ministry of Culture

Folk culture. In 1959 the Institute of Ethnology and Folklore was created within the Academy of Sciences of Cuba, with the aim of collecting and classifying the Cuban cultural heritage. It formed the National Folklore Group, which performs Afro-Cuban dances throughout Cuba and abroad and gives international folklore laboratories each year. The activities of the folklore group are complemented by the Institute of Literature and Linguistics of the Academy of Sciences. The revolutionary government has made a special effort to promote study of the African roots of Cuban culture. The Guanabacoa Museum is the main repository of Afro-Cuban artifacts.

Literature. A national Cuban literature first began to emerge after the end of the 18th century. In the early 19th century several writers gained prominence espousing intellectualism and the concept of freedom. These ideas gained perhaps their greatest intensity in the writings of José Martí, a Cuban of poor Spanish background who led the Modernist movement in Cuban literature. He inspired a whole school of writing devoted to winning freedom from Spain. Writers whose works reflected social protest in the pre-Castro period include Nicolas Guillén, a leader in founding the Afro-Cuban school of literature, and Jose Z. Tallet, both activist poets. In the contemporary period, short stories became the predominant prose form, but some novelists, including Alejo Carpentier and José Lezama Lima, the latter also a poet, produced notable works. Many literary magazines have been established, and the National Union of Cuban Writers and Artists has a large membership that promotes literature and the arts.

Visual arts. Cuba has galleries, art museums, and community cultural centres that regularly display the works of Cuban painters. The most important are the National Museum of Fine Arts, the Haydée Santamaría Gallery of the House of the Americas, the Gallery of Havana, and the Fortress Castle. There Cuba's foremost contemporary artists—Wifredo Lam, René Portocarrero, Mariano Rodríguez, Servando Cabrera Moreno, Raúl Martínez—share space with younger artists. The Ministry of Culture provides most of the materials needed by artists and also guarantees jobs to graduates of the Higher Institute of Art. Painters in Cuba tend to work in many genres: they design fabrics (called by the trade name Telarte); sets for movies and theatre; and posters for films, books, cultural events, and community campaigns. The posters are one of Cuba's best-known cultural exports. The Ministry of Culture promotes numerous art exchanges and sends exhibits of Cuban art throughout the world. As with film, the government works to promote Third World art, primarily through the Havana Biennial, which started in 1984.

Museums and galleries

Music and ballet. Cuban music has Spanish and African roots, a blend that has contributed to a uniquely Cuban sound in both traditional and popular music. The Cuban *son, rumba, guaracha, bolero, danzón, conga,* and *cha-cha-cha,* as well as the Nueva Trova (New Minstrels) movement, have influenced much of the hemisphere. Musicians Pablo Milanés and Silvio Rodríguez, among the founders of the Nueva Trova movement, are acclaimed composer-singers. Festivals of Cuban music and song are held throughout the year, encompassing works of every genre from every period, including the internationally popular Afro-Cuban jazz. Classical music plays a relatively minor role in the Cuban musical scene, but there is a National Symphony Orchestra, which has a chamber orchestra and instrumental ensembles.

One of Cuba's foremost contemporary artistic figures is Alicia Alonso—a dancer of international acclaim, the prima ballerina and founder (1948) of the National Ballet of Cuba, and the head of its school. The Ballet of Camagüey, under the direction of Fernando Alonso, was established in 1971, and a second Havana company was founded in the mid-1980s. Besides classical ballet, there is the Modern Dance Company in Havana, the Tumba Francesa (a black folk group) in Santiago de Cuba, and dozens of smaller troupes.

Theatre. Cuban theatre has been state-supported since 1959, coming mostly under the direction of the Ministry of Culture. There are several national dramatic groups, whose directing councils create their own repertoire, most importantly the Studio Theatre. Provincial theatre groups are also well established. Cuban theatre reached a new maturity in the 1980s, producing plays focusing on contemporary social problems as well as developing efforts to integrate music and dance. National and international theatre festivals feature Cuban companies and troupes from the rest of the Americas. The National Theatre has an excellent library, and House of the Americas (Casa de las Américas), an international cultural institution, sponsors regular *encuentros* (meetings) with theatre professionals. Increasingly, Cuban theatre troupes travel abroad as part of an active exchange program.

Film. Cuban filmmaking since 1959 has been stimulated by the Cuban Institute of Cinematographic Art and Industry, which produces feature and documentary films. The institute also has an extensive film library, and its movie house, the Charles Chaplin Theatre, regularly shows the best of both world and Cuban cinema. The institute provides a variety of support services throughout

the hemisphere and sponsors an annual Festival of Latin American Film, Television and Video, which is the largest in Latin America. The New Latin American Cinema is headquartered in Havana under the direction of the Colombian writer and Nobelist Gabriel García Márquez.

Recreation. Sports activities are generally under the direction of the National Institute of Sports, Physical Education, and Recreation. A major objective of the ministry is to provide opportunity for everyone to participate in sports and recreational programs. Baseball is the national sport and is widely played throughout the country. Boxing is another popular sport, often dominated by Cuba in the Olympic Games. Cuba has carnivals, but, unlike other Latin-American countries', these celebrations are not based on religious holidays. The main carnival period is the last two weeks of July and is celebrated most notably at Santiago de Cuba. During this period Cuba celebrates what is perhaps its most significant holiday, commemorating Castro's attack on Fort Moncada, July 26, 1953.

Baseball

Press and broadcasting. The mass media in Cuba are government organs. The three main newspapers are the Communist Party daily, *Granma; Juventud Rebelde,* the paper of the Communist Youth; and *Trabajadores,* published by the Cuban Federation of Workers. These are supplemented by provincial newspapers, such as the *Tribuna de la Habana* and *Sierra Maestra* in Santiago de Cuba, that focus on local issues. Among the most widely read magazines are the weekly *Bohemia,* which covers all aspects of the news and is the oldest periodical in Cuba; the monthly *Opina,* aimed at a younger audience, with information on available consumer goods; and *Mujeres,* published by the Federation of Cuban Women. A number of specialized cultural magazines and newspapers also have wide readerships.

Two television stations broadcast nationally, and there are several national radio networks and one international; all of these are administered by the Cuban Institute of Radio and Television. Programming generally includes news, sports, educational programs, and serials. (S.H.L.)

For statistical data on the land and people of Cuba, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

At the time of the Spanish exploration of Cuba, the native population formed two groups totaling 50,000. The Ciboney and Guanahatabey occupied western Cuba. The more numerous Taino, who occupied the rest of the island, were highly developed agriculturalists and a peaceful people, related to the Arawakan peoples of South America who had migrated to the Greater Antilles. Their houses, called *bohíos,* formed villages ranging from single families to communities of 3,000 persons. They had pottery, polished stone implements, and religious spirits called *zemis,* which were represented by idols of wood, stones, and bones. The Taino diet included potatoes, manioc, fruits, and fish.

Pre-Columbian Cuba

The Spanish colonial regime. Christopher Columbus discovered Cuba for Spain during his first voyage, on Oct. 27, 1492. Diego Velázquez began permanent settlement in 1511, founding Baracoa on the northeastern coast with 300 Spaniards and their African slaves.

Within five years the island had been divided into seven municipal divisions, including Havana, Puerto Príncipe, Santiago de Cuba, and Sancti Spíritus. Each municipality had its own cabildo, or town council, governing its legal, administrative, and commercial affairs. From 1515, elected representatives of each cabildo formed a body that defended local interests before the royal council, especially matters such as the slave trade and the encomienda (a system that granted to conquistadors a certain number of Indians in a specified area from whom tribute could be exacted). A bishopric, subordinate to Santo Domingo, was founded at Baracoa in 1518 but later moved to Santiago de Cuba.

The island's limited gold deposits discouraged early settlement. The colony became a staging ground for the mainland exploration of Yucatán, Florida, and the Gulf Coast. Such expeditions as that of Cortés, which attracted 400 Spaniards and 3,000 Indians, depleted the colonial population. The small number of permanent resident Spanish colonists used the Indians in encomienda. But by 1550 the encomienda was no longer feasible, because the island's Indian population had declined dramatically to about 5,000 because of social dislocation, maltreatment, epidemic diseases, and emigration. Throughout the 17th century, colonial life was made more difficult by the ravages of hurricanes, epidemics, the attacks of rival European countries trying to establish bases in the Caribbean, and pirates. By 1700 peace had returned, and the population had grown to about 50,000. The *flota* system (regularly scheduled fleets between Spain and Spanish America) increased the commercial and strategic importance of Havana, while ranching, smuggling, and tobacco farming occupied the colonists. The administrative costs depended, however, on irregular subsidies from New Spain until 1808.

The 18th century brought intensified agricultural development. The main changes came with the growing dependence on sugarcane cultivation and the importation of African slaves for work on the plantations. In 1740 the Havana Company was formed to stimulate agricultural development by increasing the importation of slaves and regulating the export trade. The company was unsuccessful, selling fewer slaves in 21 years than the British sold during a 10-month occupation of Havana in 1762. Reforms of Charles III of Spain at this time further stimulated the development of the sugar industry.

Dependence of the economy on sugarcane

Between 1763 and 1860 the island's population increased from less than 150,000 to more than 1.3 million. Slaves made the most dramatic growth, increasing from 39,000 in the 1770s to some 400,000 in the 1840s. In the 19th century Cuba imported more than 600,000 Africans, most of them after an Anglo-Spanish agreement to terminate the slave trade in 1820. The Cuban insistence on the slave trade raised considerable diplomatic controversy between Spain and Great Britain between 1817 and 1865.

In 1838–80 the Cuban sugar industry became the most mechanized in the world, utilizing steam-powered mills and narrow-gauge railroads. Expanding *ingenios* (sugar mills) dominated the landscape from Havana to Puerto Príncipe, expelling small farmers and destroying the island's famous large hardwood forests. By 1850 the sugar industry accounted for 83 percent of all exports, and in 1860 Cuba produced nearly one-third of the world's sugar. The sugar revolution propelled a new rich class of slave owners to political prominence. Mexican Indians and Chinese augmented the labour force, and in 1865 the African slave trade ended, although slavery was not abolished until 1886.

The demands of sugar—labourers, capital, machines, technical skills, and markets—strained interracial relations, aggravated political and economic differences between metropolis and colony, and laid the foundation for the break with Spain in 1898. Spanish colonial administration had been corrupt, inefficient, and inflexible. The United States had shown a lively and growing interest in the island, and expeditions by U.S. filibusters won support in the United States, especially in the Southern slave states. After the 1860s the United States tried many times to purchase the island.

Spain's failure to grant political autonomy, while increasing taxes, led to the outbreak of the first war of independence—the Ten Years' War (1868–78)—which led to a military stalemate. The rich sugar producers of western Cuba and the vast majority of the slaves failed to rally to the nationalists, themselves divided over the questions of slavery, complete independence, and annexation to the United States. Spain promised to reform the island's political and economic system at the Convention of Zanjón (1878). Many Cubans, including the nationalist leader Antonio Maceo, however, refused to accept the Spanish conditions.

The Ten Years' War

By 1895 the political and economic crisis had grown more severe. U.S. investment had reached $50 million, and its annual trade with Cuba amounted to about $100 million. Cuban political organizations in exile were coordinated and mobilized by the poet and propagandist José Martí. War broke out again on Feb. 24, 1895.

Spain deployed more than 200,000 troops. Both sides killed civilians and burned estates and towns. By 1898 commercial activity had come to a standstill. Excited by the "yellow press" and a mysterious explosion aboard the USS *Maine* in Havana's harbour, the United States declared war on Spain on April 25, 1898. In August Spain signed a peace protocol in Washington, ending hostilities.

U.S. occupation, 1899–1901. Cuban independence, granted by the Treaty of Paris (Dec. 10, 1898), began Jan. 1, 1899, under U.S. occupation. The military governor, General John Brooke, tried to exclude Cubans from government. He disbanded the Cuban army and conducted a census before being replaced by General Leonard Wood, a former military governor of Santiago City. Wood sought to mitigate political division and supervised elections that gave Cuba its first elected president, Tomás Estrada Palma.

The military occupation restored normality. The Americans built a number of schools, roads, and bridges; they modernized Havana and deepened its harbour. But Americans were primarily interested in preparing the island for incorporation into the U.S. economic, cultural, and educational systems, and the franchise was designed to eliminate Afro-Cubans from politics. The Platt Amendment (1901) gave the United States the right to oversee Cuba's international commitments, economy, and internal affairs and to establish a naval station at Guantánamo Bay.

The Platt Amendment

The Republic of Cuba, 1902–58. A republican administration that began on May 20, 1902, under Tomás Estrada Palma faced difficulties over U.S. influence. Estrada Palma tried to retain power in the 1905 and 1906 elections, which were contested by the Liberals, leading to rebellion and a second U.S. occupation on Sept. 29, 1906. U.S. secretary of war William Howard Taft failed to resolve the dispute, and Estrada Palma resigned. For the United States Charles Magoon administered a provisional government of Cuban civilians under the Cuban flag and constitution. An advisory law commission revised electoral procedures, and on Jan. 28, 1909, Magoon handed over the government to the Liberal president, José Miguel Gómez. Meanwhile, Cuba's economy grew steadily, as sugar prices rose continually until the 1920s.

The Gómez administration (1909–13) set a pattern of graft, corruption, maladministration, fiscal irresponsibility, and social insensitivity—especially toward Afro-Cubans—that characterized Cuban politics until 1959. The Afro-Cubans, led by Evaristo Estenoz and Pedro Ivonet, organized to secure better jobs and more political patronage and to protest a ban of political associations based on colour and race. In 1912 government troops put down large demonstrations in Oriente. The pattern of corruption was followed by Mario García Menocal (1913–21), Alfredo Zayas (1921–25), Gerardo Machado (1925–33), Fulgencio Batista (through puppets 1934–39 and himself 1940–44 and 1952–59), Ramón Grau San Martín (1944–48), and Carlos Prío Socarrás (1948–52). Machado and Batista, who overthrew Machado in 1933 with U.S. support, were the most notorious, holding power through manipulation, troops, and assassins.

The income from sugar was augmented by vigorous tourism based on hotels, casinos, and brothels; Havana became especially attractive during the years of U.S. Prohibition (1919–33). Yet the prosperity of the 1920s, '40s, and '50s enriched only a few Cubans. For the majority, poverty (especially in the countryside) and lack of public services were appalling: with a national per capita income of $353 in 1958—among the highest in Latin America—unemployment and underemployment were rife, and the average rural worker earned $91 per year. Foreign interests controlled the economy, owning about 75 percent of the arable land, 90 percent of the essential services, and 40 percent of the sugar production. Nevertheless, there was no widespread discontent on Jan. 1, 1959, when Fidel Castro supplanted Batista.

The Castro regime. Batista's fall resulted as much from internal decay as from the challenges of Fidel Castro's 26th of July Movement (commemorating Castro's attack on the Moncada military fortress in Santiago on July 26, 1953) or from the Federation of University Students (later absorbed into the Young Communists Union) and other groups. Castro had been a candidate in the aborted elections of 1952. His defense of his part in the Moncada attack, edited and published as "History Will Absolve Me," was a political manifesto. Released from prison in 1955, Castro and some friends went to Mexico to prepare for the overthrow of the Cuban government. An enlarged group, including the Argentinian revolutionary Ernesto (Che) Guevara, landed in Cuba in December 1956 and were almost annihilated in their first attack. From the Sierra Maestra the survivors fought a guerrilla campaign. When the Fidelistas took control on Jan. 1, 1959, they numbered fewer than 1,000.

The 26th of July Movement

The 26th of July Movement had vague political plans, relatively insignificant support, and totally untested governing skills. They quickly forged a following from among the poor peasants, the urban workers, the young, and the idealistic of all groups. The Communist Party of Cuba, dating to 1925, assumed the dominant political role, and the state modeled itself on the Soviet-bloc countries, becoming the first socialist state in the Americas.

The first stage of the new regime was dominated by the progressive dissolution of capitalism, between 1959 and 1963. In those confusing and difficult years, the government eliminated the remnants of Batista's army as well as the former labour unions, political parties, and associations of professional persons and farmers. New institutions emerged: the Confederation of Cuban Workers (reconstituted 1970), the National Institute of Agrarian Reform (founded 1959), the Cuban Institute of Cinematic Art and Industry (1959), the Central Planning Board (1960), the Committees for the Defense of the Revolution (1960), the Federation of Cuban Women (1960), the National Association of Small Farmers (1961), the Revolutionary Armed Forces (1961), the National Union of Cuban Writers and Artists (1961), the Young Communists (1962), and others.

Nationalization of U.S. interests

The nationalization of hundreds of millions of dollars in U.S. property and private businesses provoked retaliatory measures by the U.S. government, including a trade embargo, an unsuccessful invasion by Cuban exiles at the Bay of Pigs in south central Cuba (April 1961), and unexecuted plots to assassinate Castro.

Within Cuba the erratic drift toward socialism and the growing economic dependence on the Soviet Union divided both the leadership and the country at large. Hundreds of thousands of Cubans, especially from among the skilled and the wealthy, emigrated to Spain, the United States, and other countries. The National Institute of Agrarian Reform tried and failed to diversify the economic base, and the constant mobilization for war frustrated effective long-term planning. Attempts to foment revolution elsewhere, especially in the Dominican Republic, Venezuela, and Bolivia, alienated Cuba from most of the other Latin-American states.

Castro visited the Soviet Union during 1963, but the next two years witnessed a period of ideological instability as the government consolidated its domestic position. A second agrarian reform ended attempts to diversify the economy. Shortages became acute. A professional army replaced the militias as the bastion of national defense. The meeting of Latin-American communists in Havana in November 1964 and the civil war in the Dominican Republic in April 1965 (which brought U.S. military intervention) renewed the Cuban intent to export their revolution.

Between 1965 and 1970 the revolution experienced a third, more radical phase. Cuba began to assume a significant leadership role among the so-called Third World countries; in 1979 Cuba was host to the summit conference of nonaligned nations and was its chairman from 1979 to 1982. By 1968 there was a strong campaign against bureaucrats and a renewed attack on private property, as hundreds of small businesses were nationalized. Military officers moved into the highest ranks of government, industry, and the party.

Material conditions improved slightly during the 1970s. The revolution institutionalized itself along orthodox Soviet lines. Bottlenecks and shortages were substantially eliminated, and diplomatic isolation gave way to technical, commercial, or military assistance between Cuba, the Soviet Union, and the states of Africa, Latin America, and

Political reorganization

the Caribbean. The political system was reorganized. A new family code was introduced in 1975, and the following year a new constitution and a new electoral code created 14 provinces (instead of six) and 169 municipalities (including the Isle of Pinos, now Isle of Juventud). Fidel Castro became president of the Council of Ministers and the State Council (the latter office combining the offices of president of the republic and prime minister). Nationwide elections in 1976 returned municipal assemblies, which then elected members to the provincial assemblies and the National Assembly.

Castro reestablished full diplomatic relations with the Soviet Union in 1960, and thereafter that country was the major trading partner and source of funds and military supplies for Cuba. From 1960 the Soviet Union bought the major portion of the Cuban sugar crop, generally at a price above that of the free world market. Soviet aid to Cuba in loans, petroleum, war matériel, and technical advice amounted to several billions of dollars annually. The dominant Soviet role forced the Cubans to support the Soviet Union in its dispute with China, although pro-Chinese sentiment was strong in Cuba in the mid-1960s. In 1968 Cuba endorsed the Soviet invasion of Czechoslovakia, but Cuba lost considerable influence among the nonaligned nations when it supported the Soviet invasion of Afghanistan in 1979.

Soviet military support was crucial in the early years, and Soviet maneuvers often aroused strong antagonism from the United States. The installation of Soviet missiles in Cuba in 1962 seemingly brought the United States and the Soviet Union to the brink of war as the U.S. government set up a naval quarantine of the island and demanded the removal of the missiles. In 1979 the United States objected to the presence of Soviet combat troops in Cuba.

The wide-ranging Cuban foreign policy relied on Soviet support. In the late 1960s the Cubans began to redefine themselves as an "Afro–Latin-American people." By the 1980s this new definition was accompanied by assistance to a number of countries in Africa, Latin America, and the Caribbean. Cuban military assistance probably determined the outcome of the civil wars in Angola and Ethiopia, but civilian personnel made contributions elsewhere in Asia and Latin America.

A new twist was added to Cuba–U.S. relations by the emigration of Cubans in 1980 to the United States. Emigration restrictions were lifted on April 21, 1980, and not reimposed until September 26, during which time some 125,000 Cubans left the port of Mariel by boat for the United States. Among these, it turned out, were a number of criminals and mentally ill persons, whose repatriation became another point of contention. In 1987 the two countries signed an agreement allowing for the emigration of 20,000 Cubans annually to the United States and for the gradual return of the undesirables to Cuba. In the meantime Cuba–U.S. relations had worsened when the United States invaded the island of Grenada, in 1983, killing a number of Cubans and expelling the remainder of the Cuban aid force from the island.

U.S.–Cuba emigration agreement

Although some overall improvement in Cuba–U.S. relations could be noted since the revolution, the U.S. trade embargo imposed in the early 1960s remained essentially in force, and there were few signs of thaw into the late 1980s. U.S. probes concerning the condition of political prisoners in Cuba and the propaganda broadcasts beamed from the U.S. Radio Martí since 1985 continued to cause agitation. Cuba began a phased withdrawal of its troops from Angola in 1989.

Cuba continued to lean heavily upon the Soviet Union for support of its economy, although that country also tightened its trade policy with Cuba, compounding already stringent economic conditions on the island. Cuba–Soviet relations deteriorated somewhat as a result of the liberalization in the late 1980s of Soviet political, economic, and social policies, which the Cuban hierarchy did not approve, despite movement in that direction by some other Eastern-bloc nations. (F.W.Kn.)

For later developments in the history of Cuba, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Dominican Republic

The Dominican Republic (República Dominicana) occupies the eastern two-thirds of Hispaniola, the second largest island of the Greater Antilles archipelago in the Caribbean Sea. Haiti, also an independent republic, occupies the western third of the island. Hispaniola lies between the islands of Cuba to the west and Puerto Rico to the east and is situated about 670 miles (1,080 kilometres) southeast of Florida and 310 miles north of Colombia and Venezuela. The northern shores of the Dominican Republic are washed by the Atlantic Ocean, while the southern shore is bordered by the Caribbean Sea. Between the eastern tip of the island and Puerto Rico runs a channel called the Mona Passage. The republic has an area of 18,704 square miles (48,443 square kilometres; including 63 square miles of adjacent islands). The capital is Santo Domingo.

The country, although small, occupies a strategic position on major sea routes leading from both Europe and the United States to the Panama Canal. Between 1930 and 1961 the republic's history was dominated by the repressive dictatorship of Rafael Trujillo—a ruler who nevertheless maintained internal stability, liquidated the national debt, and introduced a measure of prosperity and modernization. Yet, the human costs were excessive. Since the Trujillo regime large numbers of Dominicans have remained in poverty with little promise of improvement.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Topography is complicated and varied and includes five distinct highland or upland areas running along a northwest-to-southeast axis. The Duarte Peak in the major Cordillera Central (Central Highlands) is the highest mountain in the West Indies, rising to a height of 10,417 feet (3,175 metres). There is also a smaller range, the Cordillera Septentrional, which runs parallel to the northwest coast, and two lesser ranges in the southwest. Another minor upland area, the Cordillera Oriental (Eastern Highlands), lies in the northeastern portion of the country. The extreme northwest and the extreme southwest are dry, low, and desertlike. The southeastern region consists of rolling lowlands.

Major mountain ranges

Drainage. The Yaque del Sur River empties into the Bay of Neiba, draining the Cordillera Central to the south, while the Yaque del Norte River drains the northern slopes, flowing into the Bay of Monte Cristi. The eastern part of the island is drained by the Yuna River, which flows into Samaná Bay, and by the Ozama River, the mouth of which is near Santo Domingo, on the south coast. The salt lake of Enriquillo, about 23 miles long and 11 miles wide at its widest point, located near the Haitian border is the country's largest natural lake.

Soils. Soils vary, but those of the upland areas are mostly of residual origin, deriving from metamorphic and sedimentary rocks. Soils in the lowlands are of recent alluvial origin, except in the southeastern savanna (grassland) area, where they consist of sedimentary deposits of recent marine origin. In general all the soils are quite fertile except in the far southwest, in the Pedernales region, where the sedimentary soils are relatively barren.

Climate. The Dominican Republic lies well within the tropical zone, but the hot, moist climate typical of this zone is tempered in many areas by the altitude and in other areas by the insular character of the republic and by the northeast trade winds that blow steadily from the Atlantic all year long. The heaviest precipitation is in the mountainous northeast, where the average rainfall is more than 100 inches (2,540 millimetres) a year. As the trade winds pass over the various mountain ranges, they lose their moisture until, in the far western part of the country, along the Haitian border, less than 30 inches of rain falls annually. The Dominican Republic is in danger of damage from tropical storms and hurricanes, which originate in the mid-Atlantic and southeastern Caribbean from August until October each year.

Temperatures

In spite of local variations, the country as a whole enjoys a relatively mild and pleasant climate. The annual mean temperature is 77° F (25° C). Temperatures rarely rise above 90° F (32° C), and, even in the heart of the highest

central mountain range, the overall mean is only about 69° F (21° C). Very cold temperatures are unknown.

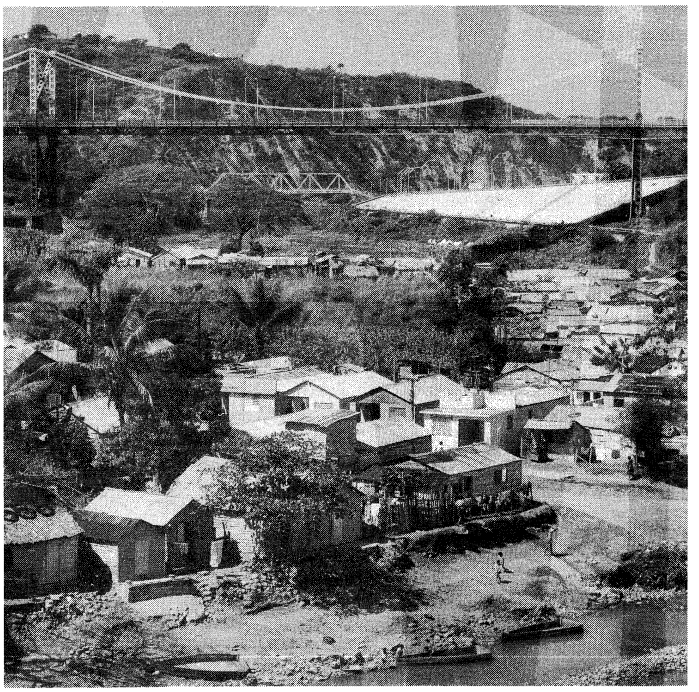
Plant and animal life. Vegetation varies considerably. The mountains are still largely forested with pines and hardwoods, although during the past century the lower and more accessible slopes were practically denuded of trees by commercial lumbering. In the drier regions low shrubs and scrubby trees predominate, but, as rainfall increases, grasslands and dense rain forests occur. The royal palm appears prominently throughout much of the country.

Forested areas

Cultivation of a wide variety of crops has largely replaced the natural vegetation in many areas—particularly in the more fertile upland valleys and on the lower mountain slopes. Mangrove swamps line some coastal areas, while elsewhere, particularly along the northern shore, sandy beaches of great beauty are to be found.

Wild animal life is not abundant, although for several centuries cattle and goats, introduced by the early Spanish colonists, ran wild on the grasslands and in the desert areas. Alligators are found near the mouths of the Yaque rivers and in the waters of Lake Enriquillo. A great variety of birds, including ducks, are hunted. Fish and shellfish inhabit the surrounding waters but have not been exploited commercially. Sport fishing, however, is an important tourist activity.

Hubertus Kanus—Photo Researchers



Housing along the Yaque del Norte River at Santiago, Dom. Rep., is spanned by a modern bridge.

Settlement patterns. The greatest population density, apart from the metropolitan area of Santo Domingo, is in the area referred to as the Cibao Valley, which extends across the north from the Samaná Peninsula in the northeast, north of the eastern and central mountain ranges and south of the Septentrional range, to Monte Cristi in the far northwest. South of the Cordillera Central lies an alluvial plain where rice is grown; its population is centred on San Juan de la Maguana. Many of the inhabitants of the town of Azua and its environs are the descendants of immigrants from the Canary Islands.

The southeastern savannas, now largely under sugar cultivation, are inhabited by settlers of predominantly European ancestry—descendants of cattle herders and families who owned small ranches. The southeastern coastline itself, however, is increasingly inhabited by blacks from other West Indian nations who have come to work on the sugar plantations, in the mills, or on the docks. Most of these are temporary or seasonal workers, many from Haiti.

Although villages exist, the more common rural settlement pattern is a scattered neighbourhood—perhaps clustered about a small store or church. Settlements frequently stretch along roadsides, with cultivated patches behind the houses; there are still many households so isolated from major or even minor roads that they can be reached only on foot or by horseback.

The Dominican Republic still has a large rural population, but it has diminished. Some 55 percent of the population is now estimated to live in urban sites, the country having experienced one of the highest urbanization rates in the world since the late 1950s. The largest of the urban centres is Santo Domingo, the capital city, followed by Santiago, which vies with the capital in political, cultural, industrial, and commercial importance. Other major population centres include San Francisco de Macorís and La Vega in the Cibao Valley, San Pedro de Macorís and La Romana on the south coast, and Puerto Plata on the north.

Urban centres

The Dominican Republic also has been subject to a high rate of out-migration, primarily to New York City. During the 15-year period 1965–80 it has been estimated that some 400,000 people, or about 8 percent of the total population, chose to emigrate, principally to improve their economic situation. Many of the emigrants found extralegal ways to gain entry to and remain in the United States. The outward flow of people alleviated the strain on home resources and at the same time provided remittances of cash and goods that assisted the welfare of those who stayed at home.

The people. The racial composition of the republic is predominantly mulatto (*i.e.*, of mixed European and African ancestry), the Amerindian element having been largely decimated by disease, warfare, and the effects of forced labour shortly after the arrival of the first Europeans.

The colonizing whites, mostly Spaniards, were joined in the 19th and 20th centuries by immigrants from East Asia and from such European countries as France, England, and Germany, as well as by small numbers of Sephardic Jews and Middle Easterners. The Mediterranean immigrants at first competed with Chinese peddlers and shopkeepers in the rural areas, but most have moved to the cities, where they occupy positions in commerce and industry. The Chinese particularly established themselves in the hotel and restaurant industry. A small group of Japanese developed truck farming in the Constanza Valley even before World War II, but their descendants are now to be found throughout the republic. Intermarriage among all of these groups has blurred, but not erased, their ethnic origins.

The exact African heritage of the large black population is unknown. Many of them arrived with or soon after the conquistadors, serving as slaves in the mines and the early sugar enterprises. Others drifted over from the French colony of Saint-Domingue (now Haiti) both before and after that area achieved its independence. Indeed, the entire country was briefly under Haitian rule early in the 19th century. There has long been a problem with Haitian squatters who cross over without immigration papers along the entire mountainous frontier between the two countries; other Haitians imported as cane cutters in the south have found ways to remain after their contracts expire. Thus, the African-American cultural influence is strong, especially among the lower socioeconomic classes, which tend to be darker-skinned than their more affluent compatriots.

The Spanish language has always been predominant, although English is becoming more common due to the massive emigration to the United States, which has been accompanied by continual visiting back and forth, plus some repatriation. Among Haitian immigrants, a French creole is spoken.

Language

Most of the population is Roman Catholic. That religion exerts a marked influence on cultural life at national, local, and family levels. The religious beliefs and practices of the rural populace are rooted in the cultures of both the early Spanish and African communities. A small percentage of the population is Protestant, although it has been growing in recent years; there are a few practitioners of Judaism and other religions.

The rate of population increase is greater than that of most of the West Indian islands. Birth rates in both the

Dominican Republic and bordering Haiti are substantially higher than those of nearly all of the neighbouring islands; but the death rate for the Dominican Republic, although higher than those of most Caribbean islands, is considerably less than that of Haiti.

The economy. Agriculture continues to be the basis of the Dominican Republic's economy, although that sector has been giving way to manufacturing. During the Trujillo regime, from 1930 to 1961, the Trujillo family largely controlled both agriculture and industry and, thus, the economy. Since 1961 most Trujillo enterprises have remained under governmental control. However, private industry has outstripped these in contributing to the gross national product, and there has been considerable diversification in recent years. As in many other Latin-American countries, this has been made possible through foreign loans, which have placed the country in serious debt.

Import duties and excise taxes constitute the largest part of public revenue. Other government income derives from income, sales, and road taxes, as well as vehicle licenses.

Trade unionism has not been important historically in the Dominican Republic but has become increasingly significant with the growth of the manufacturing sector since the mid-20th century. During Trujillo's rule, there was one so-called confederation of workers, which, in fact, was no more than a company union. After 1961 the labour movement developed rapidly, with unions tending to organize on the basis of political affiliation. Many are affiliated with international labour organizations.

There are also a number of employer associations, including the U.S. and Dominican chambers of commerce, as well as associations of cattle growers, sugar producers, and tomato growers. A number of factories and industries have associations for workers and management.

Resources. The most important resource of the Dominican Republic is the land, which is the foundation of the heavily agricultural economy. About one-third of the land is agricultural and under permanent cultivation. Pastures and meadows make up the largest land use category, more than 40 percent, and forests account for more than one-tenth of the land area. A number of minerals are known to be present but—with the exception of nickel, bauxite, gold, silver, gypsum, and iron ore—have not yet been highly developed commercially. Salt, largely from salt deposits near Enriquillo Lake, is also produced in commercial quantities. A smaller salt-producing enterprise, based on the evaporation of sea water, has also been of some importance at Monte Cristi. Other minerals of potential importance include sulfur, coal, titanium, molybdenum, cobalt, tin, oil, and zinc.

Arable land

The Dominican Republic is one of the relatively few sources of quality amber in the Western Hemisphere. Although it has not yet been extensively exploited, craftsmen produce distinctive jewelry for the local and tourist trade.

Unexploited biological resources are few, but proper management of agricultural lands, forests, and grazing areas has been encouraged to improve productivity.

Agriculture, fisheries, and forestry. Unlike most other Caribbean nations, the Dominican Republic produces much of its own basic food, as well as a considerable amount for export. The agricultural base of the economy is heavily dependent on the growth and processing of sugarcane. Raw sugar accounts for the largest share of export earnings, but there is also a smaller external market for rum, molasses, and bagasse. Tobacco, cocoa, coffee, rice, tomatoes, hides, bananas and other tropical fruits, root crops, and sorghum are other important agricultural exports. The tourist trade has stimulated the agricultural sector to improve the production of animal products, including chickens, eggs, pork, and beef, as well as dairy products in general.

Although the possibilities for a large-scale export-oriented fishing industry seem limited by the relative scarcity of marketable fish in nearby waters, the supply has been sufficient for local needs, and big-game fishing has been an additional tourist attraction. Forestry is of little consequence.

Industry. The main industry is sugar production. The processing of foods and beverages is well developed. Locally made textiles and finished clothing—particularly shoes, shirts, and hats—have replaced some imports. Wooden, metal, and plastic furniture has become important on both domestic and foreign markets. Some factories assemble products under contract with foreign, usually U.S., companies. There are many other small industries manufacturing traditional consumer goods such as soap, candles, rope, cigars, concrete blocks, cement, and tiles. Petroleum refining has grown in importance.

Manufacturing

Tourism has become, after sugar, the most important source of foreign exchange. For Americans, particularly, the country has become attractive for its liberal divorce code. But the favourable climate, beautiful beaches, restored Spanish colonial architecture, and relatively low prices draw their own clientele.

Trade and finance. The Dominican Republic's heavy dependence on sugar exports makes the balance of trade susceptible to worldwide fluctuations in sugar demand and prices. The government has taken steps to diversify the economy and to lessen the dependence on sugar exports. The largest strides in diversification have been made in the mining sector, in which exports of gold, silver, bauxite, and ferronickel account for about one-third of all export earnings, and in the light industrial sector, which exports clothing, electronic equipment, leather goods, and industrial staples. In addition to petroleum products, the country's chief imports are foodstuffs and manufactured goods. For cultural and geographic reasons, the country does not belong to any of the four major economic market integration systems of Latin America. The United States and Venezuela are the two chief suppliers of imports, while the United States receives the bulk of Dominican exports.

The monetary system is managed by the Central Bank, which maintains a gold and foreign currency reserve and administers exchange rates relative to the U.S. dollar. The private banking system is well developed, and several financial institutions, loan companies, and insurance agencies operate in the urban centres.

Transportation. Santo Domingo is the hub of a transport system that facilitates the flow of both people and goods to virtually all parts of the republic. Although the highway between the capital and the Cibao region is heavily traveled and in poor condition, secondary roads are generally good for a developing country.

Modern buses as well as a large fleet of private taxicabs provide transportation both within and between cities. Most goods are transported by truck to the important market centres.

A government-owned railroad line runs through the eastern half of the Cibao from La Vega to the port of Sánchez on Samaná Bay; it carries only freight. Most of the country's other railway lines are privately owned and serve the sugar industry. There is no passenger service.

There are international airports at Punta Caucedo, about 15 miles east of Santo Domingo, and at Puerta Plata on the northern coast. A secondary airport in Santiago handles small jets and propeller planes. Other airfields around the country are open to small civilian craft. The government-owned airline, Dominican Aviation Company, operates flights to San Juan (Puerto Rico), Curaçao in the Netherlands Antilles, New York City, and Miami and also makes local flights. Airlines now handle most passenger traffic to and from the Dominican Republic, but goods are exported and imported primarily by sea. Samaná Bay is one of the finest and largest natural harbours in the entire Caribbean. Until the 20th century the primary commercial ports lay along the northern coast, but, with the rise of the sugar plantations in the south, the ports of Santo Domingo, San Pedro de Macoris, and La Romana have increased in importance. Most general goods pass through Santo Domingo, but sugar is largely exported through the ports of San Pedro de Macoris and La Romana. Important historically, the ports of Monte Cristi and Sánchez in the north are now almost defunct. Only Puerto Plata in the north retains its commercial importance; it is still viable largely because of the tobacco, coffee, and cacao interests in the Cibao region. Barahona ships bauxite, gypsum, and salt but receives few imports.

Airports

Administration and social conditions. *Government.* The executive power is vested in a president, who is also com-

mander in chief of the armed forces, and a vice president. There is a Senate and a Chamber of Deputies. The Senate is composed of one representative from each province and one from the National District. The Chamber of Deputies reflects the size of the population but has no fewer than two representatives from each province and two from the National District. The present (1966) constitution, like its numerous predecessors, guarantees human rights, prescribes the division of governmental powers, and provides for popular sovereignty. It also accords suffrage to all Dominicans of either sex over 18 years of age, unless they are members of the armed forces or the police. During the Trujillo regime the armed forces were used to preserve the dictatorship, and even afterward the armed forces have played a role in government.

Historically, the various constitutions have provided special emergency powers for the president that have made it possible for the executive to supersede the legislative and judicial branches of the government should the president deem it necessary. While retaining provisions for emergency executive powers, each successive constitution has, nevertheless, expanded the social and economic rights guaranteed in earlier documents. A formal relationship between the Roman Catholic church and the government, incorporated in earlier constitutions, has now been eliminated. The right to private property is guaranteed but is limited by the right of the state to expropriate for the general good. Terms for national elected offices are four years, and incumbents may seek reelection.

Local government

The nation is divided into 29 provinces and one Distrito Nacional ("National District"). The central government administers the outlying provinces through governors appointed by the president. Each province elects representatives to the bicameral national congress. Internally, each province is subdivided into municipalities that elect their own councils and enjoy considerable local autonomy.

Justice. The legal system is based upon the Napoleonic Code. There is a series of regular courts, the judges of which are appointed by the Senate and may have no other public employment. These courts, with the exception of the land and commercial courts, have jurisdiction over both criminal and civil matters.

In criminal cases the judicial process begins with an investigation, generally conducted by an investigating judge. This is followed by the trial proper, conducted by the appropriate court. Appeals are made to a superior court and may finally be considered by the Supreme Court, composed of nine justices.

Education. The more isolated and rural the population, the less accessible are educational institutions, although education is free and compulsory for children between the ages of seven and 14. Primary schooling ordinarily lasts six years, although less may be offered in rural areas. This is followed by a two-year intermediate school and a four-year secondary course, after which a diploma called the *bachillerato* is awarded. Relatively few lower-income students, however, succeed in reaching this level, since the system is designed to encourage middle- and upper-income students to prepare for admittance to a university. Most wealthier students attend private schools, which are frequently sponsored by religious orders. Some public and private vocational education is available, particularly training in the field of agriculture, but this too reaches only a very small percentage of the population.

The universities

The Autonomous University of Santo Domingo, founded in 1538, is the oldest university in the New World. It is free of both governmental and religious control, although most of its funds are from the national budget. Costs are low and even poor students may attend, if they have been fortunate enough to have secured the requisite primary and secondary preparation.

The Pedro Henríquez Ureña National University, located in Santo Domingo, enjoys the support not only of the Roman Catholic church but also of private endowments and receives funds from the national government as well. The Madre e Maestra Catholic University in Santiago is similarly supported both by the Roman Catholic church and through private and public endowments. The Central University of the East in San Pedro de Macoris and the Technological University of Santiago were founded in 1970 and 1974, respectively.

Health and welfare. Health conditions among the poorer classes in both rural and urban zones are characterized by a generally unsanitary environment, inadequate health services, and poor nutrition. As a result, infectious and parasitic diseases are common, and infant mortality is high. Hospital and trained medical personnel are available only in the larger cities and towns. In the rural areas, home remedies and the professional services of practitioners of traditional local medicine are the only means of preserving or restoring health. Cases of severe illness may be taken to the nearest urban centre, where hospitalization is free. The tendency, however, is to take this measure only in extreme cases, often when death is already imminent.

Rural housing

Housing is considered by Dominican planners to constitute one of the most serious problems in the country. On the sugar plantations in the south, barracklike housing is provided for temporary workers, but more permanent employees frequently have their own small huts, or *bohios,* often on company-owned land. These may be little more than a lean-to of palm leaves and bamboo. Others, more sturdy, may have double-reed walls filled with rubble and plastered with mud.

In the Cibao, a relatively prosperous zone, houses are built solidly of palm board or pine and are frequently painted and decorated, with shutters and lintels in contrasting colours. Roofs are most often covered with sheets of zinc or tin but, in poorer households, may be thatched. A prosperous family may have a concrete floor, but most are of packed earth.

In the cities are the squatter settlements and poverty-stricken inner-city ghettos characteristic of most rapidly urbanizing underdeveloped countries. Dwellings may be built of cardboard, discarded inner tubes, and any other materials the inhabitants may scavenge.

Also in the cities are districts with well-appointed modern houses, occupied by members of the new commercial and industrial elite, as well as by the more traditional land-based oligarchy. Government programs, often funded with international loans, have financed housing construction for lower- and middle-income families.

Social conditions in the Dominican Republic resemble those found in other underdeveloped Latin-American nations. Small farmers rarely eke out more than a subsistence crop and most often must supplement this by the sale of handicrafts; products include baskets, pottery, rocking chairs, straw hats, and foodstuffs. These items are either sold to middlemen, who market them in towns, or are displayed and sold along the roads and highways.

By far the largest proportion of the population belongs to the lower-income group. This category includes the small farmers, as well as landless agricultural workers, itinerant merchants, and unskilled manual labourers.

Although most Dominicans, as well as some outside observers, have insisted that there is no traditional oligarchy, there nevertheless exists an intellectual and economic elite that constitutes a tiny fraction of the total population. An emergent middle-income group has also developed.

Cultural life. *Folk and fine arts.* It is difficult to define any particular and unique cultural tradition that may be labeled Dominican. There are, nevertheless, some cultural items worthy of special mention. Music, especially when accompanied by dancing, is important at all social levels and in all regions. The most typical forms are those with clear African antecedents, especially in their rhythms. There are also folk songs and tunes deriving from Spain and the Middle East. The merengue is a particularly popular dance, followed closely by the bolero. The guitar is probably the most popular instrument, but in some rural areas flutes and homemade marimbas are also common. Young people have been influenced by reggae and other modern African-American musical developments, as well as by the different forms of rock music. There is no highly identifiable Dominican national costume.

Musical tradition

The universities, as well as numerous private literary and cultural organizations, have long fostered an interest in the classical European arts of music, painting, drama, and literature. The country fosters a symphony orchestra, legit-

imate theatres, and art museums. Numerous local painters produce canvases ranging from exuberant Haitian-style primitives to abstracts and Impressionism.

Literature. A national spirit in Dominican literature began to develop during the 19th-century Haitian occupation and was led by Félix Maria del Monte, who was best known for his poetry. This trend was continued in the struggle against Spain, that experience being classically represented in the work of Manuel de Jesus Galván, who depicted Spanish settlers' cruel treatment of the Indians. In the early 20th century, writers such as Américo Lugo and Gaston Gernando Deligne were influenced by modernism and broke from the 19th-century styles and themes although the U.S. occupation revived some nationalist expression. Contemporary writers have focused to some degree on everyday Dominican life, but a social protest literature also emerged as represented by the works of the reformer Juan Bosch, written largely from exile.

Recreation. Important holidays are largely defined by the calendar of the Roman Catholic church, but the way in which they are celebrated reveals a mixture of official church and ancient folk traditions. Carnival, observed during several weeks preceding Lent, is especially colourful in Santiago. Masked and costumed men and boys circulate in the streets, hitting each other with inflated pig and goat bladders and chanting traditional rhymes intended to provoke each other to what is today usually only playful retaliation. Their masks are usually homemade and constitute a recognized art form in the country. During the final few days of carnival, organized and elaborately costumed groups of players of both sexes perform in the streets in return for handouts of rum and cash.

Organized team sports are popular, especially baseball, which was introduced by Americans during their occupation of the country in 1916–24. Major league players from the United States often spend their winters in the Dominican Republic, playing in local professional leagues. Many of the professional baseball players in the United States are themselves of Dominican birth. Cockfighting remains a traditional and popular spectator sport among men, although it is disapproved by the authorities.

Press and broadcasting. During the long Trujillo regime, freedom of expression in the press was severely restricted, but since that time the constitutional guarantees have been more or less upheld. Among the island's several daily newspapers, the most important are published in Santo Domingo and include *El Caribe,* founded in 1948, and *Listín Diario,* founded in 1889. A variety of U.S. newspapers and periodicals are also available. Radio and television broadcasting is supervised by the general director of telecommunications. The main stations are also in Santo Domingo, one of them—Dominican Radio Television—being owned and operated by the government. (N.L.G.)

For statistical data on the land and people of the Dominican Republic, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

At the time of Columbus' first landing in 1492, the Caribs, a people who had apparently originated on the South American mainland and migrated to the Greater Antilles (and for whom the Caribbean Sea is named), were preying upon the Taino (Arawak), who had previously settled there. These Indians were less advanced socially and culturally than were the large-scale Indian civilizations in Mexico, Guatemala, or Peru.

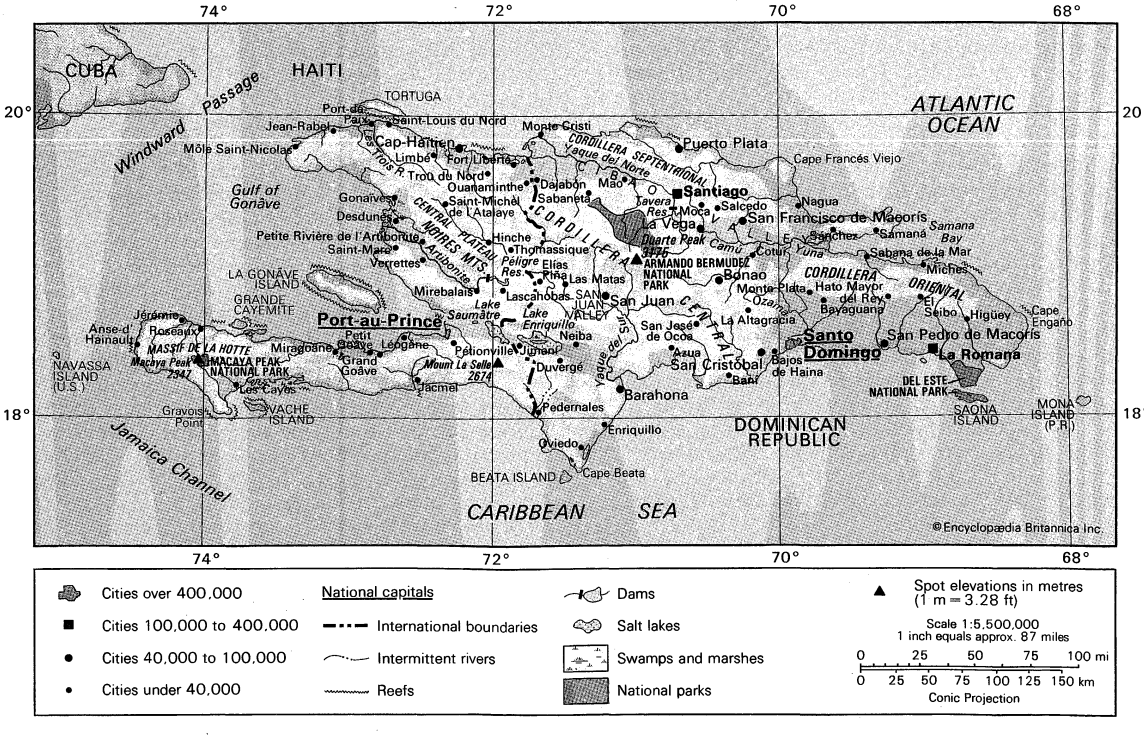
Hispaniola was visited by Columbus on his maiden voyage. A colony was established on the north coast, but the first settlers were slaughtered by the Indians. Returning, Columbus established a second colony; but reports of abundant gold farther south quickly led to the abandonment of the northern outpost and to the founding of Santo Domingo city on the Caribbean coast.

Period of Spanish rule

Colonial era. Hispaniola was the first area in the New World to receive the full imprint of Spanish colonial policy. The oldest cathedral, monastery, and hospital in the Americas were established on the island, and the first university was chartered there. The earliest experiments in Spanish imperial rule were conducted here. Class and caste lines were rigidly drawn; the Roman Catholic church served as the strong right arm of temporal authority. A cruel, exploitative, slave-based society and economy came into being. The first "revolution" in the New World was also recorded on Hispaniola.

During the first half century of Spanish rule, Hispaniola flourished, for its rich mines and lush lands yielded abundant wealth, and it served as the administrative centre for Spain's burgeoning American empire. But the more lucrative conquests of Mexico and Peru soon turned it into a poor way station. Its Indians were decimated, gold and silver were more easily available elsewhere, and the more ambitious Spaniards emigrated.

For the better part of the next three centuries, Hispaniola remained a neglected, poverty-ridden backwater of



the Spanish empire in the Americas. Successive raids by British, Dutch, and French marauders and buccaneers devastated the island still further. Socially and economically, it retrogressed. Eventually, French claims to the western third of the island were recognized, and a prosperous sugar-producing colony based on black slavery grew up in Saint-Domingue, the area that was later to become the independent nation of Haiti. As a by-product of Haiti's prosperity, the Spanish colony also experienced a modest boom in the 18th century.

In 1795, as a result of its defeat in the wars that had been raging in Europe, Spain ceded the eastern two-thirds of Hispaniola to France. Meanwhile, inflamed by the revolutionary currents then sweeping France and stirred to revolt by the inhuman conditions under which the slaves were forced to labour, a slave uprising had begun in Haiti. Led by Toussaint-Louverture, the blacks not only succeeded in throwing off French rule but soon overran the previously Spanish eastern end of the island as well, instilling terror in the white ruling class. With the aid of the British fleet, the Haitians were driven back, and in 1809 the colony was reunited with Spain. In 1821, following the lead of the countries on the mainland, the Dominican Republic declared its independence. The new republic comprised approximately the eastern two-thirds of the island.

The Dominican Republic to 1930. Within a matter of weeks, Haitian columns under Jean-Pierre Boyer (president of Haiti, 1818–43) again overran the entire island. Dominican historians have portrayed this occupation (1822–44) as cruel and barbarous. Haitians held the highest offices, closed the university, severed the church's ties with Rome, forced out the traditional ruling class, and all but obliterated the western European and Hispanic traditions. But Boyer also freed the slaves, and his administration was generally efficient.

Haitian occupation

In the 1830s Juan Pablo Duarte—known as the father of Dominican independence—organized a secret society to fight the Haitians, and in 1844, after a long struggle and aided by the outbreak of civil war in Haiti itself, independence was finally achieved. But Duarte and the other idealistic independence fighters were soon exiled, and the new nation quickly fell into less noble hands.

From 1844 until 1899 the Dominican Republic was dominated by a succession of dictatorial "men on horseback," who prevented the growth of democracy and who were not above selling out the country to foreign and commercial interests. Pedro Santana and Buenaventura Báez emerged as the two most prominent figures, alternating in the presidency for nearly 30 years. In order to ward off continuous assaults by Haiti, Santana returned the country to Spain (1861–65) and arranged to have himself named governor-general. After a series of battles, Spain withdrew its troops, and Báez approached the United States with a protectorate plan. President Ulysses S. Grant favoured annexation, but, after the questionable activities of a U.S. land-speculating company became public, the Senate failed to ratify the treaty.

During the 1870s the instability continued. Báez returned to the presidency for the fifth time, and the country's first, but short-lived, democratic government came to power. The instability culminated in the emergence of Ulises Heureaux, who dominated the country from 1882 to 1899. Heureaux presided over a period of unprecedented stability and national growth. New roads were built, production rose, and foreign capital entered, although under dictatorial auspices. Following Heureaux's assassination in 1899 the country returned to the chaotic politics of the past. New leaders took over and were in turn forced out of the presidential palace. Even the accession of the archbishop Adolfo Nouel to the presidency in 1912 failed to stem the disorder; within four months he, too, was forced to resign.

Meanwhile, the deteriorating financial situation plus the expanding interests of the United States in the Caribbean area had drawn that country ever deeper into Dominican affairs. The United States had replaced Europe as the major importer of the republic's products as well as the chief supplier of its imports. U.S. private investments were also rising rapidly. In 1905, threatened with the possibility that European creditors would use force to collect some unpaid Dominican debts, the United States took over the administration of the Dominican customs revenues. In 1916, as the fragile political structure collapsed again, the United States assumed complete control.

U.S. influence

During the U.S. occupation (1916–24), roads, schools, communications and sanitation facilities, and other projects were built. But the occupation forces had assumed arbitrary control and frequently abused their authority. In addition, the U.S. Marine Corps created a modern, unified, military constabulary that provided the instrument by which future Dominican strongmen could seize power.

In 1924, in U.S.-supervised elections, Horacio Vásquez was elected president. His rule eventually proved incompetent and corrupt, and the stock market crash of 1929–30 undermined the Dominican economy. In 1930 a revolution was launched against his rule. The military forces, now under the firm control of Rafael Trujillo, held the balance of power; but, rather than defend the government, they stood by and let the revolution succeed. Then Trujillo moved to take power himself.

The Dominican Republic since 1930. The dictatorship of Rafael Trujillo (1930–61) was one of the longest, cruelest, and most absolute in modern times. For more than three decades Trujillo ruled with an iron hand; virtually everything he touched had to belong to him. Trujillo dominated the armed forces, government, economy, church, education, every element of Dominican society. Under Trujillo, as under Heureaux, the economy prospered.

Following Trujillo's assassination in 1961, his heirs and followers attempted to remain in power, but they were also driven out, and the country embarked on a more democratic course. In 1963 Juan Bosch and his moderately reformist Dominican Revolutionary Party took power, the first democratically elected and progressive government in the country's history. Conservative forces remained strong, however, and after seven hectic months Bosch was overthrown. The country's return to conservative rule fomented a popularly based and democratic social revolution in 1965. Fearing a second Cuba, however, the United States again occupied the country (1965–66) and snuffed out the revolution. New elections were then scheduled.

Effects of Trujillo's assassination

The winner of the 1966 elections was Joaquín Balaguer, a former Trujillo puppet now presenting himself as a moderate conservative and a symbol of orderly change. The reelections in 1970 and 1974 of the conservative regime of Balaguer reflected the power of the business, commercial, and industrial oligarchy. Under Balaguer there were strong economic gains and some social reforms, but the level of tension remained high. In 1978 Balaguer was defeated by Antonio Guzmán Fernández and the Dominican Revolutionary Party. Guzmán moved cautiously to implement reforms, but conservative elements remained powerful and the economy fragile. Hurricane David devastated the country in 1979, and the faltering economy led to inflation, strikes, and depressed conditions.

Guzmán was succeeded by another Dominican Revolutionary Party candidate, Salvador Jorge Blanco, who served as president in 1982–86. The country thus was able to complete eight years of true democratic government, the longest in its history. But Jorge Blanco was faced with falling prices for sugar, widespread corruption in the government bureaucracy, a declining economy, and hence the need to impose an unpopular austerity program that produced strikes and food riots. As a result, the aging (and now blind) Balaguer was elected president again in 1986. Two years later, while out of the country undergoing medical treatment, Jorge Blanco was convicted of corruption that allegedly occurred during his presidency. Although Balaguer's advanced age and declining health gave rise to fears of instability and despite continuing economic difficulties, he was reelected in 1990, by an extremely narrow margin, over the former president Juan Bosch.

The Dominican Republic had developed a stronger base for democracy (political parties, interest groups, larger middle class, greater literacy) throughout the 1980s, but authoritarian impulses remained powerful. Balaguer's administration continued firmly in control despite periodic civil disorders protesting a stagnated economy, cost of living increases, and inflation. (H.J.Wi.)

For later developments in the history of the Dominican Republic, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Haiti

Haiti (Haitian Creole: Repiblik Dayti; French: République d'Haïti) is a republic in the Caribbean, situated in the western part of the island of Hispaniola. The native American Indian inhabitants called the island Ayti, meaning "Mountainous Land." Haiti's area of 10,714 square miles (27,750 square kilometres) occupies slightly more than a third of the island. The republic is located only 565 miles (910 kilometres) from the U.S. mainland. It borders the Windward Passage, a 50-mile-wide corridor between the northwest peninsula of Haiti and Cuba.

The coastline of Haiti is irregular and forms a long, slender peninsula in the south and a shorter one in the north. The two peninsulas are separated by the triangular-shaped Gulf of Gonâve, in the middle of which lies Gonâve Island. The border with the Dominican Republic runs north–south at a right angle with the main relief structures. The capital is Port-au-Prince.

Haiti won its independence from France in 1804, becoming the second country in the Americas, after the United States, to win freedom from colonial rule. It also became the world's first black republic, most Haitians being of African descent.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief, drainage, and soils.* Haiti is mountainous and rugged. Plains, which are quite limited in extent, are the most productive agricultural lands and the most densely populated areas. Rivers are numerous but short and most are not navigable. The seas around Haiti are noted for their striking coral reefs. The shores are generally rocky, rimmed with cliffs, and indented by a number of excellent natural harbours.

Chief mountain ranges

The backbone of the island of Hispaniola consists of four mountain ranges that extend in a west-northwest–east-southeast direction. In Haiti the most northern range occurs only on Tortuga Island ("Turtle Island"), which is shaped like a turtle shell. This island has an area of 69 square miles and in the 17th century was a stronghold of privateers and pirates from various nations.

Farther south the Dominican Cordillera Central becomes the Massif du Nord, a series of massive parallel ranges. On the top of one of its peaks, overlooking the city of Cap-Haïtien and the narrow coastal plain, stands the Citadelle Laferrière fortress built by King Henry Christophe at the beginning of the 19th century.

An interior basin, known as the San Juan Valley in the Dominican Republic and the Central Plateau in Haiti, occupies the centre of the island, south of the Massif du Nord. This plateau has an average elevation of 1,000 feet (300 metres), and access to it is difficult through winding roads. Through the basin and across a cut flows the Artibonite River, the longest river of the island. At Péligre a dam was built in the 1950s for the regulation of floods. A hydroelectric power plant began operation at Péligre in 1971. At the end of its course and before its delta mouth in the Gulf of Gonâve the Artibonite is used for partial irrigation of the triangular Artibonite Plain.

The third range, known as the Sierra de Neiba in the Dominican Republic and the Trou d'Eau Mountains (Chaîne du Trou d'Eau) in Haiti, towers over the narrow Cul-de-Sac Plain, which is immediately adjacent to Port-au-Prince and is occupied by the brackish Lake Saumâtre on the Dominican border. South of the Cul-de-Sac the fourth mountain range is called Sierra de Bahoruco in the Dominican Republic and Massif de la Selle in Haiti. It reaches its highest point at the Mount la Selle (8,773 feet). The prolongation of the range, the Massif de la Hotte (Massif du Sud), tops off at Macaya Peak (7,700 feet) farther west. Small plains lie north and south of these ranges.

Generally the mountains of Haiti are calcareous, although some volcanic formations can be found, particularly in the Massif du Nord. Karstic features (caves, grottoes, and subterranean rivers) are present in many parts of the country. Stairlike terraces on headlands are evidence of the correlative movements of the land and sea. A long fault line crosses the southern peninsula and passes just south of Port-au-Prince. Haiti is subject to periodic seismic activity, and Port-au-Prince and Cap-Haïtien were destroyed by earthquakes in 1751 and 1842, respectively.

The soils in the mountains are thin, and they lose fertility quickly when cultivated. The lesser hills have red clay and loam. Alluvial soils are found in the plains and valleys. Deforestation has caused much soil erosion.

Climate. Haiti has a warm, humid tropical climate characterized by daily variations of temperature that are greater than the annual variations. Average temperatures range from 75° F (24° C) in January and February to 83° F (28° C) in July and August. Temperatures vary with altitude. The village of Kenscoff at an elevation of 4,700 feet has an average temperature of 60° F (16° C), while Port-au-Prince, at sea level, has an average of 79° F (26° C). In winter frost can occur at high altitudes and mists are common.

Temperatures

Haiti is located on the leeward side of the island, which means that the influence of humid trade winds is not as great as in the Dominican Republic. The more humid districts are the windward northern and eastern slopes of the mountains. The leeward districts are drier, and some portions of the island receive less than 28 inches (700 millimetres) of rainfall per year. The northwest peninsula and Gonâve Island are particularly dry. There are two rainy seasons, lasting from April to June and from August to October. Some regions have only one rainy season, from May to November. Annual variations of rainfall can be large, causing droughts and famine. The southern peninsula is the part of the country that is most vulnerable to hurricanes. Hurricanes Allen (1980) and Gilbert (1988) were particularly destructive.

Plant and animal life. Much of the natural vegetation has been destroyed through clearing for agriculture, grazing, and the exploitation of timber during the last three centuries. This process has accelerated during the 20th century under population pressure. The virgin forests that once covered the country have been reduced to less than 10 percent of the land area. Patches of forest remain in the Massif de la Selle, where one can see stands of handsome pines, and in the Massif de la Hotte, where an evergreen forest with giant tree ferns and orchids is preserved on the steep slopes of Macaya Peak. The *bayahonde* (a type of mesquite), cacti, and acacia form thorn brushwoods in dry plains. On the coast the mangrove swamps have suffered from an overexploitation for firewood and charcoal.

With the retreat of natural vegetation, wildlife has lost its habitat and shelter. Wild boars, guinea fowls, and wild ducks are no longer present, but caimans still inhabit rivers of the southern peninsula and flamingos can be seen around Gonâve Island. Flamingos are hunted without restriction.

Little has been done to conserve plant and animal life. No national or regional parks for the preservation of natural life have been established. The lack of conservation measures has been particularly unfortunate for coral formations and the animal life that surrounds them.

Settlement patterns. Haiti is densely populated. About three-quarters of the people live in rural areas and are dependent upon agriculture. Rural densities are above what is considered safe for good management of the environment and for the well-being of the people, given available technology. Migration to the cities takes its toll on the rural population, particularly drawing the young but population is still increasing in the countryside.

High rural density

The plains are more densely populated, but the hills and the steep mountains are covered with cultivated plots and human settlements. Although only 30 percent of the land is considered suitable for agriculture, more than 40 percent is under cultivation. Most farms are very small and are worked by their owners. Rural settlement is generally of the dispersed type, sometimes organized along the roads. The basic element is the homestead around which small plots are arranged. The small wooden-frame houses vary in design according to the region but are everywhere enclosed within a compound of four mud-daubed wattle

walls. The roof is either thatched or made of corrugated iron sheets. The furniture inside most houses remains restricted to necessities. Wealth is invested in land and cattle or is used toward the cost of voodoo ceremonies or to pay the school fees for children. Individual means of transportation are nonexistent. Peasant women usually walk to the nearest market "town" (*bourg*) or in some cases use crowded public trucks and buses.

These *bourgs* are characterized by their administrative and political functions. Most have a Roman Catholic church, police barracks, a court of justice, and a general store surrounding a square. Peasants and particularly peasant men have always been wary of visiting the *bourg,* where abuses in the form of fines, taxes, and forced military enrollment were all too frequent in the past.

Real urban life is limited to the capital and to five or six large towns. Port-au-Prince has more than six times the population of the second city, Cap-Haïtien. Port-au-Prince was founded in 1749 by royal order and became the capital of the colony of Saint-Domingue in 1770 because it was thought that its central location was more suitable for future development, defense, and commerce than the position of Cap-Français (later Cap-Haïtien) on the north coast. Because of fires and war damages, the city has retained few buildings from the colonial period and the early 19th century. Wooden gingerbread-style houses are a testimony of Victorian influences in the formerly fashionable districts of Bois-Verna and Turgeau. Since the 1960s the city has expanded. Pétionville, a middle-class suburb in the hills to the west, is now part of the metropolitan area. The vast majority of Port-au-Prince residents live on meagre incomes, and the signs of poverty are ever present. The sight of the shantytowns that surround the city, the squalor associated with the markets, and the general lack of hygiene are often disturbing to visitors.

Urban architecture

Cap-Haïtien, which was once the capital of the colony, was founded in 1670. Its neat gridiron plan encompasses small blocks of old-fashioned houses with courtyards. The pace of life is much quieter than in Port-au-Prince, but its charm cannot conceal entirely the poverty demonstrated by hordes of beggars. The other major towns are Gonaïves, Les Cayes, and Jacmel.

The people. *Ethnic composition.* About 95 percent of Haiti's population is black; mulattoes account for most of the remainder, and whites make up a very small percentage. Because of its unique history, Haiti is different ethnically and culturally from other Latin-American countries and from the Spanish-speaking Caribbean nations. It shares with Guadeloupe, Dominica, Martinique, and Saint Lucia elements of a Creole culture, which is a mixture of Western and African influences.

Linguistic composition. Haitian Creole and French are the official languages. Creole is spoken by all Haitians and, with French, is used in drama, music, radio, television, politics, and religion. But written Creole is not widely accepted because the school system retains French as the main language of instruction. Creole is normally used in daily life, and French—mastered by perhaps 10 percent of the people—is used in more formal circumstances. Most of the vocabulary of Haitian Creole is derived from French, but its syntax is similar to that of some African languages and resembles the syntax of other creole languages of the Caribbean and the Indian Ocean.

Religions. The official religion of Haiti is Roman Catholicism, but the constitution allows the free choice of religion. About 80 percent of the population is Roman Catholic, and some 15 percent is Protestant. The Catholic clergy has gradually been "Haitianized" through the appointment of local priests and bishops. Although the official religion is respected, most Haitians are believers of voodoo (voudou, or *vodun*), a religion whose gods (*loas*) are derived from West African religions. Most practitioners do not find any contradiction between voodoo, which is above all a family cult, and Roman Catholicism.

Protestant ministers, however, consider their religion to be incompatible with the practice of voodoo. There are many Protestant sects because, in addition to the older denominations established in the early 19th century (Methodists, Episcopalians, and Presbyterians), new sects came to Haiti during and after the U.S. occupation (1915–34), including Baptists, Seventh-day Adventists, and Mormons.

Demographic trends. Population growth is quite rapid in Haiti. Since the beginning of the 20th century there has been a fourfold increase in population. This expansion took place despite high rates of mortality, particularly infant mortality, that are due to poor health services. Life expectancy is among the lowest in the Americas.

Population growth

Population growth is reduced by external migration. Since the early 1960s Haitians have been emigrating at a rate of some 50,000 per year. The main destinations for Haitian emigrants have been Cuba and the Dominican Republic. In the Dominican Republic there is a permanent population of persons of Haitian descent that lives under semiclandestine conditions. *Bracero* schemes allow temporary migrants for agricultural work and menial jobs. The United States has accepted qualified emigrant Haitian professionals since 1957 and industrial workers since 1965. Migration to the United States and Canada is still substantial. The high rate of emigration is due to poverty and to the narrowness of the job market.

The economy. Haiti is the poorest country in the Western Hemisphere. The limited resource base has been depleted, first through intensive colonial exploitation and later through unplanned development and corruption. Both the private sector and the state share responsibility for this state of affairs. A few multinational corporations are active in the country.

Resources. Gold and copper are found in the north of the country. Extraction of bauxite by Reynolds Metals Company at Miragoâne in the southern peninsula was discontinued in 1983. Drilling of sedimentary formations in the Gulf of Gonâve has not been followed by oil discoveries. Haiti is heavily dependent on energy imports. Hydroelectricity is not sufficient to satisfy current needs.

Agriculture and fishing. Agriculture is the largest sector of the Haitian economy, employing more than two-thirds of the labour force but accounting for barely one-third of the gross domestic product. Soil erosion and loss of fertility threaten the sustainability of agriculture. Cultivation on steep slopes poses a special problem since terraces are not usually constructed. Soils are often washed away after heavy rains. Recurrent drought and an absence of irrigation have kept production down. The ecological stress seems highest in the central region north and south of Port-au-Prince and in the northwest.

Export agriculture has traditionally been favoured by farmers and the state alike because it provides cash and hard currency. Coffee is the most important commercial product. The Haitian variety is an arabica coffee of the mild type. Coffee is sold through a system of intermediaries, speculators, and merchant houses, and in the past it was heavily taxed. Cocoa, mangoes, and essential oils for the cosmetics and pharmaceutical industries are also produced for export.

Haitian coffee

Sugarcane is the second major cash crop, but since the late 1970s Haiti has become a net importer of sugar. Only one sugar mill is still functioning, at Chancerelles, a suburb of Port-au-Prince. Most sugarcane is ground in rural distilleries that produce a substandard rum called clairin.

Population pressures have forced many farmers to concentrate on subsistence crops. Corn (maize), rice, sorghum, beans, and fruits and vegetables are the main food crops. Local food trade is conducted in rural markets and along the roads, but the size of these exchanges remains small.

Supply of food products has not kept pace with demand. The smuggling of food from the Dominican Republic has lowered consumer prices and injured Haitian farmers. Contraband is believed to account for 20 percent of Haiti's consumption. Following an outbreak of African swine fever in the early 1980s, the entire pig population was exterminated; this deprived many peasants of their only asset. New resistant breeds were subsequently introduced. Cattle breeding for milk and meat remains marginal, but there is some poultry production. Fishing is of limited importance because the fleet is ill-equipped.

Industry. The small domestic market has been a constraint on the growth of the manufacturing sector. After the overthrow of the Duvalier dictatorship in 1986 many

barriers to international trade were abolished and local industry has had to compete with imports from the Dominican Republic and the United States. Most industry is of the assembly type for reexport to the United States. The main items assembled are electronic components, baseballs, and clothing. The construction industry has flourished because of the high demand for housing.

Finance and trade. The financial situation of the country is precarious. The local currency, the gourde, is protected from devaluation by a convention with the United States signed in 1919 that established a fixed parity of five gourdes to one dollar. U.S. currency circulates freely in the country. The National Bank of the Republic of Haiti is the sole bank of issue and the principal commercial bank; there are a number of private and foreign banks. Total foreign debt is high. Government finances and the economy are largely dependent on aid from international agencies and from countries such as the United States, France, Canada, and West Germany.

Import and export figures reflect a structural trade deficit. Exports of coffee have continued to dwindle rapidly and exports of assembled goods vary from year to year according to competition. Major imports are food, machinery and vehicles, fuel, and textiles. About two-thirds of the external trade is conducted with the United States, the rest being shared by the other Caribbean countries, France and other western European countries, and Japan. Internal trade, which is very active, has been disrupted by smuggling.

Trading partners

Transportation. The roads from Port-au-Prince to Cap-Haïtien, Les Cayes, and Jacmel have been paved, but generally inland transportation, hampered by rough roads, is difficult and often unpredictable. Trucks and buses offer irregular and costly service from Port-au-Prince to the provincial towns. Police checkpoints on the roads slow down traffic. Coastal boat service is minimal except from Port-au-Prince to the Jérémie district.

The Port-au-Prince harbour was modernized in the 1970s and '80s, and container facilities there handle most of Haiti's foreign trade. The Cap-Haïtien harbour has also been upgraded. There are a number of minor ports. The international airport at Maïs Gâté, 10 miles north of Port-au-Prince, provides direct service to North and South America, Europe, and the Caribbean.

Administration and social conditions. *Government.* A constitution approved by referendum in 1987 incorporated features of the U.S. and French constitutions. It provides for a president who is directly elected by universal adult suffrage for a five-year term. The head of government is the prime minister, appointed by the president from among the members of the majority party of the parliament. The parliament consists of two houses, a Senate and a House of Representatives. Senators are elected for six years and representatives for four.

Hubertus Kanus—Photo Researchers

Vendors displaying their wares at the crowded Iron Market in Port-au-Prince, Haiti.

Law is based on the French Napoleonic Code, modified by legislation enacted during François Duvalier's presidency. The judiciary consists of four levels: the Court of Cassation, courts of appeal, civil courts, and magistrate's courts. Judges of the Court of Cassation are appointed by the president for 10-year terms.

The 1987 constitution provides safeguards against military rule. Military officers are disqualified from becoming candidates for political office, and the military and the police are no longer a combined force. In 1986 the National Security Volunteers—popularly known as the Tontons Macoutes—were officially disbanded. Nevertheless, the country remains highly vulnerable to military takeovers, which have played a large role in its history.

Education. Education is officially compulsory for children between the ages of seven and 13, but, because of a lack of facilities and staff, only about 40 percent of children in this age group attend school. More than 60 percent of the adult population is illiterate; the rate of illiteracy is higher in the countryside. The curriculum is based on the French model. This system has created a small elite, who have made distinguished cultural contributions, while most of the society remains relatively uneducated. In Port-au-Prince there is a small university, the State University of Haiti. Many students attend universities in Europe and North America.

Illiteracy

Health and welfare. Haiti has the highest rate of infant mortality in the Americas. The diet is notoriously insufficient, and infectious diseases abound because of improper shelter, unsafe water, and unsanitary living conditions. Malaria and acquired immunodeficiency syndrome (AIDS) are recognized as diseases of national seriousness. There is a chronic shortage of health care personnel. Generally speaking, the state does not provide adequately for the basic needs of the population.

Cultural life. In contrast with economic difficulties and political instability, Haiti's cultural life displays variety and originality and shows authentic achievements in many areas. Artists have drawn from the vitality of the people and the rich folk life to create internationally recognized works of art. Port-au-Prince is the centre of cultural and intellectual life, and the nation's most important museums, libraries, and entertainment facilities are located there.

Visual arts. Haitian paintings have received attention since the 1940s when a school of "naive" or "primitive" artists developed in Port-au-Prince and Cap-Haïtien. The works of the earlier and more famous artists are exhibited in major galleries in the United States and France. Primitive influence is also evident in such handicrafts as wood carvings and tapestries, which are manufactured in Haiti but sold throughout the Caribbean.

Music. African drum rhythms are combined with early 19th-century European dance music to create the merengue, a relatively slow, rhythmic music played by bands at public and private functions. The music of Haiti has generated original dance movements and expressions.

Literature. Haitian literature is written almost exclusively in French. Efforts are being made to establish Creole as a literary language, however, and novels, poems, and plays have been written in Creole. Haiti has had some writers of international status, including Jean Price-Mars, who evaluated the black African heritage in Haitian culture; Jacques Roumain, a brilliant poet, essayist, and novelist; Jacques-Stephen Alexis, who wrote novels and other works dealing with Haitian society; and René Depestre, noted for his poetic creations in an elegant French.

Creole influence in literature

Recreation. Haiti is a country of intense poverty, and its citizens do not generally have access to many types of organized recreational activities prevalent in other countries. Nevertheless the colourful Lenten carnival—although perhaps not as elaborate as some—is widely celebrated. The only team sport of note is soccer, which draws

sizable crowds to matches in Port-au-Prince. Cockfighting, however, may be Haiti's most popular activity, attracting, particularly, the gambling enthusiast.

Press and broadcasting. Technical and political difficulties have severely affected publishing in Haiti. Book publishing is limited, and French publishers have taken the lead in the publication of French literature. Censorship has been frequently enforced, and many writers have emigrated to Montreal, New York City, and Paris. Several daily newspapers circulate in Haiti, most of them published in the capital. There are two television stations, one of them government-owned, and a number of radio stations whose broadcasts are received throughout the island. (C.A.G.)

For statistical data on the land and people of Haiti, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

Christopher Columbus sighted the island that now includes Haiti and the Dominican Republic on Dec. 6, 1492, and named it La Isla Española. By the end of the 16th century, most of the island's original Arawak Indian population had disappeared from conquest or warfare or by being worked to death or killed by disease. Spanish settlement was thin and restricted mainly to the eastern end of the island; French pirates, based in Tortuga and other islands, had an almost unimpeded run of the western end. The pirates began to establish plantations there; in 1664 they founded Port-de-Paix in the northwest, and the French West India Company took possession.

Arrival of the French

The French colonial regime. In 1697, by the Treaty of Rijswijk, the western third of the island was formally ceded to France by Spain and was renamed Saint-Domingue.

Saint-Domingue's population increased greatly during the 18th century. It became the most prosperous New World colony, exporting sugar, coffee, cocoa, indigo, and cotton cultivated by African slave labour. By 1789 nearly two-thirds of France's foreign investments were based on Saint-Domingue, and in a good year its trade needed more than 700 oceangoing vessels.

Saint-Domingue had a population in 1789 of 556,000; of this, 500,000 were black slaves, 32,000 were whites, and 24,000 were free blacks. On Aug. 24, 1791, stimulated by the French Revolution, the slaves rose in rebellion. In order to maintain the island as a French possession, slavery was abolished in 1794. In 1795, by the Treaty of Basel, Spain ceded the rest of the island to France, but war in Europe precluded the actual transfer of possession.

In May 1801 Toussaint-Louverture, a former slave, became governor-general, but he would not declare the colony independent. Napoleon sent his brother-in-law, General Charles Leclerc, with an experienced force, including several mulatto officers in exile from Saint-Domingue, to restore the old regime. After several months of struggle against Leclerc's forces, in May 1802 Toussaint came to terms with the French expedition. The French, however, broke the arrangement and imprisoned him in France. He died on April 7, 1803.

Toussaint-Louverture

In the face of a rumour that Napoleon intended to restore slavery in Saint-Domingue as he had done in other French possessions, Jean-Jacques Dessalines and Henry Christophe led a black army against the French in 1802. The French commander and a large part of his army were defeated, and on Nov. 9, 1803, the remnant of the French expedition, under the Viscount de Rochambeau, surrendered. Under the armistice signed on November 18, the French withdrew, but they maintained a presence in the eastern part of the island until 1809.

Independent Haiti (1804–1957). On Jan. 1, 1804, the entire island was declared independent under its original Arawak name of Haiti. The war with France had utterly laid waste the country and destroyed the economy. In October 1804 Dessalines assumed the title of Emperor Jacques I; on Oct. 17, 1806, he was killed while trying to put down a mulatto revolt, and Henry Christophe took control of his kingdom. Civil war broke out between Christophe (later Henry I) in the north and Alexandre Sabès Pétion, based at Port-au-Prince in the south. In 1809, with British help, Spanish rule was restored in the eastern part of the island (Santo Domingo).

Christophe managed to improve the country's economy, but he had to force peasants to work on the plantations. He built a spectacular palace, Sans Souci, as well as an imposing fortress, the Citadelle Laferrière, in the hills to the south of Cap-Haïtien, where, with mutinous soldiers almost at his door, he committed suicide in 1820.

Jean-Pierre Boyer, who had succeeded to the presidency of the mulatto-led south on Pétion's death in 1818, became president of the entire country after Christophe's death. In 1822 he invaded and conquered Santo Domingo, which had declared itself independent from Spain the previous year and was now engaged in fighting the Spaniards. Boyer abolished slavery and confiscated church property in Santo Domingo; it was not until 1844 that the Haitians were expelled from Santo Domingo by a popular uprising.

Haitian independence was recognized by France in 1825, in return for an indemnity of nearly 100 million francs, to be paid at an annual rate until 1887. Britain recognized the state in 1833, and the United States in 1862, after the secession of the Southern slave states.

Boyer was overthrown in 1843. Between then and 1915 a succession of 20 rulers followed, 16 of whom were overthrown by revolution or were assassinated. Faustin-Élie Soulouque, who became president in 1847 and emperor for life in 1849, was extremely repressive. He turned on his mulatto sponsors, and his regime was in some ways a return to power of the black party. He tried unsuccessfully to annex the Dominican Republic (formerly Santo Domingo) and was overthrown in 1859 by one of his generals, Fabre Geffrard. Geffrard tried to reduce repression, encouraged educated mulattoes to join his government, and established Haitian respectability abroad. The 1890s saw an increase in U.S. attempts to gain military and commercial privileges in Haiti. In 1905 the United States took Haiti's customs into receivership, and, before World War I, U.S. business interests had gained a secure financial foothold and valuable concessions.

Fabre Geffrard

U.S. occupation (1915–34). From 1915 to 1934 Haiti was occupied by U.S. marines. The United States claimed legal justification on the grounds of humanitarian intervention and under the Monroe Doctrine. Many Haitians believed that the marines had really come to protect U.S. investments in the country and to establish a base to protect the approaches to the Panama Canal. Haiti signed a treaty with the United States—originally for 10 years, but it lasted until 1934—establishing U.S. financial and political domination. In 1918, in an election supervised by the marines, a new constitution was introduced that permitted foreigners to own land in Haiti.

One effect of the marine occupation was the nominal reestablishment of the mulatto elite in control of the government. Many Haitians resented the occupation, which they believed excluded them from public office and subjected them daily to racist indignities at the hands of the marines. The marines revived an old law of Christophe's time, which enabled them to employ forced labour on the roads. This resulted in a revolt of *cacos* (guerrillas), which was suppressed. The program of public works undertaken by the marines in health clinics, sewage, and roads hardly satisfied the Haitians, who felt the effort inadequate.

In October 1930 a national assembly, the first since 1918, was elected. Controlled by nationalists, it in turn elected as president Sténio Joseph Vincent. In August 1934 the U.S. president, Franklin D. Roosevelt, withdrew the marines; but direct U.S. fiscal control continued until 1941, and indirect control continued until 1947.

Presidential regimes (1934–57). In 1935 a plebiscite extended Vincent's term to 1941 and amended the constitution so that future presidents would be elected by popular vote.

In October 1937, troops and police from the Dominican Republic, with popular support, massacred thousands of Haitian labourers living near the border. The following year the Dominican government agreed to pay $3.4 million in compensation to relatives of those slain, but only part was actually paid. The enmity between the two countries had long historical roots. The Dominicans,

with their Spanish culture and greater admixture of European blood, looked disdainfully upon the black Haitian labourer. Haitian labour was, however, necessary to the Dominican economy.

In 1946, Haitian workers and students held strikes and violent demonstrations in opposition to the president, Élie Lescot, who had succeeded Vincent in 1941 and had altered the constitution to enable himself to serve a further term. Three military officers seized power, and under their supervision Dumarsais Estimé was elected president. In 1950, after Estimé sought to extend his term, the same three military officers took control. That October one of them, Colonel Paul E. Magloire, was elected president in a plebiscite. He also sought to remain in power, but in December 1956 the army forced him to resign the office.

Military coups

Haiti since 1957. In September 1957, after considerable unrest and several provisional presidents, François Duvalier (called "Papa Doc")—formerly employed on a U.S. medical aid project and a student of voodoo—was elected president. He promised to end domination by the mulatto elite and to bring political and economic power to the black masses. Violence continued, and, after an unsuccessful attempt to overthrow Duvalier in July 1958, he organized a gang of violent adherents—the Tontons Macoutes—who terrorized the population. In 1964 Duvalier, by then firmly in control, had himself elected president for life. Haiti under Duvalier was, in effect, a police state.

Near the end of his life, faced by a contracting economy, withdrawal of most U.S. aid, and a decline of tourism, Duvalier relaxed some of the severe repression and terror that had characterized his early regime. Before his death in 1971, he designated his son, Jean-Claude, aged 19 and called "Baby Doc," to succeed him as president for life.

The regime of Jean-Claude Duvalier sought international respectability. Repression, while present, diminished; tourism, U.S. aid, and the economy revived somewhat, and the younger Duvalier, after an uncertain start, proved to be a surprisingly adroit politician. Opponents, however, saw little change in the basic nature of the regime.

In 1980 Duvalier married Michèle Bennett and transferred to her many of the privileges held by his mother, Simone Duvalier. The regime subsequently followed an uneven policy of lifting political controls and then restoring all restrictions as soon as dissent was expressed.

A series of popular demonstrations throughout Haiti in late 1985 and early 1986 against high unemployment, poor living conditions, and the lack of political freedom could not be repressed by the brutality of the Tontons Macoutes (estimated at 15,000 men). On Feb. 7, 1986, Duvalier fled Haiti, with U.S. assistance, for France. A five-member civilian-military council, the National Council of Government, led by Lieutenant General Henri Namphy, took charge of Haiti and promised democratic reforms and elections by late 1987. Violence by pro-Duvalier elements, with government connivance, led to the cancellation of these elections. New elections on Jan. 17, 1988, were widely considered fraudulent, and in light voting the declared winner was Leslie Manigat, who was overthrown by Namphy on June 20. (C.L.R.J./M.J.MacL.)

After a massacre during a church service on Sept. 11, 1988, disgusted junior officers sent Namphy into exile and installed Lieutenant General Prosper Avril as president. He promised economic reforms and a return to democracy.

Two coup attempts shook the government in 1989, and in early 1990 Avril began a crackdown on the opposition. In March, however, following a week of nationwide antigovernment protests, Avril was forced to resign, and Ertha Pascal-Trouillot, the only woman on the country's supreme court, became the provisional president. Legislative and presidential elections were to be held later in 1990.

Migration from Haiti to The Bahamas and Florida, which was particularly strong in the 1970s, continued into the early 1990s because of the miserable economic conditions and the unstable political climate. (C.L.R.J./M.J.MacL./Ed.)

For later developments in the history of Haiti, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

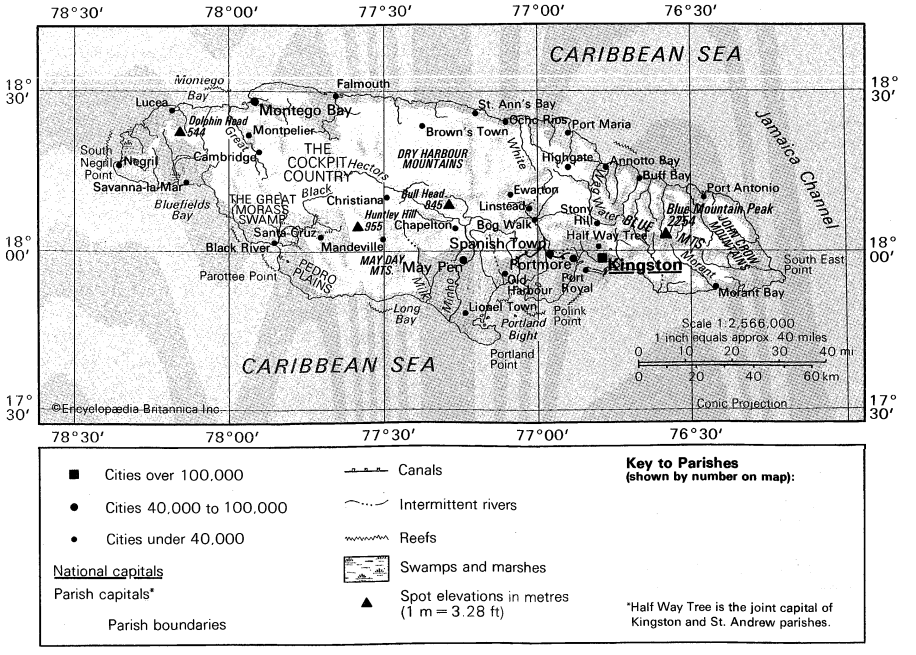
Jamaica

Jamaica, a parliamentary state within the Commonwealth, is the third largest island in the Caribbean Sea after Cuba and Hispaniola. With an area of 4,244 square miles (10,991 square kilometres), Jamaica is about 146 miles (235 kilometres) long and varies between 22 and 51 miles wide. It is situated some 100 miles due west of Haiti, 90 miles south of Cuba, and 390 miles northeast of Cape Gracias a Dios, Nicaragua, the nearest point on the American continent. The national capital is Kingston.

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

Christopher Columbus

Agriculture, the largest single employer of labour, is a main contributor to the national income, together with industry—notably bauxite mining—and tourism. Jamaica has made great strides in economic development since its independence in 1962. The national motto, "Out of many, one people," describes a multiracial society whose integration can be described as profound and enviable.



PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* In general, the topography consists of coastal plains encircling an island that is bisected from east to west along its length by mountains and plateaus. The mountains form the chief physical feature. Almost half of Jamaica's surface is more than 1,000 feet above sea level. The chief range rises in the east to Blue Mountain Peak at 7,402 feet (2,256 metres) and then loses altitude as it sweeps westward. The mountains contribute to the great diversity of scenery for which the island is famous, ranging from the stunted, elfin forests of the highest peaks to the dry, sandy, cactus-growing areas of the south. There are rain forests in the highlands and flat, alluvial plains located chiefly on the south side of the island. The rolling limestone hills and plateaus in the central and western areas include the unusual, trackless karst (a limestone region broken by ridges, depressions, and caverns) region of the Cockpit Country, covering 500 square miles.

Scenic diversity

Drainage and soils. There are numerous rivers and streams issuing from the central mountainous area, but many disappear intermittently into the sinkholes of the karst region. Few are navigable for any great distance because of their rapid descent from the mountains. Some of the larger rivers have alluvial plains in their lower valleys, and some have deltas. The 44-mile-long Black River in the west is the longest and most important river; it is navigable by small boats for about 25 miles from its mouth.

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

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

Sea breezes

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

Rains are seasonal, falling chiefly in October and May, although thunderstorms in the summer months, from June to September, can bring heavy showers. The average annual rainfall for the entire island is 82 inches (2,095 millimetres), but regional variations are considerable. The mountains force the trade winds to deposit more than 130 inches a year on the eastern parish of Portland, while little precipitation occurs on the hot, dry savanna lands of the southern and southwestern plains. Jamaica is susceptible to hurricanes during the summer but after 1951 it was not struck until Hurricane Allen in 1980 and Hurricane Gilbert in 1988. Earthquakes have caused serious damage only twice—in 1692 and 1907.

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

Reduction of forests

Jamaica has few indigenous quadrupeds. The coney, a member of the rodent family, once very numerous and prized as food before the European immigration, is now much reduced in range and number because of subsistence hunting and habitat destruction. The widespread mongoose, a small carnivore that feeds on rats and snakes, was introduced in 1872. The bat species are the most numerous of the mammals. The native crocodile may be in some danger of extinction, but this has not been confirmed. The main freshwater fish is the mountain mullet; there are four species of crayfish.

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

Settlement patterns. With the emancipation of the island's black slaves in 1838, a large proportion of the freed population—often with the aid of Nonconformist (non-Anglican) missionaries—left the large plantations. They moved to the hills where land was plentiful and created their own villages and communities. This migration laid the foundation of the present pattern of rural settlement.

Most of the cities and chief towns are located on the coastal plains, where the main commercial crops are grown. Kingston, the national capital, is located on the southeastern coast. It stands on the Liguanea Plain with the sea to the south and the St. Andrew Mountains, which form part of the ranges of the parish of St. Andrew, to the north. It is the commercial, administrative, and cultural centre of the island and the focus of its transportation services. Important centres in the interior are Spanish Town, the old capital 13 miles west of Kingston, and Mandeville, high in the Manchester mountains. Other important towns include Montego Bay, Ocho Rios, and Port Antonio on the north coast. Their fine white-sand beaches and exquisite mountain scenery make them popular tourist resorts.

The people. *Ethnic and linguistic composition.* The aboriginal Arawak Indians were exterminated by the Spanish colonists by the time the English invaded the island in 1655. The Spaniards themselves disappeared as a population element shortly afterward. With the large-scale introduction of African slaves to work the sugar estates, the English settlers were soon greatly outnumbered. Today the population consists predominantly of the black and mulatto descendants of those slaves. Small minority elements originate from the United Kingdom, India, China, Syria, Portugal, and Germany.

Origin of the population

English is the official language, but a local creole is also widely spoken and has achieved considerable status. It is basically English in vocabulary and grammar but contains features derived from a variety of African languages, as well as from Spanish and, to a lesser degree, French. This creole is a language of value and charm and is an important communication element on the island. There is a wide spectrum of dialects, however, some quite removed from standard English.

Religion. Freedom of worship is guaranteed by Jamaica's constitution. No single religion has a majority of adherents, but more than half are in Protestant denominations. Less than 10 percent of the population attend the Anglican church, which, as the Church of England, was the established church of the country until 1870. The largest denomination is that of the Church of God. Other sizable groups include the Baptists, Seventh-day Adventists, Pentecostals, and Roman Catholics. Since the late 1970s there has been a steady increase of Evangelicalism. Smaller denominations, such as the Moravians, Disciples of Christ, Society of Friends, and United Church are important because of their social welfare work.

Nearly every Christian denomination and sect is represented, and the Jewish community is one of the oldest in the Western Hemisphere. There is a Hindu community, a Muslim mosque, and a branch of the Ethiopian Orthodox church. Some of the revivalist sects base their beliefs on Christianity, but their forms of worship differ widely from those accepted by most orthodox denominations. The central feature of the Pocomania sect, for example, is spirit possession; the Cumina sect has rituals characterized by drumming, dancing, and spirit possession.

Demographic trends. Through the centuries the popu-

lation rose steadily, despite considerable emigration. Peak growth rates came in the 1950s and '60s, resulting in an unusually large number of persons reaching their 20s in the 1980s. The youthful population is expected to reach about three million by the year 2000. Birth and death rates have both declined since the 1970s.

Migratory patterns

The pattern of migration began to take shape with the first wave of emigrants from Jamaica to Panama in the 1850s to help build a railway there. The French attempt to build a sea-level canal across the Isthmus of Panama in 1879 again drew Jamaican workers to Panama. The project failed, but the successful U.S. venture in the early 1900s to build the lake-and-lock canal accounted for another large emigration of Jamaican workers. The development of the banana industry in Central America drew still further numbers of Jamaicans, as did the need for workers in the sugar and coffee plantations of Cuba. Greater than all of the other emigrations combined was that to the United States, which began at the close of the 19th century, diminishing only after enforcement of quota regulations. The tide of emigration then turned toward Canada and, later, the United Kingdom and resulted in some 200,000 Jamaicans migrating to Britain during the period 1950–60; again immigration laws reduced the flow. Since the mid-1960s the United States and Canada have become the primary destination of Jamaican migrants.

The highly mobile character of Jamaica's population is further apparent in the high level of internal migration. The development of the bauxite and tourism industries was a significant factor in this regard. Between 1969 and 1974, for instance, some 28 percent of the population changed their parish of residence.

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

Resources. Among the minerals found on the island, bauxite, gypsum, silica sand, ceramic clays, marble, and limestone are of commercial interest. The bauxite is found in an area of about 1,000 square miles in central Jamaica; the gypsum and marble are in eastern Jamaica; clays are in the west; and limestone is found throughout the island.

Agriculture, forestry, and fisheries. Agriculture continues to be one of the main bases of the island's economy. The two major crops are sugar—with its by-products of rum and molasses—and bananas. Other important crops are citrus fruit, coffee, pimento, cocoa, tobacco, and ginger.

Forestry production is insufficient to meet the country's needs; most of the wood, cork, and paper consumed is imported. The government encourages afforestation.

Fishing is a major enterprise, which is estimated to support some 150,000 persons. The island shelf is the traditional fishing area. Mechanized boats sail about 60 miles southwest of Jamaica to Pedro Bank, while some fishermen travel 300 miles to fishing grounds.

Industry. Since 1952 mining has played an increasingly significant role in the country's economy. Bauxite mining has become the island's main mineral-industry enterprise. There is decreasing involvement of foreign companies, who have traditionally run the mines, while government involvement has increased. The production of silica sand is directed at local glass-container manufactures, while most of the gypsum is mined for export. Cement is largely used in local construction. Other mineral resources include limestone, marble, clay, peat, sand, and gravel, and to a lesser extent, lignite, black sands containing titanium, copper, lead, zinc, and phosphates.

Manufacturing is increasingly important, both in providing employment and in satisfying the growing demand for manufactured goods. Processed foods, textiles, and metal products are the most important manufactures. Other significant categories are sugar, rum, and molasses processing, printing, chemical production, and cement and clay products. Industrial growth has been stimulated by the activities of Jamaica Promotions Ltd. (JAMPRO), a statutory body that promotes and facilitates both foreign and local investment, as well as Jamaican exports abroad.

Increasing reliance is being placed on tourism, which has become the country's largest source of foreign exchange. The traditional attraction of Jamaica to the tourist is the pleasant climate, fine beaches, and superb scenery.

Electricity is supplied from both public and private sources, about two-thirds coming from the public sector. Privately owned generating plants supply the power needs of major industries.

Finance. Commercial banks, some of which are subsidiaries of Canadian, British, and U.S. banks, dominate the financial sector. Savings and credit services are also offered by life insurance companies, building societies, and credit unions. The central bank, called the Bank of Jamaica, founded in 1960, controls money and credit and promotes economic development. A number of development banks and special development funding institutions provide loans for industry, housing, tourism, and agriculture.

Trade. The principal exports are alumina and bauxite; sugar, bananas, coffee, and other agricultural products; and clothing. Jamaica's principal trading partners include the United States, the United Kingdom and other members of the European Economic Community (EEC), Canada, member states of the Caribbean Community, certain Latin-American countries, Japan, the Soviet Union, and some East European countries. Jamaica enjoys preferential trading relationships with the member states of the EEC under the Lomé Convention, with the United States under the Caribbean Basin Initiative, with Canada under the Caribcan arrangement, and with other Caribbean states as a member of the Caribbean Community (Caricom). The Jamaican currency is maintained at a relatively stable exchange rate relative to the U.S. dollar.

Transportation. Generally, the transport systems follow the coastline or cut across the central mountains from north to south. The main roads encircle the island, loop into the plains areas, and cross the mountains at three major north-to-south crossings. Public passenger services are available outside the capital, which itself has a regular bus service. There are also taxi and limousine services.

The road system

The main line of the railway system, parts of which have existed since 1845, runs northwest from Kingston to Montego Bay via Spanish Town, May Pen, and Montpelier. The system is operated by the Jamaica Railway Corporation. There are scheduled international air services at the two major airports—the Norman Manley, on the Palisadoes in Kingston, and the Donald Sangster at Montego Bay; both are named for former prime ministers. These airports, together with another at Tinson Pen, Kingston, also handle scheduled domestic flights and air-taxi services. Port Antonio, Ocho Rios, and Negril have major public airstrips, and there are privately owned airstrips throughout the island. Kingston, Montego Bay, Ocho Rios, and Port Antonio are the principal seaports. Regular shipping services link Jamaica with the Caribbean, South and Central America, Canada, the United States, Europe, and East Asia.

Administration and social conditions. *Government.* Under the Jamaica (Constitution) Order in Council of 1962, by which the island achieved independence, the monarch of the United Kingdom is titular head of state. A Jamaican governor-general is chosen by the monarch on the advice of the prime minister. The prime minister is appointed by the leading political party from its parliamentary members.

Legislature

The legislature is a bicameral parliament consisting of a House of Representatives and a Senate. The House has 60 members, who are elected by universal adult suffrage. The speaker and deputy speaker are elected by the House from its members. The Senate has 21 members, who are appointed by the governor-general—13 in accordance with the advice of the prime minister and eight on the advice of the leader of the opposition party. The president and deputy president of the Senate are elected by its members. The principal policy-making body is the Cabinet, which consists of the prime minister and at least 11 other ministers, of whom at least two but not more than four must be members of the Senate. The Privy Council is limited to advising the governor-general on the exercise of the royal prerogative of mercy and on the discipline of government officials and employees.

The island is divided into 14 parishes, two of which are amalgamated as the Kingston and St. Andrew Corporation. Local affairs in the other parishes are administered by individual parish councils whose members are elected. The chairmen of the councils are the mayors of those parish capitals that enjoy mayoral status.

Political parties

The two main political parties are the Jamaica Labour Party (JLP) and the People's National Party (PNP). General elections must be held at least once every five years. The most important trade unions are the Bustamante Industrial Trade Union (affiliated with the JLP) and the National Workers' Union (affiliated with the PNP). There are also employers' associations.

Judiciary. The legal system is based on English common law. The highest court in the Jamaican legal system is the Court of Appeals. It hears appeals from the Resident Magistrates' Court, which includes the Family Courts, the Kingston Traffic Court, Juvenile Courts, and a division of the Gun Court, and from the Supreme Court, the highest trial court.

Education. Primary education is free and in certain districts compulsory. A substantial part of the annual budget goes to the Ministry of Education. Considerable sums are devoted to the College of Agriculture, the College of Arts, Science, and Technology, the University of the West Indies (the main campus of which is at Mona, a northeastern section of Kingston), and teacher-training colleges. Education is provided by government-aided and private schools, some of which are run by religious bodies. In recent years there has been an emphasis on vocational training for school dropouts.

Health and welfare. Medical care is provided by several public hospitals, including the university hospital, and various health centres and clinics. There are also a few private hospitals. Highly successful programs of insect control and malaria eradication have been undertaken, and Jamaica participates in the hemispheric drive to increase levels of participation in immunization programs.

In the field of social development, government programs are geared toward helping persons to function as productive individuals or groups in their communities. These programs, prepared by government and private organizations, are designed to assist children, youth, women, and the community in such areas as vocational training and job placement. The government operates a compulsory National Insurance Scheme that provides retirement and other benefits and a noncontributory Social Assistance Programme to help the needy.

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

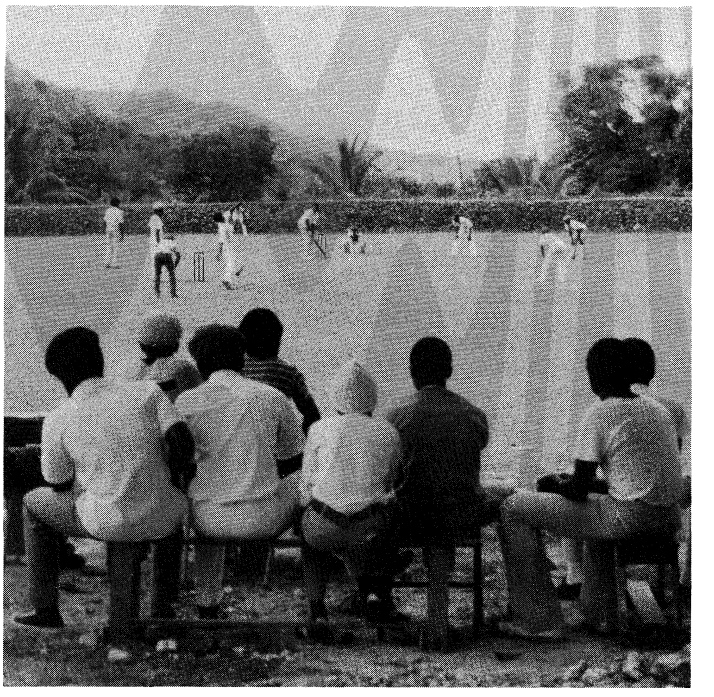
Cultural life. There is a vigorous and productive art movement in Jamaica. The works of Jamaican writers may be read in several languages. Jamaican artists have exhibited successfully abroad, and local art shows are a regular part of life. The Institute of Jamaica, an early patron and promoter of the arts, sponsors exhibitions and awards. It has responsibility for the Cultural Training Centre, which includes schools of art, dance, drama, and music, as well as for the National Library, the National Gallery, and a publishing company. The institute is also the country's museums authority. The Jamaica Library Service, the Jamaica Archives, the National Library, and the University of the West Indies contribute to the promotion of the arts and culture, as do numerous commercial art galleries.

Art promotion

Theatre and musical groups are highly active on the island. The National Dance Company, formed in 1962, has earned international recognition. Much of the country's artistic expression finds an outlet in Festival, sponsored annually by the government as part of the independence celebrations. While the festival has many features of the traditional Caribbean type of carnival, it is much wider in scope. In addition to street dancing and parades, there are also arts and crafts exhibitions and literary, theatrical, and musical competitions.

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

Folk traditions



A cricket match at a ground near Ocho Rios, Jam. The sport is highly popular in the British West Indies.
© Berlitz from TSW—CLICK/Chicago

Jamaica's beaches are perhaps the country's most highly utilized recreation facility. Cricket, however, is the most popular sport, played throughout the island and drawing large crowds to international matches held at Sabina Park. Soccer ranks next to cricket in importance.

Freedom of the press is guaranteed by the Jamaican constitution. The island has three daily newspapers, *The Daily Gleaner,* the *Jamaica Record,* and *The Star,* all published in Kingston. Numerous U.S. and other foreign newspapers and magazines are also readily available. The publicly owned Jamaica Broadcasting Corporation is the chief radio and television broadcaster. Other radio programming is provided by KLAS and Radio Jamaica Ltd.

(C.V.B.)

For statistical data on the land and people of Jamaica, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

The peopling of Jamaica seems to have begun from the islands to the east, at least as early as 5,000 BC. In about AD 600 a fresh wave of settlers, the Arawak, arrived backed by the prevailing southeasterly winds. They eventually settled throughout the island, and their economy, based on fishing and the production of corn and cassava, sustained a population of about 60,000, grouped in villages under the guidance of chieftains called caciques.

Columbus reached the island in 1494 and spent a year shipwrecked there in 1503–04. Spanish settlement began in 1509, when Juan de Esquival founded the town of Sevilla la Nueva on the north coast; the island had by then been granted to the Columbus family by the Spanish crown, but in the years that followed it became something of a backwater, valued chiefly as a base for supplies of food and hides. In 1534 the capital was moved to Villa de la Vega (later Santiago de la Vega), known today as Spanish Town, and a substantial town developed there. The Spanish impact on the Arawak was disastrous, partic-

ularly because the latter were highly vulnerable to European diseases. By the beginning of the 17th century there were virtually no Arawak left, and the population of the island numbered about 3,000, including a small number of black slaves.

In 1655 a British expedition under Admiral William Penn and his colleague General Robert Venables captured Jamaica, and the island was slowly settled by its new masters. By 1660 the last Spaniard had been expelled, but the Spaniards' former slaves held on in the mountains, eventually forming bands called Maroons, a name probably derived from the Spanish word *cimarrón,* meaning "wild" or "untamed." The Maroons were joined by runaway slaves during the British occupation and long remained a thorn in the side of the British.

The buccaneers. Jamaica was governed by military authority until 1661, when Edward D'Oyley was appointed captain general and governor in chief with an executive council. The island soon became a chief resort of buccaneers, who frequently united the profession of merchant or planter with that of pirate or privateer. The buccaneers—who preyed on Spanish ships—operated mainly from their base at Port Royal. By their relentless attacks on Spanish Caribbean cities, they kept the Spaniards occupied at a time when Britain was unable to spare a fleet for the protection of its West Indian colonies.

British rule. By the Treaty of Madrid (1670), British title to the island was recognized, and the buccaneers were suppressed. The Royal African Company was formed in 1672 with a monopoly of the English slave trade, and from that time Jamaica became one of the greatest slave marts in the world, with a thriving smuggling trade to Spanish America. By the 18th century, Jamaica had become one of the most valuable of colonial possessions. In 1672 there were in Jamaica about 70 sugar works, 60 indigo works, and 60 cocoa works. An attempt was made in 1678 to impose royal taxes and to supersede the powers of the legislature. The privileges were restored in 1682; but not until 46 years later was the question of the revenue settled, with a compromise by which Jamaica undertook to pay £8,000 (later reduced to £6,000) per annum to the crown, provided that the laws of England were made binding in Jamaica; all additional revenues had to be voted annually by the assembly. In 1692 an earthquake destroyed most of the town of Port Royal and led to the foundation of Kingston.

The Royal African Company

Economics and government under colonial rule. A threatened invasion by the French and Spanish in 1782 was averted by the victory of the admirals George Rodney and Samuel Hood off Dominica. The last attempt at invasion was made in 1806, when the French were defeated by Admiral Sir John Duckworth. During the French war, Jamaica was at the zenith of its prosperity, with coffee rivaling sugar as an export crop and with more than 300,000 slaves at work. The abolition of the slave trade in 1807, however, raised the planters' costs, and the end of the war brought a steady drop in sugar prices.

Emancipation struck a further blow at the planters' prosperity and security. All slaves were emancipated by an act of the imperial parliament in 1833. They became free in fact, after a period of so-called apprenticeship, in 1838. Many left the plantations and moved to the hills, where their descendants live as small farmers today. The planters received compensation at the rate of £19 per slave but generally were left financially exhausted and with a scarcity of labour. The abolition in 1846 of the tariff protection of colonial produce in the British market reduced the price of sugar still further and in many cases destroyed the profits of the impoverished planter.

Dissensions among the executive, the legislature representing propertied interests, and the home government, as to the means of retrenching public expenditure, created much bitterness. Although some improvement marked the administration of Sir Charles Metcalfe and the Earl of Elgin, when Indian immigration was introduced to redress the scarcity and irregularity of labour and the railway was opened, the improvement was not permanent. Along with the collapse of the plantation system, there was widespread poverty and unemployment. All this produced a crisis in 1865 that changed the old social and economic patterns. An outbreak occurred at Morant Bay in October 1865, in which the chief magistrate of the parish and 18 other white persons were killed. The rising was suppressed under martial law, and the principal instigator, G.W. Gordon, was hanged. These severities were widely applauded in the West Indies, but indignation in England led to the recall of the governor, Edward John Eyre, and to a drastic change in the government of the island.

Violence and unrest

Before his recall, Eyre had induced the Jamaican assembly, frightened by the riots, to vote its own extinction. In its place a crown colony form of government, in which the governor wielded the only real executive or legislative power, was established by an act of the British Parliament in 1866. The new governor, Sir John Peter Grant, achieved a remarkable reorganization of the affairs of the colony. He established a constabulary on the lines of that of Ireland; reconstructed the judicial establishment, substituting stipendiary magistrates for the planters-justices; established a public medical service, a public works department, and government savings bank; improved education; and irrigated the fertile but drought-stricken plain between Spanish Town and Kingston. During his government the lucrative trade in bananas was started by Captain A.W. Baker, founder of the organization that later became the United Fruit Company. Bananas soon became a principal crop, grown for export by small farmers as well as by large estates. Many Jamaicans also found employment in the early years of the 20th century on the construction of the Panama Canal and on sugar plantations in Cuba.

On Jan. 14, 1907, a violent earthquake struck Kingston. Almost every building in the capital and in Port Royal was destroyed or seriously damaged. About 800 persons were killed and a large part of the city was burned. These disasters offered an opportunity of improving Kingston, and under Sir Sydney (Lord) Olivier the public offices were rebuilt on the finest street of the city.

Representative government was restored by stages from 1884, when nine elected members were added to the legislature. Their number was increased to 14 in 1895. Further changes were discussed from 1922 to 1926, and more vigorously after 1938, partly as a result of riots in that year. By 1938 dissatisfaction with the crown colony system, sharpened by the hardships and suffering brought on by a worldwide economic depression, erupted in serious and widespread rioting. These events resulted in the formation of the first lasting labour unions as well as of political parties linked to them. A growing demand for self-determination also became apparent. In 1944 a new constitution established a house of representatives elected by universal suffrage, in which a two-party pattern soon emerged. A modification of this constitution in 1953 gave departmental responsibilities to elected ministers but retained a nominated legislative council as an upper house with limited powers and an executive council including both officials and ministers. In 1957 the official members left the executive council, which then became a Cabinet under the chairmanship of the premier. Full internal self-government was obtained in 1959.

Jamaica was relatively little affected by World Wars I and II, though many Jamaicans served in the British armed forces overseas. After World War II the island profited greatly from help under the Colonial Development and Welfare Act and from outside sources of private capital. Colonial Development grants financed the building of the University of the West Indies, an important factor in the preparation for independence, which was established in Jamaica in 1947. Many new industrial undertakings were started, including a sugar refinery, citrus factories, and a cement factory. Development was temporarily checked in August 1951 by a severe hurricane, which devastated crops and killed about 150 people. The expansion of the tourist trade and the development of bauxite mining helped increase employment opportunities on the island.

Independence. On Jan. 3, 1958, Jamaica became a founding member of the West Indies Federation, a group of Caribbean islands that formed a unit within the Commonwealth of Nations. Norman Manley, leader of the People's National Party (PNP), became prime minister af-

West Indies Federation

ter the elections held in July 1959, but in 1960 the Labour Party under Sir Alexander Bustamante pressed for secession from the federation. Widespread support was given to their views in a 1961 general referendum. At the general election in April 1962 the Labour Party was returned to power and Bustamante became prime minister. In May the federation was dissolved, and on August 6 Jamaica became independent with full dominion status within the Commonwealth. In 1963 the nation joined the International Monetary Fund, and on a state visit Elizabeth II, as queen of Jamaica, opened Parliament there on March 4, 1966. In June 1969 the country became the 24th member of the Organization of American States.

During most of 1965 and 1966 Bustamante was ill and Donald Sangster acted as prime minister. After a campaign marked by considerable violence, the first general election since Jamaica had gained independence was held on Feb. 21, 1967. The Jamaica Labour Party (JLP) gained 33 seats and the PNP 20 seats. Bustamante had announced his retirement from politics before the election, and the new prime minister was Sangster. He died shortly after taking office, however. Hugh Lawson Shearer, leader of the Bustamante Industrial Trade Union, was then elected prime minister. In the election of early 1972 the PNP obtained its first major victory, and Michael Manley, the leader of the party, was sworn in as prime minister.

Manley, the charismatic son of Norman Manley, based his winning campaign on the "politics of participation," and once in office he embarked on a number of social reforms. Censorship was eliminated, and many restrictions on civil liberties were lifted. His government pursued a program, largely successful, to wipe out illiteracy. But, although Manley's regime had such promising beginnings, the government was soon overwhelmed by the exigencies of the impoverished Jamaican populace. Economic problems undermined most of Manley's social programs.

During the crucial elections of 1976, there was virtual political warfare between the PNP and the opposition JLP. Manley's PNP won heavily. Perhaps despairing of economic betterment through close ties with the United States, Manley had proposed strengthening ties with Cuba. In 1977 the government assumed majority ownership of the bauxite mines, which had been foreign-owned.

The Edward Seaga administration

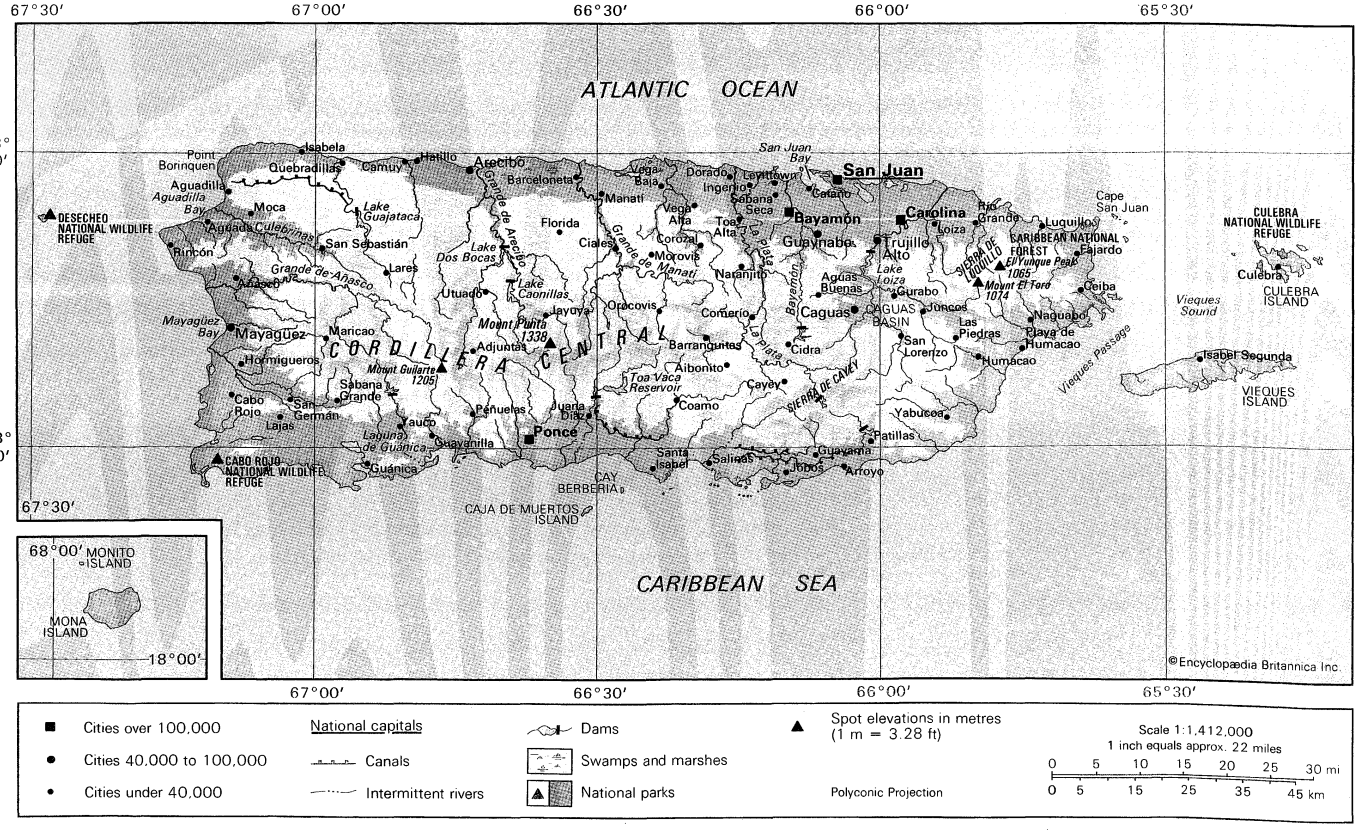
The continuing economic misery of much of the population and increasing political violence led to Manley's defeat in the 1980 election. The new prime minister, Edward Seaga of the JLP, had to contend with the widespread destruction of Hurricane Allen as one of his first acts in office. Although Seaga had disapproved of the PNP's close ties with Cuba, he at first maintained a friendly, albeit aloof, relationship with Castro. But by December 1981 Seaga officially severed diplomatic ties with Cuba. Concurrently, relations with the United States improved, and Jamaica became a major recipient of U.S. aid in the West Indies. The economy performed well at first but quickly worsened and continued in a downward trend despite an economic boost from the lowering of oil prices. In 1986 the PNP won local elections, a development proclaimed by administration foes as a sign of disapproval of Seaga's policies. Any progress toward economic recovery was wiped out in September 1988, when Hurricane Gilbert struck the island, causing even greater damage than had the earlier Allen. In elections held in February 1989 the PNP won a decisive victory, unseating Seaga and restoring Manley as prime minister. (J.H.Py./D.J.Bu.)

For later developments in the history of Jamaica, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Puerto Rico

The Commonwealth of Puerto Rico, a Caribbean possession of the United States, lies at the eastern end of the Greater Antilles, its main island about 50 miles (80 kilometres) from the island of Hispaniola to the west and about 40 miles from the Virgin Islands to the east. The U.S. mainland lies about 1,000 miles to the northwest. Puerto Rico covers an area of 3,515 square miles (9,104 square kilometres), including offshore islands. It has a mountainous tropical ecosystem with little flat land, few mineral resources, and a rapidly growing population.

Christopher Columbus discovered the island in 1493. Under Spanish authority from that time until 1898, Puerto Rico's lack of resources resulted in neglect and minimal investment by the Spanish. The capital city of San Juan, however, has one of the best harbours in the Caribbean,



and the Spanish built fortifications to protect this asset for their vital oceanic trade routes. When the United States acquired Puerto Rico in 1898 as a result of the Spanish-American War, it found itself in control of a poor island whose inhabitants were mostly involved in small-scale agriculture. The social system at that time was Spanish and conservative; the people were mostly rural, poor, uneducated, Roman Catholic, and resistant to change. The sudden intrusion of capitalistic ideas and values assured a high degree of social and cultural conflict.

Modern Puerto Rico is generally well-off by Latin-American standards. Beginning in the 1940s, a political coalition between the Puerto Rican leader Luis Muñoz Marín and the U.S.-appointed governor, Rexford Guy Tugwell, was forged to promote a self-help program, called "Operation Bootstrap," of economic development and social welfare. In a little more than four decades, much of the territory's crushing poverty was eliminated. This was done partly through emphasis on the development of manufacturing and service industries, the latter related to an enormous growth in tourism. Improvements have been made largely with the cooperation of the United States, but relationships with that country have also become a focal point of political turmoil. Various factions have bitterly disputed the political status of the island, and, although a majority voted to retain its commonwealth relationships, strong minorities have continued to push for statehood or—at times with violence—independence.

"Operation Bootstrap"

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Puerto Rico's mountainous backbone is the easternmost extension of a tightly folded and faulted ridge that extends from the Central American mainland across the northern Caribbean to the Lesser Antilles. While the highest point on the island reaches only about 4,389 feet (1,338 metres) at Mount Punta, there is a marine trough north of San Juan that plunges to more than 30,000 feet (9,144 metres) below sea level, one of the lowest ocean depths. The great difference in crustal elevations illustrates the strong tectonic forces that have operated in geologic history to create these features. Puerto Rico still occasionally suffers from earthquakes, reflecting the ongoing geologic processes. Rectangularly shaped, the island measures, at most, only about 111 miles from east to west and a mere 40 miles from north to south. Two important islands off the east coast, Vieques and Culebra, are also parts of Puerto Rico, as is the island of Mona to the west.

Terrain

Most of Puerto Rico is either mountainous or hilly terrain, with nearly a fourth of the island composed of steep slopes that inhibit agriculture and increase building costs. The highest mountain range, the Central Cordillera, with altitudes exceeding 3,000 feet, trends east–west and is located off-centre, closer to the south coast. Slopes rise abruptly from the south coast to the highest peaks and descend more gently toward the north. The Caguas Basin is the largest of several basins in the mountains that provide level land for settlements and agriculture. There is a continuous but narrow lowland along the north coast. Most people live along the coastal lowlands, and migration from the mountainous rural areas to the coastal cities continues to empty out the island's interior.

Drainage and soils. Because most of Puerto Rico's precipitation falls on the north-facing mountain slopes, the majority of the permanent rivers flow from the interior to the north coast. The river courses on the south coast are dry most of the year, carrying water only after rainfall. No river is large enough for navigation, but several north-flowing rivers have been dammed for hydroelectric power and to provide drinking and irrigation waters. Irrigation along the south coast is essential for agriculture. Some pockets of alluvial soils on the south coast have better than average fertility, but all agricultural areas need fertilizer for commercial production and most need irrigation or conservation techniques to preclude destructive erosion. Many worn-out areas in the mountains have been set aside for forest preserves.

Climate. Puerto Rico's tropical climate varies according to elevation and exposure to rain-bearing winds. Northeast trade winds bring heavy rainfall to the north coast, while the south coast is in the rain shadow of the mountains. San Juan receives 60 inches (1,524 millimetres) of rain; El Yunque Peak farther east receives 180 inches (4,572 millimetres); while Ponce on the south coast receives only 36 inches (914 millimetres). Rain falls each month of the year, but the heaviest rainfall occurs between May and December. The average daily humidity is 66 percent, which makes the 78° F (26° C) average lowland temperatures feel warmer. Highland temperatures average about 5° F less, and the annual range between the hottest and coolest months in highlands or lowlands is less than 6° F (3° C). Hurricanes occur between June and November, sometimes exacting high tolls in property damage and loss of life. The delightful climate and beautiful beaches are an important resource for tourism.

Plant and animal life. Plant life is abundant and varied. Tropical rain forest is one of the main types of natural vegetation on the north side of the island, while light precipitation on the south side restricts growth to mostly thorn and scrub vegetation. Most of the island's original vegetation has been removed through centuries of agricultural exploitation, and planted varieties of trees, shrubs, and grasses now predominate. The scarlet- and orange-flowered royal poinciana, or flamboyant (*Delonix regia*), and the African tulip (*Spathodea campanulata*) are among the trees that dot the mountains with patches of vivid colour against the lush green background. Within the Caribbean National Forest southeast of San Juan are preserved rare species of orchids and an unusual species of small green parrot. Birds are abundant in numbers and types, but land animals are mostly confined to nonpoisonous snakes, lizards, mongooses, and the *coqui,* a frog that has become a symbol of Puerto Rico. Numerous varieties of fish abound in the surrounding waters, but edible and nonedible species mingle together, making commercial fishing more difficult.

Vegetation

James P. Rowan

The dense Caribbean National Forest, southeast of San Juan, P.R.

Settlement patterns. Since the 19th century the dispersed farmsteads in the hills of Puerto Rico have given way to an urban society with scattered villages in the mountains. Puerto Rico is now about 70 percent urban. Because there is an efficient network of good roads, many urban areas have merged. A virtually continuous urban area has developed from Caguas to San Juan, and along the north coast from Fajardo through San Juan to Arecibo. Ponce on the south coast and Mayagüez on the west are

other urban cores. There is hardly a place on the island that is as much as an hour's drive from a major urban area. The largest metropolitan area is that of San Juan, which includes several municipalities in addition to the city proper. Each urban centre includes modern shopping centres and other services found in comparably sized cities on the U.S. mainland. Rural villages receive piped water, electricity, and other modern amenities. The government has had a long-standing policy to upgrade rural areas with programs that provide low-cost housing and incentives to disperse factories throughout the island.

The people. *Ethnic composition.* In 1493 Columbus found the island of Borinquen (the Amerindian name for Puerto Rico) to be inhabited by Taino Indians, a subgroup of the Arawak. About 200 years later the Taino population of some 30,000 had been reduced to only 2,000. Only a few African household slaves were brought to Puerto Rico in the early years of settlement, because the plantation system was not introduced until the early 19th century. For some 300 years Spanish males constituted the largest group of immigrants. The Spanish interbred with Amerindians and Africans with little social stigma. When slavery was abolished in 1873, only about 5 percent of the population was pure African. Some Chinese, Italians, Corsicans, Lebanese, Germans, Scottish, and Irish also found their way to the island in the mid-19th century. During this time the population grew steadily, becoming racially and culturally homogenous. People from the United States arrived after 1898, and the most recent large-scale immigration was the arrival of 23,000 Cubans after Fidel Castro came to power and an almost equal number of largely job-seeking people from the Dominican Republic. Although Puerto Rico's population today is racially mixed and there is no overt racism, pure Spanish bloodlines are still socially important for the elite group.

Major immigrations

Linguistic composition. Many English words have found their way into the Spanish of Puerto Rico, and numerous attempts at bilingual education have been made throughout the 20th century, but Puerto Rico remains predominantly Spanish-speaking.

Religion. Puerto Rico is primarily Roman Catholic in religious affiliation, a legacy of its centuries as a Spanish colony. In the 19th century the church's loyalty to Spain eroded much of its popular support. After 1898 Protestantism began to grow, and the single largest Protestant church is the Pentecostal Church of God. Puerto Rico's constitution guarantees freedom of religion.

Demographic trends. Overpopulation in Puerto Rico had become a critical problem by 1940. Health improvements as a result of the U.S. occupation had drastically reduced death rates, contributing to a population explosion that resulted in a 21-percent increase between 1930 and 1940. Because the island is relatively small and has limited physical resources, its already fragile economy and the quality of life were threatened. Birth control and mass migrations to the U.S. mainland slowed population growth beginning in the 1950s, but crowded conditions have continued to put pressure on the economy. Infant death rates have been reduced and life expectancy increased so that both figures approach U.S. standards. Birth rates, death rates, and educational achievement also compare favourably with the United States. Improved conditions have prompted a small movement from the mainland back to Puerto Rico, the rate of which has at times exceeded that of emigration.

The economy. Manufacturing and services produce about three-fourths of Puerto Rico's wealth. Before 1955, agriculture was the leading sector of the economy. Operation Bootstrap produced spectacular although not totally successful results. After the plan's initial effort to increase employment in cooperative agricultural enterprises and labour-intensive industries had failed, the government sought to bolster the infrastructure and to promote private enterprise. Low wage rates on the island, advantageous tax breaks, and government-supported start-up costs induced hundreds of manufacturing firms from the U.S. mainland to establish operations in Puerto Rico. These firms included those producing textiles, processed food, shoes, clothing, ceramics, tobacco, and wood products. Puerto Rico's Economic Development Corporation (Fomento in Spanish) guided economic development. To extend the early success of the Bootstrap program, Fomento in the 1960s made the decision to promote petrochemicals and other high-technology industries. Income from U.S. federal agencies operating in Puerto Rico and various social welfare programs also became vitally important in maintaining an adequate standard of living. Nevertheless, the unemployment rate has been high.

U.S. manufacturing investment

Resources. Other than picturesque beaches and a tropical climate, Puerto Rico is resource-poor. Limited flat land and good soils considerably handicap agriculture. Only clay, silica sand, stone, and limestone exist in sufficient quantities to allow industries to produce goods for domestic use and export. No commercial deposits of coal, oil, or natural gas are present, and the forests have limited value. Local waters produce good game fish, but commercial tuna fleets must fish off the coast of Africa or even the Pacific coast of the Americas to bring in their haul.

Agriculture, forestry, and fishing. Once the mainstay of the economy, primary activities now employ less than 4 percent of the labour force and contribute less than 3 percent of the gross national product (GNP). Plantation sugar production was firmly entrenched by U.S. corporations, and the industry dominated employment and income until the 1940s. Sugar production has become less significant but is still important for the molasses and rum industries. Coffee, tobacco, and milk are traditional farm products, but more land is being put into new, specialized products. Pineapples and other tropical and citrus fruits, beef, pork, poultry, and eggs are produced for local and export markets. Bamboo and tropical hardwoods support a small furniture industry. Local waters supply a variety of seafood for local markets. The tuna industry is a large-scale international operation that brings its catch from faraway grounds to plants around Ponce for processing and export.

Cash crops

Industry. Manufacturing, largely through investment by U.S. firms, is dominated by high-technology industries such as those producing petrochemicals, electronics, machinery, chemicals, pharmaceuticals, and medical equipment. More than 2,000 manufacturing establishments operate on the island. Most industrial components must be imported and final products are exported. Puerto Rican wage rates now equal those of the United States, and the island is no longer competitive in labour-intensive industries with many other Latin-American or Asian countries. Tourism has developed into a major industry.

Finance and trade. Finance and trade are dominated by U.S. firms. More than one-third of all U.S. investment in Latin America is located in Puerto Rico. Banks, retailers and wholesalers, restaurants, insurance companies, hotels, airlines, and many other firms in trade or commerce have profitable operations on the island. Guarantees of the U.S. Constitution, ready access to New York City's credit markets, exemption from U.S. taxes on profits, and being part of the U.S. monetary system all make investment easy and profitable. Puerto Rico's most important trading partner by far is the United States. The chief exports include petroleum products, food products, and chemicals; the main imports are machinery and transport equipment, manufactured goods, and food products.

Transportation. Puerto Rico has no public railway system for passengers or freight, but modern paved roads crisscross the island and rim the coasts. Most international commerce is carried on by ocean transport. San Juan's international airport, 5 miles outside the city, handles passenger and freight traffic, and it is conveniently located to serve the industries and hotels clustered there. (R.J.Ta.)

Administration and social conditions. *Government.* Puerto Rico's political status is officially described in its constitution as a "freely associated state" within the federal system of the United States. Puerto Ricans are U.S. citizens, and they elect a resident commissioner to the U.S. House of Representatives, who is allowed to speak but may not vote except in committees. (Thus, because they are without representation, Puerto Ricans do not pay federal taxes.) The 1952 commonwealth constitution was patterned after its U.S. counterpart, and it sets up executive, legislative, and judicial branches in a democratic

framework. The constitution may be altered by the commonwealth so long as the articles are not in conflict with the U.S. Constitution or the legal stipulations of Puerto Rican-U.S. relations.

Executive, legislature, and judiciary

The governor, who is elected by direct popular vote to a four-year term, heads the executive branch. The legislature comprises the Senate and the House of Representatives, whose members are elected for four years. There are eight senatorial districts (with two senators each) and 40 representative districts, and in addition 11 senators and 11 representatives are elected at large. A formula assures proportional representation of minority parties.

Customs taxes on foreign goods imported into Puerto Rico and excise taxes on goods sold in the United States are collected by the federal treasury. Relations between Puerto Rico and the United States are defined in the Puerto Rico-Federal Relations Act, which retains many provisions of the Foraker (1900) and the Jones (1917) acts. Local government—excluding San Juan, which has a city-management rule—is run by a popularly elected mayor and council.

Elections are held every four years, supervised by an electoral board comprising representatives from majority and minority parties. There are four principal registered parties: the New Progressive Party, the Popular Democratic Party, the Puerto Rican Independence Party, and the Puerto Rican Socialist Party. The two leading parties are the pro-statehood New Progressive Party, and the Popular Democratic Party, which supports the continuation of commonwealth status. The Puerto Rican Independence Party and the Puerto Rican Socialist Party advocate independence.

Justice. Puerto Rico has a unified court system, which is administered by the island's Supreme Court, whose justices are appointed by the governor with the advice and consent of the commonwealth Senate. A federal district court has jurisdiction over the application of federal laws in Puerto Rico, and appeals may be carried to the U.S. Supreme Court in Washington, D.C.

Education. Puerto Rico is deeply committed to the expansion of public education. The island invests about 7 percent of its GNP in education. Literacy has increased to about 90 percent with most children completing at least eight years of education; schooling is compulsory for children between six and 16 years of age. A serious problem is the high dropout rate. Vocational technical education has been stressed to combat the high rate of unemployment. The main institution of higher learning—the University of Puerto Rico, whose Rio Piedras campus was founded in 1903—has three campuses and provides programs ranging from agricultural science to medicine. There are also several private universities.

Literacy rate

Health and welfare. Urban clinics and rural health centres have been created to provide basic medical and health care. Health conditions are almost up to the standards of the U.S. states. The U.S. Medicare and Medicaid programs have contributed to improving health quality, as have various other social programs.

Puerto Rico has made impressive strides in meeting its housing shortage, although the pressures continue due to population growth. The Urban Renewal and Housing Corporation is in charge of broad and diversified housing programs, with concentration on low-income projects.

Cultural life. Puerto Rican cultural life is a blend of modern North American and traditional Latin and African forms. Dance, music, art, literature, and sports all display influences from the United States and Caribbean neighbours. The heritage from Amerindians is minimal because their numbers were small and they were quickly wiped out by the colonizers. African influences are evident in food, dress, music, and art. Music festivals, good museums in Ponce and San Juan, and theatre performances encourage "Hispanidad," or Spanish customs. The preservation of its Latin heritage, while at the same time embracing U.S. economic and social modernness, has become a Puerto Rican dilemma that often surfaces in political debate. The idealized folk hero of Puerto Rico is the *jíbaro,* the rugged, independent hill farmer of yore, who has a status similar to the gaucho of Argentina in local song and story.

Literature and theatre. Contemporary poets, novelists, short-story writers, and essayists keep alive the traditions of such 19th-century forerunners as the novelist Alejandro Tapia, the essayist Eugenio María de Hostos, and the poet José Gautier Benítez. New playwrights and artists have also received considerable encouragement from the Institute of Culture.

Music. Nineteenth-century composers included Manuel Tavárez and Juan Morel Campos, both known for their dance melodies. The popular 20th-century songwriter Rafael Hernández is still revered, and singers such as Antonio Paoli, in the early part of the 20th century, and José Feliciano achieved international reputations. The noted Spanish cellist Pablo Casals, whose mother was Puerto Rican, moved to the island in 1957 and founded the world-famous music festival that bears his name and that attracts international performers to San Juan every June. Latin and African folk music is preserved by many groups.

Pablo Casals music festival

Recreation. Typical of Latin-American nations, Puerto Rico celebrates numerous religious holidays with colourful festivals, but as a U.S. territory it also observes many of the secular holidays of the mainland. Baseball is Puerto Rico's national sport, and the island's professional league is a testing ground for U.S. major league prospects. Basketball is also popular among the people, and the usual Caribbean activities are broadly available to tourists. Cockfighting is a popular gambling activity.

Press and broadcasting. The commonwealth has a free press that may be independent or politically partial. Both local and major U.S. newspapers are widely available. The major dailies published on the island include *El vocero de Puerto Rico* ("The voice of Puerto Rico") and *El Nuevo Día* ("The new day"). Radio and television are well established, the first local radio broadcast dating from 1923 and television from 1954. Both commercial and public channels offer programming comparable to that of the U.S. mainland. (A.M.-C./R.J.Ta.)

For statistical data on the land and people of Puerto Rico, see the *Britannica World Data* section of the BRITANNICA WORLD DATA ANNUAL.

HISTORY

The first inhabitants of Puerto Rico, probably from the Florida Peninsula, reached the island more than 1,000 years before the arrival of the Spanish. These primitive inhabitants collected food from the seashore and wild fruit from the land. By the year AD 1000 Arawak Indians, who developed the Taino culture, had arrived by way of the Lesser Antilles from the tropical forests of South America. The Arawak, living in small villages, were organized in clans and led by a cacique, or chief. They were a peaceful people who, with a limited knowledge of agriculture, lived on such domesticated tropical crops as pineapples, cassava, and sweet potatoes supplemented by seafood. Anthropologists estimate their numbers to have been between 20,000 and 50,000. On a fertile island the Arawak lived an easy life disturbed only by occasional visits from their Carib neighbours on the islands to the south and east. At the time of discovery, Carib Indians occupied most of the Lesser Antilles, the Virgin Islands, and Vieques Island.

In 1493 Christopher Columbus left Spain on his second voyage to the Indies with an elaborate expedition of 17 ships and about 1,500 men. At the island of Guadeloupe the Spaniards rescued several Arawak Indians who had been taken from Boriquén, the Indian name for Puerto Rico, by the Caribs. Columbus agreed to return them to their island, and on Nov. 19, 1493, the expedition anchored in a bay on the west coast of Puerto Rico. Columbus formally took possession of the island in the name of Ferdinand and Isabella, the rulers of Spain, and named it San Juan Bautista. Two days were spent on the island before the ships moved westward to Hispaniola, where the first settlement in the New World was established.

Columbus' visit to Puerto Rico

Spanish rule. *Early settlement.* For 15 years the island was neglected except for an occasional visit by a ship putting in for supplies. In 1508 Juan Ponce de León, who previously had accompanied Columbus, was granted permission to explore San Juan Bautista in recognition of his valuable colonizing efforts in eastern Hispaniola. On the

north coast, Ponce de León found a well-protected bay that could offer safe harbour for a large number of sailing vessels, and he founded the first town, Caparra, the site of the first mining and agricultural efforts. The harbour was named Puerto Rico because of its obvious excellent potentialities. In this area was located the most important settlement on the island; through time and common use the port became known as San Juan, while the name Puerto Rico came to be applied to the whole island.

The peaceful and friendly relations with the Arawak did not last long. The Spaniards expected the Indians to acknowledge the sovereignty of the king of Spain by payment of gold tribute. The Indians were to be instructed in Christian ways. In return for this education, which was rarely given, the Arawak were expected to work and supply either more gold or provisions of food. In 1511 the Indians rebelled against the Spanish, who with their superior arms rapidly subjugated them.

Placer mining of gold was continued by Indians brought from other islands and by blacks brought from Africa by some of the early traders. After the 1530s, however, gold production markedly declined with dwindling Indian labour, and the Spanish colonists, with slaves from Africa, turned to agriculture.

Puerto Rico, however, did not prosper economically. Carib Indians from neighbouring islands made frequent raids, carrying off food and slaves and destroying property. The colony continued to lead a precarious existence, ravaged by plagues and plundered by French, British, and Dutch pirates. Repeatedly during the mid-16th century the French burned and sacked San Germán, the second community to be established on the island. People began to leave the island at every opportunity.

The fortification of San Juan

In the second half of the 16th century Spain, recognizing the strategic importance of Puerto Rico, undertook to convert San Juan into a military outpost. The fortress El Morro, built with the financial subsidy from the Mexican mines, was well constructed and perfectly located to dominate the narrow entrance to the harbour. Later, a stronger and larger fortress was built to the east and on the Atlantic side of the city. In the early 17th century the city was surrounded by a stone wall, 25 feet high and 18 feet thick, two parts of which still stand. These defenses made San Juan almost impregnable.

Sir Francis Drake attacked the town in 1595 but failed to gain the harbour. Three years later George Clifford, 3rd Earl of Cumberland, had complete military success but was forced to abandon his conquest owing to an outbreak of plague among his troops. In 1625 a Dutchman, Bowdoin Hendrik, sailed into the harbour, captured and burned the town, but failed to subdue El Morro.

San Juan, as the most exposed military outpost guarding the heart of Spain's New World empire, received political and economic attention from the mother country. The rural inhabitants of the interior of the island, however, were ignored by Spain and scorned by the presidial residents of San Juan. As the French, English, Danish, and Dutch fought over and settled the Lesser Antilles during the 17th and 18th centuries, rural Puerto Ricans, ignoring the edicts of Spain, found profit in clandestine trade. Ginger, hides, sugar, tobacco, and cattle from the island were in great demand, and while the colonial authorities of San Juan rarely ventured out of their walled defenses for fear of the reprisals of the buccaneers, the rural settlers prospered in a modest way through contacts with the non-Spanish European traders. No large plantations were established, and the farmer, with little help, cultivated his own land. Contrary to the fears of Spain, this contact with foreigners did not corrupt the islanders, who remained loyal and were willing to participate in aggressive expeditions.

The British attack on Puerto Rico

Liberal reforms. In 1797 the British general Sir Ralph Abercromby, who had captured Trinidad, unsuccessfully attacked Puerto Rico. The British considered the island—a centre of clandestine trade and of operations for quasi-piratical expeditions and a refuge for runaway slaves—a weak link in the chain of defense of the Spanish empire. The failure of Abercromby was due in part to the important economic and administrative changes in the Spanish colonial empire that were carried out in the latter half of the 18th century by representatives of the Bourbon rulers of Spain. In the case of Puerto Rico it was hoped that the island might become an economic asset rather than a financial drain on the Spanish crown. Trade relations between the island and Spain were liberalized, agricultural production was stimulated, and the island as a whole was integrated into the system of military defense.

The liberal reforms of the enlightened despotism of the Spanish Bourbons coincided with and encouraged rapid population growth, introduction of new products, and the beginning of commercial agriculture. Population was estimated in 1765 at 45,000, in 1775 at 70,250, in 1787 at 103,051, and in 1800 at 155,426. By the end of the 18th century there were 34 towns on the island. Immigrants from the Canary Islands, French settlers from Louisiana or Haiti, and Spaniards from Santo Domingo, which had been turned over to Napoleon, accounted in part for the increase in population. These newcomers brought with them new ideas and methods of producing marketable crops. Coffee, introduced into the island in 1736, became an important export item by 1776. Sugar production, which had always been small, was undertaken on a large scale by augmented slave labour. From 1765 to 1800 the slave population increased from 5,037 to 13,333.

When Napoleon invaded Spain and placed his brother Joseph Bonaparte on the Spanish throne (1808), the colonies of South and Central America asserted their right to govern themselves in the name of the imprisoned Bourbon king, Ferdinand VII. This claim to temporary self-rule eventually evolved into a revolutionary movement for independence. In Puerto Rico, however, for various reasons, the sequence of events and their results were different. The communities of the interior of Puerto Rico, with one exception, offered little objection to the strict rules of Spain's mercantilist policy, which for many decades had ceased to have effect on them. Most of the residents of San Juan, on the other hand, dependent upon administrative and military positions, were most willing to follow the orders of the central government of Spain.

As the revolutions progressed on the southern and central mainlands, loyal Spaniards reluctant to leave the colonies found refuge in Puerto Rico. In recognition of its loyalty and in a belated move to liberalize an outmoded colonial system, the Spanish government granted Puerto Rico in 1815 ample economic liberties. The island was opened to all non-Spanish Catholics, the ports were permitted to trade with non-Spanish countries, and free land was granted to the new settlers. After 1830 Puerto Rico gradually developed into a plantation economy based on three main crops: sugar, coffee, and tobacco. Sugar and molasses, sold for the most part in the U.S. market, provided an important source of income for the Spanish government. Foreign settlers contributed to economic development, though the Spanish element attempted to maintain a tight monopoly.

Economic and political development. By the end of the 19th century the population had increased to nearly one million, and the value of foreign trade had increased considerably. By 1899 the United States was buying almost two-thirds of Puerto Rican sugar production. The area devoted to sugar had been slowly expanding. Coffee, in the late 19th century, provided the principal source of income for the island.

U.S. sugar market

Political development in Puerto Rico during the 19th century was characterized by periods of liberal advance counteracted by long periods of conservative reaction. In part this was due to the changes occurring in the Spanish government, and in part due to the antiquated Spanish colonial administrative policy. During the first half of the 19th century, two short periods of relative political freedom were enjoyed. From 1809 to 1814 and from 1820 to 1823 Puerto Rico was declared an integral part of Spain with the right to elect representatives to the Spanish Cortes, or parliament. Ramón Power y Giralt, an able liberal, was selected during the first period and succeeded in revoking the absolute powers of the island's colonial governor. In the latter period Demetrio O'Daly secured the separation of the military authority from the colonial administrator. Freedom of the press was also permitted.

On each occasion moderate colonial rule was thwarted by the return of royal absolutism in Spain.

In 1837, when a fairly permanent constitutional monarchy was established in Spain, Puerto Rico failed to benefit because it was argued that the colonies were not true Spanish provinces and therefore should be governed by special laws. For more than 30 years Puerto Rico waited for special legislation to ease the despotic rule of military colonial governors. During this waiting period political thought in the island began to crystallize. A liberal current of opinion requested assimilation into the Spanish government and permission to be represented in and governed by the Cortes. A bloc of conservative opinion strongly approved of the status quo. A small third group advocated complete independence.

Movements toward self-government and independence. A local commission was elected in 1865 to draw up a report on the basis of which a governmental reform might be carried out. The majority report, which declared that the abolition of slavery was the sine qua non of any political reform, provoked a shocked reaction among the island and peninsular conservatives. The alarmed colonial government took steps to curtail what was feared to be a growing movement of rebellion. Some of the more outspoken and respected islanders were ordered to be arrested and sent to Spain for trial. Thus provoked, a small group of radicals committed to independence attempted an uprising, for which, however, inadequate preparation was made. El Grito de Lares, the abortive revolt of Sept. 23, 1868, brought forth severe reprisals on all island liberals. Though separatist elements joined the Cubans who were struggling for independence, they were unable to challenge Spanish power effectively. However, the abdication of Queen Isabella II of Spain was forced by a republican government that pardoned all political prisoners. The first Spanish republic extended to Puerto Rico its third period of constitutional government, 1868–74, during which slavery was abolished.

Reprisals against island liberals

During the 1880s a movement for political self-government under Spain led by Román Baldorioty de Castro replaced the sentiment in favour of integrating Puerto Rico into the Spanish government. Again liberal political movement, this time autonomy, was denounced as disloyal and was violently suppressed in 1887. Such treatment only served to solidify the movement for local self-government, and in 1897 the Autonomy Party, through cooperation with the Liberal Party in Spain, achieved its objective. The autonomous government granted was parliamentary in form but retained the governor-general as a representative of the Spanish king. He was empowered to disband the insular parliament and suspend civil rights. The two-chamber parliament was empowered to legislate for the island, create and control an insular tariff, and levy local taxes.

The Spanish-American War. The Spanish-American War (1898) prevented the islanders from putting into effect the new government. In May 1898 Admiral W.T. Sampson bombarded San Juan for a short time without serious results. Facing token military resistance and with general popular acceptance, General Nelson A. Miles landed a U.S. force of about 3,500 men in July, and, after a short campaign, hostilities were ended on August 12.

Under the United States. *Early years.* On Oct. 18, 1898, the island was turned over to the U.S forces, and General John R. Brooke became military governor. Puerto Rico was ceded to the United States by the Treaty of Paris, signed Dec. 10, 1898 (ratified by the U.S. Senate Feb. 6, 1899). In the work of policing the country, in the accompanying tasks of sanitation, construction of highways and other public works, accounting for the expenditure of public funds, and establishing a system of public education, the military control that lasted until May 1, 1900, proved effective in bridging the period of transfer from the control of Spain to the system under U.S. civil government. The U.S. military, however, ruled with little regard for political sensitivities. The U.S. Congress passed the Foraker Act, under which civil government was instituted in May 1900. Under this act the United States exercised the controlling power, a condition that proved distasteful to many Puerto Ricans; the organic law was subsequently amended to give a wider native participation in the government. The Olmsted Act, approved by Congress on July 15, 1909, placed the supervision of Puerto Rican affairs in the jurisdiction of an executive department to be designated by the president. The people, however, demanded a larger measure of local control. The majority also asked for U.S. citizenship and many other changes. As a result, Congress passed a new organic act (the Jones Act), which came into effect on March 2, 1917. Under its terms Puerto Rico became a territory of the United States "organized but unincorporated," and citizenship of the United States was conferred collectively on Puerto Ricans, allowing the right to retain the old status if preferred. Only 288 persons declared in favour of the latter. The local civil government, however, even with modifications, fell far short of the measure of self-government that Puerto Ricans expected in light of the democratic tradition of the United States. Key officials, including the governor, were presidential appointees and thus beyond local control.

The Olmsted Act

In spite of the legal limitations on political autonomy, a climate of freedom was slowly developed as a result of the change of sovereignty. At first this new order was sometimes mistrusted, resented, and misunderstood, but in the long run it was recognized as beneficial and assimilated by the islanders. For example, the separation of church from state, resulting in open competition for religious adherence, demonstrated the new climate in a practical way; government programs that dealt directly with the vital needs of the people for education, health and sanitation, and regulation of working conditions all reflected a change designed to remedy centuries of neglect.

Economic and social changes. Early U.S. governors were mainly preoccupied with Americanizing Puerto Rican institutions, language, and political habits but had no clear policy on the island's eventual political status. This approach created strong resistance from many native leaders led by Luis Muñoz Rivera, who had fought for autonomy under Spain. The economic reorientation of the island as a result of the change in sovereignty had almost an immediate and profound effect on all aspects of life. Included within the U.S. tariff walls, Puerto Rican agricultural products, particularly sugar, had a ready market. Aided by the adoption of U.S. currency and by unobstructed financial movement, Puerto Rico experienced within a short period a large capital investment that revolutionized the production of sugar. A sevenfold expansion in acreage between 1899 and 1939, new disease-resistant plants, rapid-transportation facilities, large and efficient cane-grinding mills, and complete corporate management within a generation converted the economy of the island into one in which 75 percent of the population directly or indirectly was dependent upon sugar. The population increased from about 950,000 in 1899 to more than 1,540,000 in 1930. Glaring inequalities of wealth contributed to sharpened social and political tensions.

Increased sugar production

The island was forced to import much of its food. Coffee was neglected at a time when weather conditions and transportation problems dictated financial and government aid. Only tobacco production experienced growth, which failed to be sustained after the 1920s when U.S. smokers shifted from cigars to cigarettes.

The shock of these economic changes might have been absorbed in spite of the island's limited resources if at the same time Puerto Rico had not been undergoing a severe social change as a result of the application of modern sanitation means and medical knowledge to a people with a very high death rate. The population was threatening to double its number in two generations. The two counterpressures—expansion of corporate control over the limited productive land and increasing population pressure—reached an explosive stage when the economic depression occupied the attention of government officials in the United States. Recurring hurricanes joined with declining exports to aggravate the economic distress of the island.

Political development. With one exception political parties that had developed since the change in sovereignty had centred their attention on modifications in the political relations between the island and the U.S. federal gov-

ernment. The Republican Party limited its program to a plea for statehood for the island. The Union Party worked for greater autonomy. In the 1920s the Nationalist Party rose to affirm the ideal of immediate independence. The one exception was the pro-U.S. Socialist Party, led by the highly respected labour leader Santiago Iglesias. This party had expressed since its foundation a concern for the plight of the labouring classes of the island. Nevertheless, its effectiveness had been hampered by insufficient popular support, due primarily to the concentration of attention upon the issue of the political status of the island.

Effects of the New Deal policies on Puerto Rico

In the mid-1930s, with President Franklin D. Roosevelt's New Deal policies radically enlarging the previously accepted concept of the function of government, Puerto Rico was not neglected. More important than the much-needed temporary relief were the steps taken by the Puerto Rican Reconstruction Administration (PRRA), designed to readjust the distribution of economic power on the island. A restrictive quota was placed over sugar production. Legal procedures were initiated to enforce a long-neglected law limiting corporate holdings to 500 acres. Thus the process of increasing the sugar acreage was to be reversed, and Puerto Ricans were to be returned to their small farms.

This radical program provoked the open opposition of the sugar interests, locally vocal through the Republican Party. The Socialists accepted the program in a tacit fashion. Their reluctance was due to the fact that the young radical wing of the heirs of the Autonomy Party, led by Luis Muñoz Marín, the son of Luis Muñoz Rivera, was recognized in Washington and on the island as the local political proponent of the economic reform.

The success of the New Deal measures was jeopardized by two unconnected factors. Unforeseen administrative and financial problems forced a curtailment of the objectives of the PRRA. No longer was a complete readjustment of the island's economic structure possible; the PRRA took on a more temporary or experimental nature. The second factor was the interjection of the status issue on the political scene by the U.S. government in answer to Nationalist violence. Taking the form of a vindictive offer of independence under adverse economic conditions, the proposal served to realign again the political parties into pro- and anti-independence groups.

The incipient political movement for economic reform originally fostered by the New Deal and temporarily sidetracked was surprisingly successful in the election of 1940.

Ascendency of Muñoz Marín

This new political movement took the form of a political party, led by Muñoz Marín, called the Popular Democratic Party (PPD). Organized to improve the conditions of the lower classes, particularly the hardworking *jíbaro* of the mountainous interior, the new party's platform was summarized by the slogan "Bread, land, and liberty." The island electorate had agreed that the political status was not in issue and that economic and social problems took precedence. Tenuous control over the island legislature and a new-style colonial governor, Rexford Guy Tugwell, allowed the PPD to initiate such economic reforms as redistribution of land, enforcement of minimum wage and hour laws, an enforced progressive income tax law, and the establishment of an economic development program. In recognition of partial fulfillment of its announced aims the PPD was overwhelmingly backed by the island electorate in 1944. In 1946 President Harry S. Truman named Jesús T. Piñero, a Puerto Rican, as governor, the first Puerto Rican to occupy that post. In 1947 the U.S. Congress amended the organic act of Puerto Rico to permit election of governors by popular vote. Muñoz Marín was elected Nov. 2, 1948, and took office in January 1949.

For more than a generation the PPD, led by Muñoz Marín, governed Puerto Rico. Muñoz served for four terms as governor and was followed by his able administrative assistant Roberto Sánchez Vilella. In 1968, because of a split in its ranks, the PPD lost control of the lower house of the legislature and also relinquished the office of the governor to the pro-statehood New Progressive Party (PNP), led by the industrialist Luis A. Ferre, who became the governor of Puerto Rico.

From 1948 to 1968, under the PPD, Puerto Rico experienced a major economic change. An economy based on agriculture was transformed into one based on industrial production. Favourable tax laws encouraged the establishment of new industries such as those for electronic and pharmaceutical products. Workers left the sugarcane fields and the coffee haciendas and moved into the coastal cities where wages, working conditions, and social services were better. Many also migrated to the large metropolitan centres in the mainland United States.

Establishment of the commonwealth. The PPD also modified the political relationship with the United States. Responding to an offer made in 1950 by the U.S. Congress, the Puerto Rican people participated in a Constitutional Assembly that drew up a document establishing the Commonwealth of Puerto Rico. Overwhelmingly approved in a referendum by the voting population and by the U.S. Congress, the commonwealth constitution was proclaimed on July 25, 1952. This process reaffirmed the post of an elected governor, created a legislative branch in which minority representation was guaranteed, and set up a modern judicial system based on civil liberties.

The commonwealth constitution

In spite of broad popular support for the autonomy of the commonwealth government and a rapidly modernizing industrial society, there were expressions of dissatisfaction. Puerto Rican extremists dramatized their desire for independence with an attempt to assassinate President Truman on Nov. 1, 1950, and on March 1, 1954, they wounded five congressmen in an attack in the U.S. Capitol.

Efforts to improve the commonwealth by expanding its local autonomy to include, for example, a greater presence in foreign, particularly regional or Caribbean, affairs failed. Nevertheless, legal reviews in the courts, both insular and federal, continued to enforce the commonwealth concept. The granting of statehood to Alaska and Hawaii, the growing dependence on federal programs for the unemployed, the aged, and the veterans of the Korean and, later, Vietnam wars, and the sizable influx of Cuban exiles all contributed to a growing sentiment for statehood; a commission was appointed to explore the political options. In a 1967 plebiscite, however, the commonwealth status was overwhelmingly approved, although the pro-statehood faction received 39 percent of the vote and the pro-independence group largely boycotted the plebiscite.

Although commonwealth status won impressive support, both the leaders of the PPD and influential members of the U.S. federal government, following the recommendations of the Status Commission, recognized that the commonwealth relationship needed to be improved and the degree of self-government broadened even further. In the 1970s there was an increasing number of expressions of dissatisfaction with the commonwealth status. Terrorist bombings in San Juan and on the mainland in the 1970s were reported to be linked to the independence movement. A special commission formed in 1970 unanimously declared that Puerto Ricans should be allowed to vote for the U.S. president and vice president.

Through the 1970s and '80s political sentiment on the island came to be evenly divided between autonomy and statehood, with the small pro-independence group, estimated at 10 to 15 percent of the population, sometimes serving to swing the balance of the political process in one direction or the other. The PPD returned to power briefly for a four-year period from 1972 to 1976, but the pro-statehood party regained power in 1976 under younger, more vigorous leadership. However, violent and abusive treatment of pro-independence supporters by the police, with tacit government approval, brought a change with the 1984 election, returning a second generation of PPD members to power by a narrow margin under the leadership of Rafael Hernández Colón. This narrow victory was repeated by the PPD in 1988. While the events of the second half of the 20th century clearly indicate that the political status issue of Puerto Rico has not been resolved, the broad and general consensus has supported a permanent relationship with the United States, be that under the existing commonwealth government or eventual statehood. (T.G.Ms.)

For later developments in the history of Puerto Rico, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

THE LESSER ANTILLES

Anguilla

Anguilla, an island in the eastern Caribbean Sea, is a dependent territory of the United Kingdom. It lies about 60 miles (100 kilometres) northwest of Saint Kitts and is the most northerly of the Leeward Islands in the Lesser Antilles. Anguilla has an area of 35 square miles (91 square kilometres). The Valley is the principal village and the administrative centre of the island.

PHYSICAL AND HUMAN GEOGRAPHY

The land. Anguilla is bare and flat and is fringed by white sand beaches. Its long thin shape (16 by at most three miles) gave it its name (French: *anguille,* "eel"). Anguilla has several small uninhabited offshore islands. The largest are Dog, Scrub, and Sombrero islands and Prickly Pear Cays.

Anguilla is of coral limestone formation; the highest point is 213 feet (65 metres). Soil is thin, but there are small pockets of red loam soil, mainly in the shallow valleys. As with most coral islands, water is a scarce commodity. The climate is tropical; the average temperature is about 80° F (27° C), and rainfall averages 35 inches (889 millimetres) per year. Hurricanes can occur from July to October. The vegetation is primarily low scrub, although there are some plantations of fruit trees.

The people. The majority of the population of Anguilla is of African descent. The official language is English, and the main religious denominations are Anglican and Methodist.

The economy. Anguilla is dependent on financial services, remittances from any emigrants, and tourism. The steady increase in tourist arrivals has also bolstered the construction industries and stimulated the improvement of transport facilities. There is some boatbuilding. Both deepwater fishing and fish farming are expanding, but there has been some decline in the lobster catch.

Executive power

Administration and social conditions. Executive power is in the hands of a governor, appointed by the British monarch. The governor is in charge of external affairs, defense, internal security (including police), and public services. The Executive Council is composed of a chief minister, other ministers, and ex officio members. The House of Assembly consists of members elected by universal adult suffrage, members appointed by the commissioner after consultation with the chief minister, and ex officio members.

Health services on the island are limited. There is a small cottage hospital on the island, but specialist care must be sought from larger islands nearby.

For statistical data on the land and people of Anguilla, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

The island was reached by Christopher Columbus in 1493, and, after becoming a British colony in 1650, it was administered as part of the Leeward Islands. From 1825 administration was largely through Saint Kitts, and Anguilla was incorporated with Saint Kitts and Nevis in 1882, a situation to which it strenuously objected. In 1967 Anguilla, complaining of domination by the Saint Kitts administration, ejected the Saint Kitts police and set up its own council, subsequently proclaiming its independence. After negotiations failed, the British intervened in March 1969 to restore legal government with troops and a temporary commissioner. The troops were withdrawn in September 1969, and the Anguilla Act of July 1971 placed Anguilla directly under British control. Anguilla formally became a dependent territory of the United Kingdom in 1980, and a new constitution became effective in 1982.

(J.D.Mo.)

For later developments in the history of Anguilla, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Antigua and Barbuda

Antigua and Barbuda are islands that form an independent state in the Lesser Antilles in the eastern Caribbean Sea. They lie at the southern end of the Leeward Islands chain and have an area of 171 square miles (442 square kilometres). There is one dependency, Redonda. The capital is St. John's, on Antigua.

PHYSICAL AND HUMAN GEOGRAPHY

The land. Antigua's coastline is intricate, with bays and headlands fringed with reefs and shoals; several inlets, including Parham and English Harbour, afford anchorage for shipping, and St. John's has a deepwater harbour. The island has an area of 108 square miles. It is mostly low and undulating, but in the west there are volcanic rocks that rise to 1,330 feet (405 metres) at Boggy Peak. An absence of mountains and forests distinguishes Antigua from the other Leeward Islands. Because there are no rivers and few springs, droughts occur despite a mean annual rainfall of some 40 inches (1,016 millimetres). The average January temperature is around 77° F (25° C); that of August, 82° F. Summer highs can reach 90° F.

Climate

Barbuda, formerly Dulcina, lies 25 miles (40 kilometres) north of Antigua. A coral island, flat and well-wooded, with highlands rising to 143 feet at Lindsay Hill in the northeast, it is 62 square miles in area. Barbuda is without streams or lakes and receives less rainfall than Antigua. Codrington, the only settlement, lies on a lagoon to the west. The climate is similar to that of Antigua.

Redonda, an uninhabited rock, lies 25 miles southwest of Antigua. It rises sheer to a height of 1,000 feet and is 0.5 square mile in area. Phosphate deposits are located there.

The people. The majority of the population is of African descent. A large part of the population lives in St. John's. The language is English, and most people are Anglican, with minorities of other Protestant sects and of Roman Catholics.

The economy. Agriculture was once the mainstay of the economy but has been largely supplanted by tourism. Sugar was long the dominant crop on Antigua, but its production is now insignificant. Barbuda was never involved in the sugar plantation system, its inhabitants always having been fishermen and subsistence farmers. Their traditional system of land tenure is threatened by tourism development. Fruits and vegetables, including mangoes, cucumbers, pineapples, and eggplants, are now cultivated on the islands. Manufacturing plays a small role in the economy: beverages, chemicals, concrete blocks, and paper products are major commodities. An international airport is near St. John's.

Administration and social conditions. Antigua and Barbuda is a constitutional monarchy. The British monarch is nominal head of state, represented by a governor-general. The constitution allows for a Senate and a House of Representatives. Executive power is vested in a Council of Ministers headed by the prime minister. Primary and postprimary education is compulsory.

Head of state

For statistical data on the land and people of Antigua and Barbuda, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

Antigua was visited in 1493 by Christopher Columbus, who named it after the Church of Santa Maria de la Antigua in Seville, Spain. It was colonized by English settlers in 1632 and remained a British possession although it was raided by the French in 1666. The early colonizers were also attacked by Caribs, who were one of the dominant Amerindian peoples of the West Indies. At first tobacco was grown, but in the later 17th century sugar was found to be more profitable.

The nearby island of Barbuda was colonized in 1678. The crown granted the island to the Codrington family in 1685. It was planned as a slave-breeding colony but never

became one; the slaves who were imported came to live self-reliantly in their own community.

The emancipation in 1834 of slaves, who had been employed on the profitable sugar estates, gave rise to difficulties in obtaining labour. An earthquake in 1843 and a hurricane in 1847 were further economic blows. Barbuda reverted back to the crown in the late 19th century, and its administration came to be closely related to that of Antigua, eventually becoming a dependency of that island.

The Leeward Islands colony, of which the islands were a part, was defederated in 1956, and in 1958 Antigua joined the West Indies Federation. When the federation was dissolved in 1962, Antigua persevered with discussions of alternative forms of federation. Provision was made in the West Indies Act of 1967 for Antigua to assume a status of association with the United Kingdom on Feb. 27, 1967. As an associated state, Antigua was fully self-governing in all internal affairs, while the United Kingdom retained responsibility for external affairs and defense.

Independence movement

By the 1970s Antigua had developed an independence movement, particularly under its prime minister George Walter, who wanted complete independence for the islands and opposed the British plan of independence within a federation of islands. Walter lost the 1976 elections to Vere Bird, who favoured regional integration. In 1978 Antigua reversed its position and announced it wanted independence. The autonomy talks were complicated by the fact that Barbuda, long a dependency of Antigua, felt that it had been economically stifled by the larger island and wanted to secede. Finally, on Nov. 1, 1981, Antigua and Barbuda achieved independence, with Vere Bird as the first prime minister. The state obtained United Nations and Commonwealth membership and joined the Organization of East Caribbean States. Bird's party won again in 1984 and 1989 by overwhelming margins, giving the prime minister firm control of the islands' government.

(R.To./D.L.N./J.D.Mo.)

For later developments in the history of Antigua and Barbuda, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Barbados

Barbados is an independent island nation in the Caribbean, situated about 100 miles (160 kilometres) east of the Windward Islands. Roughly triangular in shape, it measures 21 miles from northwest to southeast and about 14 miles from east to west, with a total area of 166 square miles (430 square kilometres). Its capital is Bridgetown, the only seaport.

Barbados is not part of the Lesser Antilles, although it is sometimes grouped with this archipelago. The island is of different geologic formation; it is less mountainous and has less variety in plant and animal life. The geographic position of Barbados has profoundly influenced the island's history, culture, and aspects of its economic life. In the era of sailing ships, access to the island was difficult because of the prevailing winds from the northeast. Outward-bound ships from Europe had to gain the island while heading west, for it was difficult for them to turn and reach its shores by sailing eastward against the wind. It remained a British possession without interruption from its settlement in the 17th century to 1966, when it attained independence. As the first Caribbean landfall from Europe, Barbados has functioned since the late 17th century as a major link between western Europe (mainly Britain), eastern Caribbean territories, and parts of the South American mainland. Because of its long association with Britain, the culture of Barbados is probably more British than that of any other Caribbean island. Since independence, however, cultural nationalism and regional awareness have tended to increase.

PHYSICAL AND HUMAN GEOGRAPHY

The land. The rocks underlying Barbados consist of sedimentary deposits, including thick shales, clays, sands, and conglomerates, laid down approximately 70 million years ago. Above these rocks are chalky deposits, which were capped with coral before the island rose to the surface. A layer of coral up to 300 feet (90 metres) thick covers the island, except in the northeast physiographic region known as the Scotland District, covering 15 percent of the area, where erosion has removed the coral cover. The government has adopted a conservation plan to prevent further erosion.

The coral surface

Relief, drainage, and soils. Mount Hillaby, the highest point in Barbados, rises to 1,115 feet (340 metres) in the north central part of the island. To the west the land drops down to the sea in a series of terraces. East from Mount Hillaby, the land declines sharply to the rugged upland of the Scotland District. Southward, the highlands descend steeply to the broad St. George Valley; between the valley and the sea the land rises to 400 feet to form Christ Church Ridge. Coral reefs surround most of the island.

There are no rivers or lakes and only a few streams, springs, and ponds. Rainwater percolates quickly through the underlying coralline limestone cap, draining into underground streams that discharge off the leeward coast. These streams are the main source of the domestic water supply.

Barbados has mainly residual soils. They are clayey and rich in lime and phosphates. Soil type varies with altitude; thin black soils occur on the coastal plains, and more-fertile yellow-brown or red soils are usually found in the highest parts of the coral limestone.

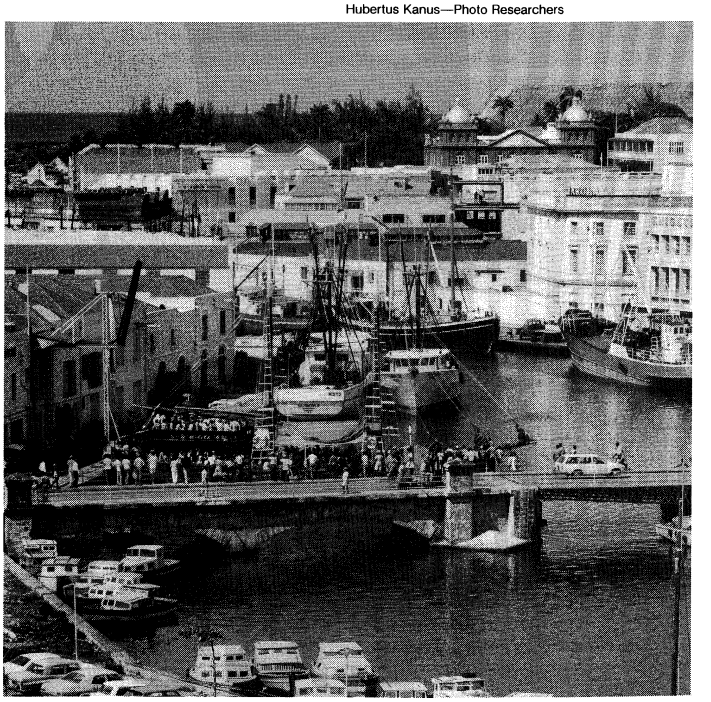
Climate. The climate is generally pleasant. The temperature does not usually rise above 86° F (30° C) or fall below 72° F (22° C). There are two seasons: the dry season, from early December to May, and the wet season, which lasts for the rest of the year. Average rainfall is about 60 inches (1,525 millimetres) a year, but, despite the small size of the island, rainfall varies, rising from the low-lying coastal areas to the high central district. Barbados lies in the southern border of the Carribean hurricane zone, and hurricanes have caused great devastation.

Rainfall

Plant and animal life. Very little of the original vegetation remains on Barbados; the pale green of sugarcane has become the characteristic colour of the landscape. Tropical trees, including poinciana (flamboyant), mahogany, frangipani, and cabbage palm, are widespread, and flowering shrubs adorn parks and gardens.

The few wild animals, such as monkeys, hares, and mongooses, are considered pests by farmers. Birds include the dove, hummingbird, sparrow, egret, and yellow breast. The marine life includes flying fish, sprat, green dolphin, kingfish, barracuda, mackerel, and parrot fish.

Hubertus Kanus—Photo Researchers



The cramped docking facilities of the Careenage, the old harbour, at Bridgetown, Barbados.

Settlement patterns. Barbados is densely populated. More than one-third of the population is concentrated in Bridgetown and the surrounding area. Most of the farmland is owned by large landowners or corporations. As a result, "tenantries" are as common as villages. Tenantries are clusters of wooden houses—locally known as chattel houses—located on the borders of the large estates; they are usually owned by the occupants but stand on rented ground from which they may easily be removed. Most of them have electricity and running water. The largest town is Bridgetown. In its commercial and administrative centre, multistory buildings are altering the features of the 19th-century town. Apart from Bridgetown, Oistins, Holetown, and Speightstown are the largest towns.

The people. Blacks make up more than 90 percent of the population; the remainder consists of whites, persons of mixed African and European descent, and East Indians. English is the official language, and a nonstandard English called Bajan is spoken. The Anglican church has the largest congregation. About a quarter of the population belongs to other Protestant churches, and there is a small number of Roman Catholics.

Demographic trends

Since the 1950s the rate of population growth has been slowed by a successful family-planning program and by emigration, now mostly to other parts of the Caribbean and to North America. In the same period the death and infant mortality rates declined sharply, and life expectancy rose above 70 years.

The economy. Barbados has a small, market-oriented, developing economy. Services, manufacturing, and agriculture are the main productive sectors. Although Barbados had a relatively high per capita growth rate in the 1980s, unemployment, especially among the youth and women, has been a serious problem. Most of the employment is in services and distributive trades, the greater part of which has been unionized.

Resources. Apart from some small deposits of oil and natural gas, Barbados has few natural resources. Sustained exploitation of the climate and beaches for their tourist potential has been the most impressive feature of ongoing economic activity. An overly abundant population may also be considered one of the island's resources. This has always provided a cheap labour source, and the population working abroad has made significant contributions to the economy through remittances.

Agriculture and fishing. About three-quarters of the land is arable, and most of it is planted with sugarcane. Sugar production dominated the economy until the 1950s, but that industry has since declined in importance. Agricultural production remains dominated by large farm units, but the pattern of production has changed, mainly as a result of falling sugar prices and of government-sponsored programs of agricultural diversification and limited land settlement. As a result, there has been significant growth in food production (vegetables, fruits, and livestock), mainly for local consumption. Cotton is also grown.

Fishing has always been part of the island's basic economy, and the government has supported the industry with modernization programs.

Industry. Apart from some quarrying of clay, limestone, and sand, the mining industry is limited to oil and natural gas production and to the refining of imported crude oil for local needs. Crude oil production accounts for about one-third of local needs. Manufacturing industry, stimulated by government incentives, is one of the main growth areas of the economy. Tourism is a fast-growing segment of the economy and the chief foreign-exchange earner.

Finance and trade. Barbados' banking system consists of commercial banks (mostly branches of international banks), a central bank, and various development-oriented financial institutions. A small securities exchange, trading in the stock of locally owned companies, has operated since 1987. During the 1980s there was considerable growth in the offshore financial sector.

Exports

The chief exports include electrical components, processed foods, clothing, furniture, and chemicals. Principal imports include food products, machinery, and fuels. Barbados' main trading partners are the United States, the United Kingdom, and Trinidad and Tobago and other members of Caricom (Caribbean Community and Common Market).

Transportation. The island has a network of good roads. Bridgetown has a deepwater harbour, and several international airlines and British West Indian Airways offer regular services to Grantley Adams International Airport near the southern coast.

Administration and social conditions. *Government.* The constitution of 1966 established a governmental structure based on the British parliamentary system. The British monarch is the head of state and is locally represented by a governor-general. A prime minister, a cabinet, an elected House of Assembly, and a nominated Senate are the main governmental institutions. The Democratic Labour Party (founded in 1955) and the Barbados Labour Party (founded in 1938) are the main political parties. The legal voting age is 18. The Supreme Court of Judicature consists of the High Court and Court of Appeal. Magistrates' courts have civil and criminal jurisdiction.

Education. Barbados has a literacy rate of 98 percent, which is attributable to its comprehensive, mainly government-funded primary-school system. The government places high priority on education; it allocates more than 20 percent of its budget to education, and all education in public institutions is free. Facilities for secondary, technical, and vocational education have expanded rapidly since the 1960s; a polytechnic, a community college, and several new secondary schools have been established. Most training at the university level is done at the University of the West Indies, which maintains a campus at Cave Hill in Barbados.

Health and welfare. Social conditions have been upgraded by political changes since World War II and by improvement in the economy. Sustained efforts by government agencies in sanitation, public health, and housing have significantly improved health conditions. The diseases associated with poverty and underdevelopment have been eliminated or controlled. Health care is provided by both public and private agencies. Other areas of social welfare, notably child care, family life, pension plans for the elderly and disabled, public aid, and the status of women, have benefited from government attention. Community centres and playing fields have been established in most parts of the island.

Cultural life. Barbados has a museum and a public library system. There are daily newspapers and local radio and television stations. The country also has dramatic groups, schools of dancing, and art exhibitions. Barbados is internationally known for cricket.

For statistical data on the land and people of Barbados, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

Original inhabitants

Little of the island's prehistory is known, but archaeological investigation indicates that Amerindians probably lived on the island from about AD 500 to 1500. The first contact with Europeans may have occurred in the early 16th century when Spaniards visited Barbados on one of their raids for slave labourers. By the mid-16th century—largely because of the island's small size, remoteness, and depopulation—the Spanish had effectively abandoned their claims to its possession.

British rule. English colonists established a settlement in 1627 without challenge from either Amerindians or Spaniards. The early period of English settlement was marked by the insecurity resulting from infrequent provision of supplies from Europe and the difficulty in establishing a profitable export crop. This was complicated by bitter squabbles over the claims of rival lords proprietors and over the question of allegiance to either king or Parliament, resulting in the entrenchment of representative governmental institutions by the 1660s.

The search for a profitable export crop ended in the 1640s when Dutch assistance enabled the colonists to convert from tobacco and cotton cultivation to sugar production. This decision had momentous social, economic, and political consequences. Sugar needed a larger labour force than was available and larger farm units than had

previously existed. The importation of African slaves was intensified, and the small farms were amalgamated into plantations. The character of the population changed: in the early 1640s there were probably 37,000 whites and 6,000 blacks; by 1684 there were about 20,000 whites and 46,000 blacks; and, when slavery was abolished in 1834, there were 15,000 whites and 88,000 nonwhites. Sugar was a scarce and therefore valuable commodity in European markets, and Barbadian sugar planters, particularly in the 17th century, reaped huge profits out of the early lead that the island established in sugar production. Increasing wealth brought consolidation of political power for a planter elite. Though slaves continually resisted their bondage, the effective authoritarian power of planter-slave-owners ensured that, apart from the 1816 slave rebellion, there was no effective threat to their control.

Sugar dominance

Sugar remained "king" in Barbados even through the 19th-century crises caused by slave emancipation, free trade, and beet sugar competition. This was mainly because a dense population provided cheap labour, and because the white planter–merchant elite's political power ensured that government resources would be used to rescue the industry in any emergency. The workers therefore carried the burden in low wages and minimal social services. This situation encouraged emigration (often frustrated by the elite) and occasional, futile political protests such as the Confederation Riots of 1876.

By the 1930s the social and political pressures from below could no longer be contained. Population increase, the closing of emigration outlets, the economic effects of the worldwide depression, and the spread of socialist ideology and the black nationalist movement of the Jamaican leader Marcus Garvey had created conditions for a labour revolt. By then, middle-class reformers had begun to agitate against the restricted political franchise (the right to vote was limited to males and restricted by income and property qualifications) and the inadequate social services.

Out of the labour disturbances of 1937 emerged a clear challenge to the existing order. The British government's response assisted this successful challenge. The West Indies Royal Commission (Moyne Commission) dispatched in 1938 to report on social and economic conditions in the British West Indies, endorsed some of the political and social reforms that were advocated by the leaders of the new mass organizations, particularly the full legalization of trade unions and the extension of the political franchise. The implementation of these reforms during the 1940s provided the essential base for the institutionalization of mass political organizations, which became the principal means through which the elite's political power was curtailed. In Barbados black political leaders gained ascendancy by 1944, the same year in which women received the vote; universal adult suffrage was adopted in 1951, and full internal self-government was achieved in 1961.

Independence

Independence. Barbados became independent on Nov. 30, 1966, after joining the ill-fated West Indies Federation (1958–62). By then the economy was expanding and diversifying, mainly as a result of the policies pursued by the governments formed after the planter–merchant elite lost power.

Barbados is a member of the Commonwealth and continues to play a leading role in the establishment of regional cooperation. In 1968 Errol Barrow, the prime minister in 1966–76 and 1986–87, helped form the Caribbean Free Trade Area, which became Caricom in 1973. The island has also established close ties with Third World countries. Throughout the 1980s Barbados had one of the most stable political systems in the English-speaking Caribbean.

(C.S.J./W.K.M.)

For later developments in the history of Barbados, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Dominica

Dominica, a Caribbean island, lies between the French islands of Guadeloupe and Marie-Galante to the north and Martinique to the south. It has been a member of the Commonwealth since independence in 1978. It is 29 miles (47 kilometres) long, has a maximum breadth of 16 miles, and is 290 square miles (750 square kilometres) in area. The capital and chief port is Roseau.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief, drainage, and soils.* The island is of volcanic formation, signs of activity including solfataras (volcanic vents) and hot springs. A range of high, forest-clad mountains runs north to south, broken in the centre by a plain drained by the Layou River, which flows to the west; the highest points are Mount Diablotin (4,747 feet [1,447 metres]) and Mount Trois Pitons (4,670 feet [1,424 metres]). In the south, Boiling Lake lies 2,300 feet above sea level; its waters are often forced three feet above normal by the pressure of escaping gases. The soil is rich, and the numerous rivers are all unnavigable.

Climate. Dominica has a pleasant climate, particularly during the cool months from December to March. Summer temperatures reach an average high of 90° F (32° C); winter temperatures are not much lower, the average high being anywhere from 84° to 86° F (29° to 30° C). The dry season is from February to May, and the rainy season is from June to October, the most likely period for hurricanes. Rainfall varies, being especially heavy in the mountainous interior. Average annual coastal rainfall varies from about 60 inches (1,500 millimetres) to 145 inches (3,700 millimetres), but in the mountains average rainfall can reach 250 inches (6,350 millimetres).

Plant and animal life. Dominica is the most heavily forested island of the Lesser Antilles. The forest is the habitat of a considerable variety of birds and animals. Two parrots—the imperial parrot, or sisserou (*Amazona imperialis*) and the smaller red-necked parrot (*Amazona arausiaca*)—are found only in Dominica. There are many hummingbirds, of which the blue-headed (*Cyanophaia bicolor*) is native only to Dominica and the neighbouring island of Martinique. Large frogs, known as crapaud or mountain chicken, are a culinary delicacy.

The people. The population is mainly of African descent, with some Europeans, Syrians, and Caribs. Dominica is the only island with a relatively large and distinctive group of Carib Indians, descendants of the people who inhabited the island before European colonization. Most of the remaining Caribs, a small number of whom are pure-blooded, live in the 3,000-acre (1,214-hectare) Carib Reserve. English is the official language, but a French patois is commonly spoken, and the original Carib language is evidenced in a number of place names. The majority of the population is Roman Catholic, but there also are Methodists, Pentecostals, and Seventh-day Adventists. Dominica experienced out-migration throughout the 1970s, a trend that culminated with a massive exodus after Hurricane David in 1979. This trend, however, reversed in the 1980s.

The economy. Dominica is one of the poorest of the Caribbean nations, its economy dependent upon agriculture, which is intermittently destroyed by hurricanes. Attempts to diversify have had minimal success.

Resources. Pumice, a volcanic rock used chiefly for building purposes, is the most important commercial mineral. There are also deposits of clay and limestone.

Agriculture, forestry, and fisheries. Agriculture remains the most important sector of the economy, in terms of both employment and contribution to the gross national product. The main crops are bananas, citrus fruits, and coconuts. Bananas account for nearly half of Dominica's export earnings. Cocoa, coffee, and vegetables are also produced. The forests have potential for marketable timber. The fishing industry was devastated by Hurricane David, when nearly all of the island's fishing boats were destroyed. Recovery has been slow.

Industry. Most of the main products and exports are derived from the agricultural industry; they include copra, coconut oil, soap, bay oil, and fruit juices. Wood products, including furniture, are produced from local timber. Portsmouth is the main boatbuilding centre. Imports include food, mineral fuels, and manufactured goods.

Tourism

Tourism has been slow to develop because of poor transport and the lack of hotel facilities and good beaches. The

island has sought to develop preserves of its unique flora and fauna to attract tourists.

Transportation. High rainfall and rugged terrain have impeded road building in Dominica. The first road across the island was not completed until 1956, and it was not until 1984 that a major road rehabilitation project was launched to greatly improve accessibility. The main airport is at Melville Hall, 36 miles from Roseau. A second airport at Canefield, closer to the capital, was opened in 1982. Larger vessels use the deepwater port at Woodbridge Bay near Roseau, but Portsmouth remains the major banana-shipping port.

Administration and social conditions. *Government.* Dominica's government is a parliamentary system, with the parliament consisting of the chief executive and the House of Assembly. Most of the House members are elected, but some, usually called senators, may be either elected or appointed. The chief executive is the president, who has the responsibility of appointing the prime minister, an elected member of the parliament with the support of the majority of its members. Terms of office are for five years, and there is universal adult suffrage.

Education. Primary education is compulsory and free in government-run schools. There are many secondary schools, and a university centre is operated by the University of the West Indies.

Health and welfare. There are several major hospitals. Local medical needs are handled by health centres throughout the island. Intestinal diseases, diabetes, anemia, tuberculosis, and sexually transmitted diseases constitute the major health problems of Dominica.

Cultural life. Carib material culture remains evident in the production and use of dugout canoes and intricate woven baskets. The Department of Culture has encouraged revival of slavery-era traditions, which had almost died out, including Afro-French dances, drama, music, and costumes. The Botanical Garden, although it has lost its collection of exotic plants, provides an idyllic setting for the island's main sport of cricket.

For statistical data on the land and people of Dominica, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

Before colonization the island was a stronghold of the Carib Indians who had migrated from South America, driving out the earlier Arawak Indians. It was named by Christopher Columbus, who sighted it on Nov. 3, 1493, a Sunday (Latin: *dies dominica,* "the Lord's day").

The first colonists

The French and British colonial period. The first colonists (1632) were French, but, with the Treaty of Aix-la-Chapelle (1748), Great Britain and France agreed to treat the island as neutral ground and leave it to the Caribs. From this time until 1805, Dominica went back and forth between France and Britain. French planters continued to settle in Dominica until 1759, when the British captured the island. It was formally ceded to Britain in 1763. In 1778, French forces from Martinique captured Dominica. The British recaptured the island in 1783. The French, coming this time from Guadeloupe, again failed to capture the island in 1795. The final French assault on the island was in 1805, and although they burned the capital, Roseau, they were forced to withdraw.

At first administered as part of the Leeward Islands, in 1771 Dominica was made a separate colony. It was rejoined administratively to the Leewards in 1883 and remained thus until 1940, when it was transferred to the Windwards as a separate colony. In 1958 Dominica joined the West Indies Federation. After the federation was dissolved in 1962, discussions for alternative forms of federation took place. These were settled by the West Indies Act of 1967, which gave Dominica the status of association with the United Kingdom. Under the 1967 constitution the island became fully self-governing in internal affairs.

Independence. On Nov. 3, 1978, Dominica achieved full independence, with Patrick Roland John as its first prime minister. John's government was implicated in a rumoured invasion of Barbados that was to have been launched from Dominica. In the ensuing Cabinet crisis Oliver Seraphine emerged as the new prime minister (May 1979).

Hurricane devastation

Hurricane David severely damaged the island in August 1979, virtually wiping out the nation's agricultural economy. The hurricane carried away most of the island's topsoil, and it was estimated that it would take 20 years to rebuild what had been destroyed. The economy was set back by Hurricane Allen a year later and in 1989 by Hurricane Hugo.

The winner of the 1980 elections, Eugenia Charles, became the Caribbean's first female prime minister. She had initially formed her party, the Dominica Freedom Party, to oppose legislation limiting freedom of the press. More conservative in her approach than either of her predecessors, she moved Dominica toward closer ties with Barbados. Her government faced several coup attempts in 1981, but these were perhaps of less significance than the plight the country faced in attempting to recuperate from the two hurricanes. Under Charles's administration, however, Dominica made marked advances toward recovery, with considerable decreases in unemployment and inflation. Her party was returned to power in 1985 and became more firmly entrenched in the 1986 elections. Dominica joined with other eastern Caribbean states and the United States in the 1983 invasion of Grenada.

(D.L.N./J.D.Mo.)

For later developments in the history of Dominica, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Grenada

The island of Grenada, also known as the Isle of Spice, is the southernmost of the Lesser Antilles in the eastern Caribbean Sea about 100 miles (160 kilometres) north of the coast of Venezuela. In 1974 it attained independence within the Commonwealth and membership in the United Nations, the first of the six West Indies Associated States to do so.

Oval in shape, the island is approximately 21 miles (34 kilometres) long and 12 miles wide, with an area of 120 square miles (311 square kilometres). The southern Grenadines—the largest of which is Carriacou, about 20 miles north-northwest, with an area of 13 square miles—are a dependency.

The capital

The capital, St. George's, on the southwest coast, is also the main port, having a fine natural harbour as well as picturesque pastel-coloured houses that rise up the hillsides from the waterfront. The waterfront itself is known as the Carenage because island schooners were once careened (beached for cleaning or repair) there. St. George's is the yachting and charter-boat centre of the eastern Caribbean.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Grenada is volcanic in origin, with a ridge of mountains running north and south—the steeper slopes to the west and a more gradual incline to the east and southeast. The highest point is Mount St. Catherine (2,757 feet [840 metres]) in the northern part of the interior. The landscape is scenic, with fairly deep, steep-sided valleys and about 10,000 acres (4,000 hectares) of forest.

Drainage and soils. Several short, swiftly flowing streams supply all towns and most villages with piped clean water. A further source of water is Grand Etang, a lake covering 36 acres in the crater of an extinct volcano at an elevation of 1,740 feet. The fertile soils are chiefly volcanic, with some limestone in the north.

Climate. The island has equable temperatures varying with altitude and averaging 82° F (28° C). Rainfall is adequate, except in the Point Salines area in the southwest; it varies from an average of 60 inches (1,500 millimetres) in coastal districts to more than 150 inches in the mountainous regions. The rainy season lasts from June to December. November is the wettest month, but showers occur frequently during the other months. Grenada lies south of the usual track of hurricanes, but when they do occur, as in 1955, 1979, and 1980, they often cause extensive damage.

Plant and animal life. The island is verdant, with a

year-round growing season and a wide variety of tropical fruits, flowering shrubs, and ferns. There are also forests of teak, mahogany, saman (known as the rain tree), and blue mahoe (a strong-fibred tree) in the interior.

Forests

The animal life is varied and includes such wild animals as the mona monkey (a small, long-tailed, West African species that was introduced by slaves), the manicou (a species of opossum), the agouti (a rabbit-sized rodent, which is brown or grizzled in colour), the iguana, the mongoose, and a variety of turtles and land crabs.

The people. Most of the population is black, having descended from African slaves, and there is a large minority of mulattoes and other mixtures. There are also small minorities of East Indians, descendants of indentured labourers brought to replace the freed slaves; descendants of the old French and British settlers; and more recent immigrants from North America and Europe. Although English is the accepted language, a form of patois is still spoken by older people in the villages. A majority of the population is Roman Catholic; other Christian denominations include Anglicans (more than a fifth of the population), Methodists, and Seventh-day Adventists. Although densely populated, Grenada's population grew slowly during the 20th century.

The economy. Agriculture and tourism are the most important sectors of the economy, although fishing and agriculturally based industries are becoming more significant. Grenada relies on financial support from the United Kingdom and other sources to bolster the economy.

Agriculture, forestry, and fisheries. To a greater extent than in most West Indian islands, Grenada's arable land is divided into small holdings on which peasant proprietors cultivate diversified crops. Because of these small holdings and the generally hilly terrain, mechanical tilling is rare. The major agricultural export crops—cocoa, bananas, nutmeg, and mace—in the past were controlled by cooperative associations, but these associations have begun to come under greater government control. Banana exports depend upon preferential terms given by the United Kingdom and are affected by the policies of the European Economic Community. Exports of mace and lime juice provide substantial earnings. Copra and, increasingly, other products processed from the coconut are also exported, and a wide variety of tropical fruits—mangoes, passion fruit, guavas, tamarind, and citrus fruits—are grown. The government has encouraged increased production of staple vegetables, such as peas, tomatoes, sweet potatoes, pumpkins, and corn (maize).

The island's forests yield mostly teak and mahogany and the government has worked to upgrade fishing.

Tourism

Industry and trade. Tourism, a major factor in the island's economy, has been encouraged by the government. Air transport facilities have been improved, and the harbour is visited by numerous cruise ships. Other sources of employment are such secondary industries as clothing manufacture, sugar milling, brewing, rum distilling (a strong white rum being made for local consumption), food canning, copra processing, cigarette manufacturing, and soapmaking. There is a cotton ginnery on Carriacou.

The United Kingdom is Grenada's principal trading partner. Exports go largely to Trinidad and Tobago, the United Kingdom, West Germany, and The Netherlands; most imports come from the United Kingdom, the United States, and Trinidad and Tobago.

Transportation. Bus service is available between the larger towns and villages. An international airport at Point Salines was inaugurated in 1984. Pearls Airport—providing service to nearby islands with connecting flights to Venezuela—is located on the northeastern coast. An airport on Carriacou also provides flights to nearby islands.

The harbour at St. George's has berths for oceangoing vessels, as well as a yacht basin and service facilities. Several shipping lines maintain regular passenger and cargo services to North America, the United Kingdom, Europe, and neighbouring West Indian islands.

Administration and social conditions. *Government.* Grenada is governed as a constitutional monarchy, with the British monarch represented by a governor-general as the nominal head of state. Executive authority is vested in a prime minister, who is the head of the majority party in the elected House of Representatives, the lower house of the two-chamber legislature. The Senate is appointed by the governor-general on the advice of the prime minister and the opposition leader.

The legislature

Education. School attendance is not compulsory, although primary and secondary education is free. Grenada has vocational and technical schools as well as the St. George's University School of Medicine and a branch of the University of the West Indies.

Health and welfare. Grenada has several main health centres, as well as the district medical stations. Medical and dental treatment in government hospitals and clinics is free. The government has launched a program to eradicate malaria and mongoose-spread rabies.

Culture. The Grenada National Museum in St. George's is dedicated to archaeology and history and houses the Grenada Historical Society. Two Grenadian artists, Elinus Cato and Canute Caliste, have achieved international recognition for their primitive-style paintings. Several weekly newspapers are published, and islandwide radio and television broadcasting is available.

For statistical data on the land and people of Grenada, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

Grenada was sighted by Christopher Columbus on Aug. 15, 1498, when he sailed past the island without landing and gave it the name of Concepción. The origin of the name Grenada remains obscure. After its discovery, Grenada was dominated for the next 150 years by the warlike Carib Indians, who had earlier killed off the more peaceful Arawak. In 1609 British merchants attempted to form a settlement, but the Caribs forced them to leave.

French settlement. The French governor of Martinique, du Parquet, purchased Grenada from a French company in 1650 and established a settlement at St. George's. Grenada remained French until 1762, when it capitulated to the British. The island was formally ceded to Britain in 1763 by the Treaty of Paris. In 1779 it was recaptured by the French, but it was restored to Britain in 1783.

British rule. In the late 18th century the British imported large numbers of slaves from Africa to work the sugar plantations. During 1795 and 1796, when French policy favoured the abolition of slavery, a rebellion against British rule occurred, led by a French planter and supported by the French in Martinique. The rebels massacred a number of the British, including the lieutenant governor, but the uprising was quelled. The emancipation of the slaves finally took effect in 1833.

Grenada was headquarters of the British Windward Islands government from 1885 until 1958, when Grenada joined the West Indies Federation. The federation ended in 1962, after which Grenada attempted to federate with the remaining territories in the Windward Islands, as well as with Barbados and the Leeward Islands. On March 3, 1967, however, the island became a self-governing state in association with the United Kingdom.

Independence. In the general election of August 1967, the Grenada United Labour Party (GULP) defeated the Grenada National Party (GNP) and took office under the premiership of Eric M. Gairy, a trade unionist. Grenada became an independent nation on Feb. 7, 1974. The transition was marked by violence, strikes, and controversy centring upon Gairy, who was named prime minister. Opposition to Gairy's rule continued to mount, and a coalition called the New Jewel Movement (NJM), along with other opposition parties, succeeded in reducing GULP's majority in Parliament in the 1976 election. On March 13, 1979, while Gairy was out of the country, the NJM staged a bloodless coup, proclaimed a People's Revolutionary Government (PRG), and named their leader, Maurice Bishop, as prime minister. The new government faced opposition from Western nations because of its socialist principles, but it embarked on a program to rebuild the economy, which had been left in disarray by Gairy. The PRG administration was ended in October 1983 by a military coup, during which Bishop was killed.

Accession of GULP

Less than a week later, on October 25, a U.S.-led invasion of the island overthrew the coup leaders and returned power to the governor-general, Sir Paul Scoon. In December Scoon appointed Nicholas Braithwaite, a former Commonwealth official, to head a governing council until an election could be held, and constitutional government was restored. A peacekeeping force remained until 1985. The election, held in December 1984, was won by the New National Party headed by Herbert A. Blaize, who had led the government in the 1960s. The new government sought to revive tourism, but Grenada's continuing economic problems throughout the late 1980s contributed to the government's dwindling popularity. Following an election in March 1990, Braithwaite, whose National Democratic Congress fell one seat shy of a parliamentary majority, was appointed prime minister by Scoon. (E.V.B.B./Ed.)

For later developments in the history of Grenada, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Guadeloupe

Guadeloupe, a French overseas *département,* is a group of islands in the Lesser Antilles chain in the eastern Caribbean Sea. The nearest neighbours of the principal islands are the British dependency of Montserrat to the northwest and the republic of Dominica to the south. The island of Martinique, also a French overseas *département,* lies about 74 miles to the south. The main territory of Guadeloupe consists of the twin islands of Basse-Terre to the west and Grande-Terre to the east, the two being separated by a narrow channel, the Salée River; other islands in the group are Marie-Galante to the southeast, La Désirade to the east, and Saintes Islands (Terre d'en Haut and Terre d'en Bas) to the south. Two more island dependencies—Saint-Barthélemy and Saint-Martin (the southern third of which is administered by The Netherlands as Sint Maarten)—are situated about 150 miles to the northwest, lying to the northwest of the outer arc of the Lesser Antilles.

The total area of Guadeloupe is 687 square miles (1,780 square kilometres). Basse-Terre, on the island of the same name, is the seat of government. The largest urban area, however, centres around Pointe-à-Pitre on Grande-Terre, the chief port and economic hub of Guadeloupe.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief and drainage.* Basse-Terre, which has an area of 364 square miles, has a chain of mountains running north to south and culminating in Soufrière, a volcano 4,813 feet (1,467 metres) high; it erupted in 1797, 1837, and 1976 and is now a source of hot springs and sulfur springs. Other summits of note are the 4,442-foot Mount Sans Toucher and the 4,143-foot Mount de la Grande Découverte. The mountain chain forms a watershed from which rivers run down to the sea. The principal river on the island is the Goyaves; other rivers are the Grande Plaine, the Petite Plaine, the Moustique, the Lézarde, and the Rose. Basse-Terre has a beautiful coastline, indented with bays and fringed with picturesque beaches.

Principal rivers

Grande-Terre has an area of 220 square miles and is generally low-lying; it has only a few bluffs higher than 490 feet. Saint-Martin and Saint-Barthélemy are rugged and rise to an altitude of 1,391 feet and 921 feet, respectively.

Climate. The tropical climate is tempered by the northeast trade winds. The temperature on the coast varies between 77° and 82° F (25° and 28° C), with extremes of 68° and 93° F (20° and 34° C). In the mountains above 1,900 feet it may drop to 61° F (16° C), and at the summit of Soufrière to 39° F (4° C). There are two distinct seasons—the "Creole Lent," or dry season, from December to April, and the winter, or rainy season, from July to September–October. Precipitation varies with altitude and orientation. Grande-Terre receives about 39 inches (990 millimetres) of rain a year, while the mountainous parts of Basse-Terre receive more than 100 inches. Hurricanes occur occasionally, usually coming from the south.

Seasons

Plant and animal life. The heat, rainfall, and fertility of the volcanic soils produce a luxuriant vegetation diversified according to altitude. About two-fifths of the islands' area is covered by forests, most of this on Basse-Terre. Extensive mangrove swamps cover the banks of the Salée River. Dense forest grows in the mountainous regions of Basse-Terre, beginning almost at sea level on the windward slopes and at altitudes of about 750 to 3,000 feet or more on the leeward side. There chestnut trees and bracken are found, as well as such hardwoods as mahogany and ironwood. On the highest peaks some flooded basins produce a vegetation of grasses and sedges. Grande-Terre, cleared of most of its original forests, has only a few patches of woodland. The smaller islands, such as La Désirade and Saint-Martin, have a different type of vegetation, consisting primarily of dry forest with groves of latania (a kind of fan palm) and cactus.

Animal life has been modified since colonization. Raccoons are sought for their fur. The agouti (a short-haired, short-eared, rabbitlike rodent) still inhabits the heights of Capesterre, southeast of Basse-Terre. In some regions, wild ducks, waterfowl, and teal are found.

The warmth of the water around the islands is responsible for a rich variety of marine life, including lobster, crab, octopus, tarpon, snook (a basslike kind of fish), hogfish, snapper, parrot fish, and many species of ray fish.

The people. The population is composed principally of Creoles (*i.e.,* persons born in the islands), most of whom are mulatto, but on Saintes Islands the inhabitants are mainly white. The largest minorities are the black and French-Amerindian groups. The white population greatly declined during the period of the French Revolution. On the smaller islands, whites are mostly descended from 17th-century Norman and Breton settlers. While French is the official language, a local creole dialect is also widely spoken. The majority of people are Roman Catholic.

Guadeloupe's population has a low rate of natural increase compared to other West Indian islands. Its birth and death rates are lower than the Caribbean average but about the same as those of its French counterpart, Martinique. The vast majority of the population resides on the two largest islands; Marie-Galante is the next most populous island, followed by Saint-Martin, Saint-Barthélemy, Saintes Islands, and La Désirade. The islanders' standard of living is among the highest in the eastern Caribbean.

Population distribution

The economy. The economy is marked by a slow-growing agricultural sector, an embryonic industrial sector, a growing tourism sector, and a highly developed public service sector. In effect, the economy is sustained primarily by the salaries of officials and by French credits, which consist of aid in the form of allocations and grants. Tourism is the main source of foreign exchange.

Agriculture and industry. Bananas and sugarcane form the principal cash crops; coffee, vanilla, and cacao are also grown. The banana plantations suffered from a series of hurricanes in the 1960s, but the plantations were replanted with more productive types of trees. The cultivation of fresh vegetables, coconuts, pineapples, limes, mangoes, flowers, and coffee is increasingly important. Eggplants and flowers are chiefly grown for export. Most of the small fish catch is exported. An industrial zone and a free port have been developed at Jarry, near Pointe-à-Pitre. Industry is involved mostly in the processing of agricultural products, especially sugar refining and rum making.

Trade. There is a severe deficit in the balance of external trade, most of which is with France and the Franc Zone. Most imports are consumer goods. Most of the banana crop and raw sugar are exported to France. Rum, coffee, cocoa, and vanilla also are exported.

Transportation. Guadeloupe maintains regular air and sea links with France and with the North American continent. The port of Pointe-à-Pitre is equipped to handle cargoes of minerals, sugar, and cereals. The port of Basse-Terre specializes in the banana export trade. Le Raizet, north of Pointe-à-Pitre, is an international airport used by French, U.S., British, and Dutch airlines. There is a secondary airport on Marie-Galante. On the island of Saint-Martin, the town of Marigot, the capital of the French portion of the island, is an important port; the Juliana International Airport, west of Philipsburg in the Dutch sector, serves both parts of the island. Local steamers connect Basse-Terre and Grande-Terre with the other island

dependencies. The road system on the main islands is kept in excellent condition. Except for some privately owned plantation lines, there are no railways in Guadeloupe.

Administration and social conditions. *Government.* The *département* is under the executive authority of a commissioner appointed by the French government; there is an elected legislative council. Guadeloupe sends representatives to both the French National Assembly and the French Senate. Since 1974 Guadeloupe has had the status of a full *région* of France. The territory of Guadeloupe is divided into three *arrondissements,* which are in turn divided into 34 *communes,* each administered by an elected municipal council.

Executive authority

Justice. The judicial system is French. There are a court of appeal at Basse-Terre, two higher courts (*grande instance*), and four lower courts (*tribunaux d'instance*). Justices of the peace are established in each of the cantons.

Education. French is the medium of instruction. In addition to primary schools, there are lycées (secondary schools) as well as a teacher-training college. A school of humanities, a law and economics school, a school of medicine, and a school of science at Pointe-à-Pitre are part of the University of the Antilles and Guyana.

Health and welfare. The same social legislation is in effect as in metropolitan France. There is a general hospital at Pointe-à-Pitre, as well as a Pasteur Institute and a number of other hospitals and clinics.

Cultural life. Cultural affairs are developed through the Ministry of Culture. Folk culture is of considerable significance, and colourful native costumes, including the unique *madras et foulard,* may still be seen on holidays. Celebrations, particularly the annual carnival, feature Creole music and folk dances, such as the beguine. A number of museums are located in the major cities. Several newspapers are printed on the islands, and radio and television are broadcast daily.

For statistical data on the land and people of Guadeloupe, see the *Britannica World Data* section of the BRITANNICA WORLD DATA ANNUAL.

HISTORY

Visited on Nov. 4, 1493, by Christopher Columbus, the two main islands, then together known as Karukera (Island of Beautiful Waters), were peopled by Caribs, who had displaced the original Arawak inhabitants. The territory was consecrated to Our Lady of Guadeloupe of Estremadura in Spain, from whom it takes its name.

French rule. Preliminary attempts by the Spanish to establish themselves were repulsed by the Caribs in 1515, 1520, and 1523. In 1626 the Spanish, who had established themselves on the coast, were driven away by Pierre Belain d'Esnambuc, a Frenchman who established a trading company. In 1635 two Frenchmen, Léonard de L'Olive and Jean Duplessis d'Ossonville, landed and established a colony. Until 1640 the colonists fought against the Carib Indians, but thereafter the colony prospered. Four chartered companies were ruined in successive attempts to colonize Guadeloupe, and in 1674 it passed to the French crown, becoming a dependency of Martinique, which it remained until 1775. Guadeloupe benefited from the influence of Jean-Baptiste Labat (1663 to 1738), a strong personality who was the effective founder of the Basse-Terre colony and who in 1703 armed the African slaves (who had already been brought to the island) in order that they might fight against the English; he also established the first sugar refineries, thereby laying the foundations for the era of prosperity that followed. In 1759 Guadeloupe was occupied by the British for four years but was restored to France in 1763. In 1794 it was again occupied by British troops, allied with French royalists, but was recaptured by Victor Hugues, an official of the French revolutionary government, who proclaimed the abolition of slavery and had several hundred white planters massacred. When slavery was reestablished by Napoleon's government in 1802, a revolt of the slaves occurred and culminated in the heroic act of the antislavery forces, who blew themselves up at Matouba when threatened by French forces under the command of General Antoine Richepanse; Richepanse himself had been sent by Napoleon to pacify Guadeloupe, but he died of yellow fever in the same year. The British occupied Guadeloupe in 1810, but, after some changes in status, it was restored to France in 1816.

Jean-Baptiste Labat

The abolition of slavery in 1848 was the most significant development of the territory's 19th-century history. Universal suffrage was abolished during the reign of Napoleon III of France, but in 1870 colonial representation in the French Parliament was restored. In 1940 Guadeloupe gave its allegiance to the Vichy government of Nazi-occupied France, but in 1943 it adhered to General Charles de Gaulle's Free French forces. In 1946 it was given the status of a French *département,* and in 1974 it became a *région* of France.

Political and economic changes since World War II. Guadeloupe has had several independence movements since the end of World War II, but the charismatic appeals of de Gaulle, who visited the island in 1956, 1960, and 1964, managed to sidestep the separatists and convince the majority to stay within the French union. More local control has been granted the island since the 1960s, but, as progress on the autonomy talks slowed to a standstill in the 1970s, the separatist groups became increasingly violent. Several bombings were committed in Paris by Guadeloupe independence groups. Despite further acts of violence in the 1980s by these groups and their gains in local government, the French government reiterated its determination to maintain *département* status for Guadeloupe. The lack of economic improvement, however, provided some stimulus to independence groups. (R.Co./Ed.)

For later developments in the history of Guadeloupe, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Martinique

The island of Martinique is an overseas *département* of France situated in the eastern Caribbean. It is included in the Lesser Antilles island chain. Its nearest neighbours are the island republics of Dominica, 22 miles to the northwest, and Saint Lucia, 16 miles to the south. The main islands of Guadeloupe, also a French overseas *département,* lie about 74 miles to the north.

Martinique has an area of 421 square miles (1,091 square kilometres) and measures about 50 miles in length and about 22 miles at its widest extent. The second smallest (after Saint-Pierre and Miquelon) of all the French overseas *départements,* Martinique's population density is one of the highest in the Antilles. The administrative capital and chief town is Fort-de-France.

The name Martinique is probably a corruption of the Indian name Madiana ("Island of Flowers") or Madinina ("Fertile Island with Luxuriant Vegetation"), as reputedly told to Christopher Columbus by the Caribs in 1502. Empress Joséphine, consort of Napoleon I, was born on the island in 1763; she was the daughter of a Martinique planter named Joseph Tascher de La Pagerie.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief and drainage.* The mountainous relief of Martinique represents the outermost edge of what remained of the original geologic formation after the subsidence of the trench that became the Caribbean Sea. The relief of the island takes the form of three principal massifs (mountainous masses). These are a still active volcano, Mount Pelée, 4,583 feet (1,397 metres) high, to the north; the Carbet Mountains, of which Lacroix Peak reaches 3,923 feet, in the centre; and Mount Vauclin, 1,654 feet high, in the south.

The three principal massifs

The tortuous relief of the island has led to a complex drainage pattern, characterized by short watercourses. In the south, the Salée and the Pilote rivers flow down from the scarps of Mount Vauclin. In the centre the rivers flow outward from the Carbet Mountains in a starlike pattern; they include the Lorrain, Galion, Capot, and Lézarde rivers. In the north the Grande River, the Céron River, the Roxelane River, the Pères River, and the Sèche River are little more than irregular torrents.

The northern coastline of Martinique is characterized by steep cliffs; farther south, however, the cliffs become

lower, with two large bays—Fort-de-France and Marin—being located on the western coast. Coral reefs, headlands, and coves line the east coast.

Climate. The climate is remarkably constant, the average temperature being about 79° F (26° C), with average minimums of 68° to 72° F (20° to 22° C), average maximums of 86° to 90° F (30° to 32° C), and temperature extremes of 59° F (15° C) and 93° F (34° C). The northeast trade winds, which blow almost 300 days a year, temper the heat, but winds from the south are hot and humid and sometimes bring hurricanes.

The year consists of two distinct seasons—a relatively dry season, which lasts from December to June, and the rainy winter season from July to December. Rainfall is abundant, especially in July and September, but is very irregularly distributed; it varies from about 40 inches (1,000 millimetres) to almost 400 inches a year, depending upon elevation and the orientation of the relief.

Plant and animal life. The climate, together with the fertile volcanic soil, produces a luxuriant vegetation, which is divided, according to altitude, into four zones: the maritime zone, the lowlands, the former forest zone, and the upper mountain slopes. The maritime zone includes an enormous mangrove swamp, half of which is located in the bay of Fort-de-France. The beaches are invaded by morning glory, tropical twining herb, and sea grape. The lowland vegetation zone, extending from the coast to a height of about 1,500 feet, has ferns and orchids, as well as various trees, including mahogany, white gum, and other species. Above 1,500 feet is the former virgin forest zone, where large trees and bracken are still to be found. As the altitude increases the trees grow smaller. A transitional zone is characterized by peat moss. Above 3,000 feet the upper slopes are almost bare, except for some stunted forest. Forests cover about one-fourth of the total land area.

The four vegetation zones

There are relatively few kinds of animals on the island. The mongoose was introduced in the 19th century in the hope of eliminating the deadly rat-tailed viper, but without doing so. Also found are the manicon (a kind of opossum), wild rabbit, wild pigeon, turtledove, and ortolan (a small bird about six inches long, often netted and fattened as a table delicacy).

The people. In 1658 French settlers on the island numbered about 5,000. The Carib element gradually disappeared, partly as a result of conflicts and partly as a result of assimilation. The importation of slaves from Africa added a further ethnic component. Today the racial composition of the island is extremely mixed, but the mulatto element is the largest. The white Creole (locally born) element, however, controls an important part of the island's economy. A creole dialect, similar to that spoken in Haiti, is commonly heard, but French is the official language. The majority of the population is Roman Catholic.

The population of Martinique increased rapidly until the late 1970s, when, plagued by unemployment and other economic maladies, the residents of the island began to emigrate in large numbers to France and in smaller numbers to French Guiana. As a result of the high growth rate in the 1950s and '60s, almost two-fifths of the population is under age 20. About a third of the total population lives in the capital city of Fort-de-France.

The economy. Martinique has a typically Caribbean economy, depending heavily on a few agricultural products and tourism and relying on outside sources, principally France, for aid. A huge trade deficit and a high rate of unemployment are major impediments to economic progress. Nevertheless, the island enjoys one of the higher standards of living in the Caribbean.

Agriculture. The principal agricultural products are sugarcane and bananas, grown chiefly for export. Fresh and canned pineapples, cut flowers, avocados, eggplants, and citrus fruits are other exports. Grown for the domestic market are yams, cassava, sweet potatoes, and breadfruits. Fishing of crab, lobster, clams, cod, and crayfish is mainly for domestic use. The destruction of most banana plantations by Hurricane Allen in 1980 was a major setback.

Cash crops

Industry. Economic planning has laid emphasis on land reform (providing certain planters with lands that have been insufficiently reclaimed from swamp), the diversification of agricultural crops, and industrialization. Local industry includes cement, fertilizer, and polyethelyne plants and an oil refinery at Fort-de-France. Other industries include rum distilling, fish and fruit canning, sugar refining, the processing of cattle feed, soft drinks, and food, and the manufacture of pottery, wooden furniture, and chemicals. The growth of light manufacturing is being encouraged. One of the most popular tourist areas in the Caribbean, Martinique has a flourishing cruise ship business that brings tourists mainly from France, Canada, and the United States.

Trade. Martinique's economy is heavily dependent on trade with France, which provides about two-thirds of the island's imports and receives more than three-fifths of its exports. The value of imports far surpasses the value of exports, resulting in large trade deficits. Exports include refined petroleum products in addition to agricultural products; chief imports are foodstuffs, consumer goods, petroleum products, and electrical machinery.

Transportation. Martinique maintains regular air and sea links with France and North America. The main port is Fort-de-France. There is an international airport at Lamentin, to the east of Fort-de-France. The road network links Fort-de-France by an expressway with coastal towns. There are local bus services, and small coastal steamers connect various points around the island.

Administration and social conditions. *Government.* As an overseas *département,* Martinique is divided into three *arrondissements,* comprising 34 *communes,* each of which is administered by an elected municipal council. Executive authority is represented by a commissioner and other officials, and there is an elected legislative council. Martinique is represented in the French National Assembly, in the French Senate, and on the French Economic and Social Council. Since 1974 Martinique has also had the status of a full *région* of France.

Justice. The French system of justice is in force. The Court of Appeal at Fort-de-France also has jurisdiction over Guadeloupe and French Guiana. There are two higher courts (*grande instance*), two lower courts (*tribunaux d'instance*), one administrative court, and a commercial court.

Education. Free and compulsory education is provided for children between six and 16 years of age. There are primary, secondary, and vocational schools. The enrollment of children of school age is exceptionally high. Higher education is usually pursued in metropolitan France, for which a number of scholarships are available. Institutes of law and economics, letters and human studies, and Creole studies constitute part of the University of the Antilles and Guiana.

Higher education

Health and welfare. There are several general and maternity hospitals, as well as some dispensaries. Martinique receives the same social benefits as mainland France.

Cultural life. The Fort-de-France carnival, featuring a parade with masks, is an annual event. Voodoo ceremonies are far less important on the island than they are in Haiti. Cockfighting is a popular sport.

For statistical data on the land and people of Martinique, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

The Caribs inhabited the island at the time Christopher Columbus sighted it in 1493. It was not until 1502, on his fourth voyage, that he visited the island, leaving there some pigs and goats. Neglected by the Spaniards, who sought more material rewards than those the island offered, Martinique was occupied in 1635 by a Frenchman, Pierre Belain d'Esnambuc, who established 80 settlers at Fort-Saint-Pierre at the mouth of the Roxelane River. A year later d'Esnambuc, who had fallen sick, entrusted Martinique to his nephew, Jacques-Dyel du Parquet, who bought the island from the Compagnie des Isles d'Amérique and developed it into a remarkably prosperous colony. In 1654 a group of 250 Dutch Jews, chased from Brazil by the Portuguese, introduced sugarcane. In 1660 cacao trees were planted in place of cotton.

French rule. After the death of du Parquet, his widow governed the island in the name of her children but

disagreed with the settlers; and in 1658 the French king, Louis XIV, resumed sovereignty over the island, paying an indemnity to du Parquet's children. In 1664 the island was placed under the authority of the Compagnie des Indes Occidentales; in 1674 it was made part of the French crown domain, being administered according to the Pacte Colonial, a body of principles summarized in the statement, "The mother country founds and maintains the colonies; the colonies enrich the mother country." Supplies and slaves were brought out to the French Antilles by the Compagnie du Sénégal, founded in 1664; the slave ships called at Martinique before proceeding to Guadeloupe, permitting the colony first choice of the slaves. In 1723 Arabian coffee was introduced, thus further contributing to the island's prosperity. In 1787 Louis XVI granted Martinique the right to establish a colonial assembly.

Attacks by Dutch and British

At various times Martinique was subjected to attack by various foreign fleets. An attack by the Dutch was repulsed in 1674; further attacks by the English were repelled in 1693 and in 1759. In 1762, however, the English captured the island, only to return it to France under the terms of the Treaty of Paris in 1763. The English recaptured it in 1794 and occupied it until 1802; captured once more by the English in 1809, it was definitively restored to France in 1814. Slave uprisings occurred in 1789, 1815, and 1822. The abolition of slavery in 1848 created a labour problem, as a result of which labourers from India and China were introduced. Universal suffrage was proclaimed in 1848 but was abolished once more under Napoleon III; after 1870 the Third Republic of France restored representation for the island in the French Parliament.

In 1902 the volcanic eruption of Mount Pelée destroyed the town of Saint-Pierre, killing about 30,000 people. During World War II Martinique adhered to the Vichy government of Nazi-occupied France for three years before rallying to the Free French cause in 1943. In 1946, Martinique was granted the status of a French *département*, and in 1974 it was made a *région*.

Developments since World War II. Martinique, more vociferous in its demands for independence than Guadeloupe, has had its postwar politics influenced by Aimé Césaire, the Martinique writer who was one of the founders of the Negritude movement. Césaire, first elected as a deputy in 1945, had originally been a member of the Communist Party, but by 1956 he had resigned and formed his own party, the Progressive Party of Martinique. In 1957 Césaire's party won the Martinique elections by a huge margin, and it seemed that independence would be achieved.

Martinique's economy was depressed, however, and massive unemployment worked against the independence movement. Emigration to France and French foreign aid had always been a palliative for Martinique's economic problems, and demands for independence resulted only in Martinique's being given greater autonomy. Even the visits of Charles de Gaulle (1956, 1960, and 1964) could not smooth over the political unrest, however, and by the late 1970s the French government, in an apparent about-face, decided to help Martinique become economically self-sufficient in preparation for independence. Economic problems were exacerbated by the widespread destruction from hurricanes in 1979 and 1980. Liberation groups were responsible in the 1980s for several bombings in Paris and the French Caribbean islands. Despite such actions, there was little official movement toward independence.

(R.Co./Ed.)

For later developments in the history of Martinique, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Montserrat

Montserrat, an island of the Lesser Antilles chain in the Caribbean Sea, is a crown colony of the United Kingdom. The island has an area of 40 square miles (102 square kilometres). The capital and only port of entry is Plymouth.

PHYSICAL AND HUMAN GEOGRAPHY

The land. Montserrat lies 27 miles (43 kilometres) southwest of Antigua and is 11 miles long and seven miles wide. It consists of a serrated range of volcanic peaks rising in three main hill masses, the summits of which are forested. There are seven volcanically active peaks in the mountainous terrain. Chances Peak, in the southern hills above the still-active Soufrière Hills, is the highest point, at 3,000 feet (914 metres). The coastline is rugged. The average temperature on the island is 81° F (27° C) in summer and 76° F (24° C) in winter. Rainfall averages 57 inches (1,448 millimetres). Hurricanes are rare, although Hurriance Hugo in 1989 caused widespread devastation.

The people. The population is largely of black African ancestry. The official language is English, and the main religious denominations are Anglican, Methodist, and Roman Catholic, although Seventh-day Adventists and Pentecostals are of increasing importance. Montserrat's population has not changed substantially during the 20th century, largely as a result of out-migration and a low birth rate. The elderly population has been growing since 1970 as increasing numbers of people from North America and Canada have retired there.

Religion

The economy. Although agriculture has declined in importance, it continues to be significant as a supplier of products to related industries. The most important agricultural products are vegetables and cotton. The production of Sea Island cotton was revived in the 1970s to serve as the basis of garment and craft industries. Fruits, livestock, and potted plants are also important agricultural exports.

Most of the people are now employed in industry, trade, construction, and services. Tourism has become the most important sector of the economy. The construction industry has gained in importance because of increased demand for housing and other facilities related to tourism and to the resettlement of retired persons. Light industry has expanded and, in addition to food and cotton processing, includes the manufacturing of plastic bags, textiles, and electronic components. In 1956 access to Montserrat was facilitated by the completion of Blackburne Airfield, connections to which usually must be made through the international airport on Antigua; visitors previously depended upon irregularly scheduled ships. An international recording studio is located on the island.

Administration and social conditions. The constitution that took effect Jan. 1, 1960, provides for a governor appointed by the British monarch; the governor heads an executive and legislative council.

Primary education is free and compulsory to the age of 14. The Technical Training College provides vocational and technical training. A branch of the University of the West Indies and a medical school are in Plymouth. The island has radio and television broadcasting stations. The government maintains a hospital and provides free dental treatment for expectant and nursing mothers, school children, and the elderly.

Health

For statistical data on the land and people of Montserrat, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

Montserrat was sighted by Christopher Columbus in November 1493 on his second voyage to the New World; he named it for the jagged peaks he saw there. Inhabited by the Carib Indians, it was first colonized by Irish and English settlers led by Sir Thomas Warner in 1632. More Irish immigrants came from Virginia. Plantations were set up to grow tobacco and indigo, then cotton and sugar. The early settlers suffered repeated attacks from the French and Carib Indians. The French took possession of the island in 1664 and in 1667, but it was restored to England in 1668 by the Peace of Breda. The French sacked the island in 1712 and recaptured the island for the last time in 1782. Montserrat was again restored to England, this time by the Treaty of Versailles (1783).

In 1834 slavery was abolished (slaves had been introduced in 1664). The cessation of slavery and the falling price of sugar, combined with a series of devastating earthquakes and hurricanes between 1890 and 1936, brought the collapse of the stagnant plantation economy. The Montserrat Company, formed in 1857 under the direction of Joseph Sturge, bought abandoned estates, encouraged the cultiva-

Collapse of the economy

tion of limes, and sold plots of land to settlers. Much of Montserrat is still owned by small holders.

Between 1871 and 1956 Montserrat was part of the (British) Leeward Islands Colony, which included the British Virgin Islands, Saint Kitts-Nevis-Anguilla, and Dominica. The federation was abolished on July 1, 1956, when Montserrat became a colony in its own right. In 1958 Montserrat joined the West Indies Federation, which was dissolved in 1962 with the independence of Jamaica. Subsequent attempts to form a federation were abandoned in 1966. That year Queen Elizabeth II became the first British monarch to visit the colony.

In the general election of November 1978, the People's Liberation Movement (PLM), which favours complete independence, won all elective seats. In 1983 that party again won, but the opposition made some gains and the margin was again narrowed in the 1987 election. Montserrat's policy since the PLM ascendancy has been to pursue independence but at a measured rate, preferring first to achieve greater economic self-sufficiency. (J.D.Mo./Ed.)

For later developments in the history of Montserrat, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Saint Kitts and Nevis

The Federation of Saint Kitts and Nevis (also called Saint Christopher and Nevis), is a sovereign state composed of two islands of the Lesser Antilles in the eastern Caribbean Sea. Their combined area is 104 square miles (269 square kilometres). The capital is Basseterre on the island of Saint Kitts.

PHYSICAL AND HUMAN GEOGRAPHY

The land. Saint Kitts is 23 miles (37 kilometres) long and five miles wide, is oval in shape, and has an area of 68 square miles (176 square kilometres). A volcanic mountainous ridge down the centre forms a semicircle around a plain in the southeast. Mount Liamuiga (formerly Mount Misery), with a lake in its forested crater, is the highest point (3,792 feet [1,156 metres]). The soil (except in the mountains) is light and porous. Most of the beaches are of black volcanic sands. The island is well watered and fertile, with a cool, healthy climate. The average temperature is 80° F (27° C), and the annual rainfall averages 55 inches (1,397 millimetres).

Climate

Nevis, surrounded by coral reefs, lies two miles southeast of Saint Kitts across a channel called The Narrows. The island is circular, and it consists almost entirely of a mountain, Nevis Peak (3,232 feet), flanked by the lower Round Hill (1,014 feet) on the north and Saddle Hill (1,850 feet) on the south. Its area is 36 square miles (93 square kilometres). The soil of Nevis is clay studded with volcanic boulders. The climate is similar to that of Saint Kitts.

The people. The population is largely black, with a small mulatto minority and other mixtures. There are also very small East Indian and white groups. The official language is English. The main religious denominations are Anglican and Methodist, with small groups of Moravians and Roman Catholics. Both Saint Kitts and Nevis have traditionally had high levels of emigration, offsetting natural increases and enabling the islands to maintain a fairly stable population. Remittances from emigrants form an important source of foreign exchange.

The economy. The narrow coastal plain of St. Kitts, the skirts of the mountains, and the Basseterre Valley are devoted to the cultivation of sugarcane (mainly on large estates), which is the chief product and export. The government has nationalized all sugar plantations and has also purchased the sugar factories. Nevis grows chiefly cotton, vegetables, and coconuts. Light industries in Saint Kitts and Nevis produce items mainly for export from imported materials. Products include electronic equipment, batik-dyed fabrics, and other clothing and furniture. Tourism is also an important sector of the economy. The United States, the United Kingdom, and Trinidad and Tobago are the principal trading partners. There is a deepwater port at Basseterre, and Golden Rock International Airport on Saint Kitts provides service to other islands and to the United States and Canada. Newcastle Airfield on Nevis provides interisland air service.

Administration and social conditions. Since independence in 1983, the Federation of Saint Kitts and Nevis has been an independent member of the Commonwealth, with the British monarch as its head of state. An appointed governor-general represents the crown. The prime minister, who together with other ministers is a member of the Cabinet, is the head of government. The monarch and the National Assembly constitute the parliament, some of whose members are appointed. The island of Nevis enjoys a certain amount of autonomy within the federal structure; it has its own legislature, and the constitution provides for it to secede from the federation if certain procedures are followed. There is universal adult suffrage.

Administration of Nevis

Education is compulsory for all children from the age of five to 16. There are several hospitals and many health centres throughout the islands. Tropical diseases have been virtually eliminated. Most of the state's cultural activity is concentrated in the capital, Basseterre.

For statistical data on the land and people of Saint Kitts and Nevis, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

Early settlement. Christopher Columbus visited Saint Kitts on his second voyage in 1493 and found it inhabited by warlike Caribs. He named it Saint Christopher for his patron saint. The name was shortened to Saint Kitts by settlers under Sir Thomas Warner, who, arriving from England in 1623, established the first successful English colony in the West Indies at Old Road on the west coast. The French also settled on the island in 1627 under Pierre Belain d'Esnambuc. Divided during the 17th century between warring French and English colonists, Saint Kitts was given to Britain by the Treaty of Utrecht in 1713 and remained in British possession despite the capture in 1782 of Brimstone Hill by the French. The island was restored to Great Britain by the Treaty of Versailles in 1783.

Nevis was also sighted by Columbus in 1493. The island's name derives from Columbus' description of the clouds atop Nevis peak as *las nieves,* or "the snows," when he sighted the island. It was settled by the English in 1628 and soon became one of the most prosperous of the Antilles. Although it suffered from French and Spanish attacks in the 17th and 18th centuries, it maintained a sound economic position until the mid-19th century.

Federation and independence movements. The islands of St. Kitts, Nevis, and Anguilla were united by federal act in 1882 and became an independent state in association with the United Kingdom on Feb. 27, 1967. The islands were granted full internal self-government, with the United Kingdom retaining responsibility for defense and foreign affairs.

After the islands had assumed the status of associated states, Anguilla complained of domination by the Saint Kitts administration. In May 1967 the Anguillans ejected the Saint Kitts police and set up their own council; in July they proclaimed their independence. After unsuccessful negotiations, the Anguilla Act of July 1971 placed Anguilla directly under British control. On Feb. 10, 1976, Anguilla was granted a constitution and its union with Saint Kitts and Nevis was formally severed in 1980.

Political status

A constitutional conference was held in London in 1982, and, in spite of disagreement over special provisions for Nevis, Saint Kitts and Nevis became independent on Sept. 19, 1983. A drop in world sugar prices hurt the nation's economy through the mid-1980s, and the government sought to reduce the islands' dependence on sugar production and to diversify the economy. (G.E.M.M./J.D.Mo.)

For later developments in the history of Saint Kitts and Nevis, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Saint Lucia

Saint Lucia, an independent parliamentary state within the Commonwealth, is an island of the Lesser Antilles in the Caribbean Sea situated about 24 miles (39 kilometres)

south of Martinique and some 21 miles northeast of Saint Vincent. The capital and major port is Castries. Saint Lucia is 27 miles long and has a maximum width of 14 miles and an area of 238 square miles (616 square kilometres).

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief and drainage.* The island is of volcanic origin and is bisected from north to south by a central ridge of wooded mountains, the highest point being Mount Gimie (3,145 feet [958.6 metres]). Many streams flow from the mountains through fertile valleys. In the southwest are the Gros and Petit peaks (2,619 feet and 2,461 feet), two immense pyramids of rock rising sharply from the sea and enclosing a small bay. Near Petit Peak, in the crater of an ancient volcano, are the boiling sulphur springs from which the nearby town of Soufrière takes its name. A choice tourist site, the springs also contain substantial energy potential.

Plant and animal life. The forest, which has been reduced through lumbering, contains colourful orchids and anthurium. The rich birdlife includes the Saint Lucian parrot (the endangered national bird), the Saint Lucia black finch, and the Saint Lucia oriole. There is also a lizard unique to Saint Lucia, and the agouti is common.

Rainfall

Climate. Saint Lucia lies in the path of the northeastern trade winds and has a tropical maritime climate. Rainfall and temperature vary with elevation. Average annual rainfall ranges from 51 inches (1,295 millimetres) on the coast to as much as 150 inches (3,810 millimetres) in the interior. There is a dry season roughly from January to April and a rainy season from May to November. The mean temperature is about 80° F (27° C), with highs sometimes ranging into the upper 80s and lows into the upper 60s.

The people. No Caribs remain on the island; the vast majority of the inhabitants of the island are black and there is a small minority of mulattoes and other mixtures. The remainder are whites or of East Indian extraction. A French patois is spoken by most of the inhabitants but is being gradually supplanted by English, the official language. The main religion is Roman Catholicism, and Seventh-day Adventists and Anglicans are important minorities. The rate of population growth for Saint Lucia is slightly higher than the Caribbean average. The main population centres are Castries and Vieux Fort.

The economy. *Agriculture, forestry, and fisheries.* Sugarcane was formerly the chief crop, but production ceased entirely in 1964, when most of the cane fields were converted to banana cultivation. Bananas are now the principal crop. Other crops are coconuts, cacao, citrus and other fruit, spices, cassava, and yams. There is a steady local fishing industry.

Industry. Saint Lucia's manufacturing sector has been a major beneficiary of the U.S. Caribbean Basin Initiative, a program designed to promote manufacturing in the region. An industrial free zone has been established in the south of the island near Vieux Fort; there factories produce and export electronic goods and toys. Other industries produce cardboard cartons, clothing, rum, tobacco products, coconut products, concrete blocks, and beer. Tourism has developed rapidly since 1970; in 1987 a complex for cruise ships was opened near Castries.

Trade. The chief exports are bananas, cardboard cartons, clothing, coconut products, and electronic goods. Almost three-fourths of Saint Lucia's exports are to the United Kingdom and most of the rest to other Caribbean islands and to the United States. Imports include food, fuels, and manufactured goods.

Transportation. There is an international airport at Vieux Fort, at the southern tip of the island, and a smaller airport at Vigie for domestic and regional flights. International shipping lines operate from the ports at Castries and Vieux Fort. In addition, there is an oil transshipment terminal near Castries.

Administration and social conditions. *Government.* Saint Lucia is a constitutional monarchy with the British monarch as head of state, represented by a governor-general. The bicameral parliament consists of the House of Assembly elected by universal adult suffrage and the Senate, with members appointed on the advice of the prime minister, on the advice of the opposition leader in the House, and by the governor-general. The prime minister, leader of the majority party, heads the government.

Head of government

Justice. Saint Lucia has retained its association with the Eastern Caribbean Supreme Court, consisting of a court of appeal and a high court.

Education. Primary education is free and compulsory, and there is a branch of the University of the West Indies at Castries. There is also a technical college and a teacher-training college. Many of the primary schools are parochial, principally Roman Catholic.

Health and welfare. Several general hospitals and many health centres are distributed throughout the island. There is also a private hospital operated by a religious order at Vieux Fort.

Communications. Several newspapers are published on the island, and radio and television broadcasts are widely received.

For statistical data on the land and people of Saint Lucia, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

French and British territorial rivalry. The exact date of the European discovery of Saint Lucia is not known, but it is thought to have been about 1500. The first attempts at colonization were made by the English in 1605 and 1638, but they were frustrated by sickness and the hostility of the native Caribs. A successful settlement was achieved in 1650 by French from Martinique, who made a treaty with the Caribs in 1660. In 1664 Thomas Warner, son of the governor of Saint Kitts, regained the island, but it was restored to France by the Peace of Breda in 1667. In 1674 it was claimed by the crown of France and made a dependency of Martinique.

Another British settlement under a grant made in 1722 by George I to the Duke of Montague was frustrated by France, which had granted the island to Marshal d'Estrées in 1718, and the island was declared neutral. In 1743 the French resumed possession, retaining the island until the Treaty of Aix-la-Chapelle in 1748, in which the two countries again agreed to regard Saint Lucia as neutral. In 1762 it was captured by Admiral George Rodney and General Robert Monckton, only to be given up once more by the Treaty of Paris (1763). In 1778 it again surrendered to the British, who used its harbours as a naval base, but, by the Peace of Versailles, Saint Lucia was once more restored to France. Between 1782 and 1803 the possession of Saint Lucia passed several times between Britain and France, the British having to suppress a vigorous revolutionary party, which was aided by insurgent slaves, before gaining possession in 1803. Saint Lucia was finally ceded to Britain in 1814 by the Treaty of Paris, after which it became a crown colony. During 1838–85, together with the other islands of the Windward group, it was administered by the governor of Barbados.

Cessation to Britain

French influence on the development of Saint Lucia is illustrated by the dominance of the Roman Catholic church and the survival of a French patois. In the years following 1763, French planters came from Saint Vincent and Grenada and established cotton and sugar plantations. In 1834, when the slaves were emancipated, there were in Saint Lucia more than 13,000 black slaves, 2,600 free blacks, and 2,300 whites. Prosperity was impeded by the decline of the sugarcane industry. Improvement came with the increase of banana and cacao cultivation and the revival of sugarcane.

Independence. Representative government was obtained by the constitution of 1924, which introduced an elective element into the legislative council; the constitution of 1936 provided for an unofficial majority in the council.

In 1958 Saint Lucia joined the West Indies Federation, although its colonial status remained unchanged. Under the 1960 constitution the post of governor of the Windward Islands was abolished, and Saint Lucia became an autonomous unit within the federation, also achieving a greater degree of internal self-government. After the federation was dissolved on May 31, 1962, the status question was eventually settled by the West Indies Act of 1967, in

which Saint Lucia assumed a status of association with the United Kingdom on March 1, 1967.

Independence was finally achieved Feb. 22, 1979, with Saint Lucia remaining a parliamentary democracy within the Commonwealth. In the first elections following independence, the left-leaning Saint Lucia Labour Party (SLP) defeated the more conservative United Workers' Party (UWP). The SLP governments favoured the socialist regimes of the Caribbean, establishing relations with Cuba and joining the nonaligned movement. They also helped form the Organization of East Caribbean States in 1981.

The SLP's attempts at a mixed economy proved unable to deal with the staggering problems of the new country, especially after Hurricane Allen wiped out the banana crop in 1980. Rivalry within the party led to the resignation of two prime ministers in two years, and political instability caused the tourist trade to drop to one-half of its pre-independence level. In May 1982 the UWP was voted into power on a platform of inviting foreign investment and decentralizing government administration. The agricultural sector was gradually rebuilt after the hurricane, and the end of a recession in the United States increased tourism. The UWP retained power in 1987 by a narrow margin. (R.To./D.L.N./J.D.Mo.)

Hurricane Allen

For later developments in the history of Saint Lucia, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Saint Vincent and the Grenadines

Saint Vincent and its associated islands of the northern Grenadines form an independent republic within the Commonwealth. It is part of the Lesser Antilles in the Caribbean Sea, about 21 miles (34 kilometres) southwest of Saint Lucia and 100 miles west of Barbados. Saint Vincent is 18 miles long and has a maximum width of 11 miles, and its area is 134 square miles (347 square kilometres). Some of the larger islands of the Grenadines are Bequia, Canouan, Mayreau, Mustique, and Union Island. Including the Grenadines, the territory comprises a total area of about 150 square miles. The capital and major port is Kingstown, on Saint Vincent.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief, soils, and drainage.* The island of Saint Vincent has thickly wooded volcanic mountains running north to south and producing many short, swift streams. The streams are numerous but, except after heavy rains, small. There are no navigable rivers. The highest peak is the volcano Soufrière (4,048 feet [1,234 metres]) in the north, which erupted disastrously in 1812 and 1902, when the entire northern half of the island was devastated. The 1902 eruption coincided with that of Mount Pelée in Martinique. Soufrière became active again in 1979, causing massive evacuation and much damage to agriculture. The volcanic ash has produced a fertile soil that has given rise to a lush green vegetation. Birdlife on the island is especially rich. Some of the larger Grenadines have hills, and many are noted for their coral reefs and fine beaches.

The Soufrière volcano

Climate. Saint Vincent lies in the path of the northeast trade winds and has a tropical maritime climate. Rainfall and temperature vary with altitude. Average annual rainfall ranges from about 60 inches (1,524 millimetres) on the coast to 150 inches in the central mountains. The temperature at Kingstown averages between 64° and 90° F (18° and 32° C). Hurricanes occasionally pass across the island and it suffered severely in 1780, 1898, and 1980. From January to May is the dry season on St. Vincent, and the rains start in June and continue from that time to the end of the year.

The people. Nearly three-fourths of the inhabitants are black, and another fifth is mulatto. A small minority is of European descent, and there is a small group of East Indian extraction; only a few are of Carib Amerindian stock, but there is a small group of black-Amerindian mixture. English is the official language and some French patois is spoken. The main religious denominations are Anglican, Methodist, and Roman Catholic. Saint Vincent and the Grenadines has one of the highest birth rates among the West Indies, giving the islands an extremely high rate of population increase that is moderated by emigration.

The economy. *Agriculture, forestry, and fisheries.* The economy of Saint Vincent is chiefly agricultural. Traditional crops are arrowroot (of which the country is a major producer) and sugarcane, but since the 1950s bananas have become the leading export and the production of sugarcane has become negligible. Other crops include sweet potatoes, plantains, yams, ginger, and nutmeg. Nearly half of the island is forested, the woodland being used for charcoal burning. Both inshore and offshore fishing are carried on.

Industry and trade. Industry is based primarily on the processing of agricultural products. Soap and edible oils are produced, and there are also plants for processing copra, distilling rum, milling flour, building yachts, and making boxes for packing bananas. Tourism has assumed a larger role in the economy, but the small number of beaches and the inaccessibility of the Grenadines restrict development.

Development of tourism

The main exports are bananas, root crops, sweet potatoes, flour, and plantains. Carrots are exported to Trinidad and Tobago and Barbados, and plantains are also exported to Barbados. The country's main trading partners include the United Kingdom, Trinidad and Tobago, the United States, and Canada.

Transportation. The major airport is at Arnos Vale, southeast of Kingstown. Three of the Grenadines have airstrips for light aircraft. Kingstown has a deepwater port.

Administration and social conditions. *Government.* The British monarch is the head of state and is represented on the island by an appointed governor-general. The unicameral legislature is composed of the House of Assembly, containing elected representatives and senators appointed by the governor-general. The prime minister, leader of the majority party, is appointed by the governor-general to lead the government.

Justice. Saint Vincent retains its connection with the Eastern Caribbean Supreme Court. This consists of an appeals court and a high court, while the final court of appeal remains the judicial committee of the privy council.

Education, health, and welfare. Primary education is free but not compulsory. Secondary education is conducted mostly in schools administered by religious organizations. Health measures are directed primarily against infant malnutrition and gastroenteritis. The islands have a general hospital and several health centres.

Culture and recreation. The government runs a free public library system, and there is a weekly newspaper. Kingstown has a botanical garden. The Grenadines are particularly known for sailing.

For statistical data on the land and people of Saint Vincent and the Grenadines, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

Colonization. Saint Vincent may have been given its name by Christopher Columbus, who is thought to have visited the island on Jan. 22, 1498 (St. Vincent's Day). Its Carib inhabitants were left almost undisturbed until the 18th century. In 1673 the first Africans arrived, a party of slaves who were shipwrecked in the Grenadines and eventually reached Saint Vincent, intermarrying with the native Caribs. French, Dutch, and British settlements were attempted, with the French dominant until the Seven Years' War, when the British general Robert Monckton occupied it (1762). The Treaty of Paris (1763) confirmed British possession, and settlement proceeded in spite of Carib refusal to accept British sovereignty. In 1779 the island was seized by the French, but it was restored to Britain in 1783. In 1795 the Caribs rose in revolt, assisted by the French, but they were finally subdued the following year. Most of them were then deported to the Bay Islands off Honduras and Belize. The emancipation of black slaves in 1834 decreased the island's labour supply, and Portuguese and East Indian labourers were introduced late in the century.

The arrival of Africans

In the latter half of the 19th century, sugar prices fell, plunging the island into a depression that lasted through

the end of the century. The hurricane of 1898 and the volcanic eruption of 1902 hindered recovery.

In 1958 Saint Vincent joined the West Indies Federation. In 1960 it received a new constitution. It became a state in association with the United Kingdom on Oct. 27, 1969. The island had become a member of the Caribbean Free Trade Area on July 1, 1968.

Independence. Plans put forward in 1972 for the unifications of Saint Vincent with Grenada and Saint Lucia were later dissolved. Independence for Saint Vincent and the Grenadines, as the nation was called, was achieved on Oct. 27, 1979. The new government was formed as a constitutional monarchy and became a member of the Commonwealth. The Saint Vincent Labour Party won the elections of December 1979, and Milton Cato became the first prime minister. Cato was critical of the revolution in Grenada and of developments in Jamaica and Guyana. He preferred closer links with the relatively centrist governments of Trinidad and Tobago and of Barbados.

In 1979 the Soufrière volcano erupted once again, damaging agriculture and the tourist trade. Hurricane Allen virtually wiped out the all-important banana crop in 1980. Recession in the United States and the falling value of the pound sterling against the dollar further lowered tourist visits and exports of bananas.

In July 1984 the New Democratic Party, under James Mitchell, won the general elections. Mitchell began a program of reorganizing agriculture and of lowering unemployment (which stood at more than 30 percent) by encouraging the construction industry. Under his leadership the economy steadily improved, though high unemployment remained a serious problem. Mitchell began a second five-year term in 1989 after elections gave his party all 15 elective seats in the legislature. (R.To./D.L.N./Ed.)

For later developments in the history of Saint Vincent and the Grenadines, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Virgin Islands

The Virgin Islands are a group of about 90 small islands, islets, cays, and rocks in the West Indies, situated some 40 to 50 miles (64 to 80 kilometres) east of Puerto Rico. The islands extend from west to east for about 60 miles and are located west of the Anegada Passage, a major channel connecting the Atlantic Ocean and the Caribbean Sea. Their combined land area is about 195 square miles (505 square kilometres).

British and U.S. groups

The islands are administered in two groups—the British Virgin Islands and the Virgin Islands of the United States. The former is a British colony consisting of four larger and 32 smaller islands and islets. Their total area is 59 square miles, and they lie to the north and east of the U.S. islands. The latter group, administered by the U.S. Department of the Interior as an incorporated territory, consists of three larger islands and some 50 smaller islets and cays, with a total area of 133 square miles.

The islands are noted for their inviting subtropical climate, which attracts a large number of tourists each year to swim in the warm aquamarine waters and frequent the sandy beaches and harbours. Apart from the tourist industry, the islands have few economic resources; financial aid is provided by the United Kingdom and the United States, respectively. Fresh water is scarce. In recent years some tension has arisen between the inhabitants of the islands and people from other parts of the Caribbean who have immigrated, particularly to the Virgin Islands of the United States, to seek jobs and better living conditions.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Although the Virgin Islands form the easternmost extension of the Greater Antilles, they are often included in discussions of the Lesser Antilles because of their size and proximity to that island chain. The Virgin Islands themselves are the peaks of submerged mountains that rise from a submarine plateau. While the Caribbean deepens to a 15,000-foot trench in the area between the island of St. Croix, to the south, and the remainder of the group to the north, the greater part of the plateau is covered by not more than 165 feet of water. Most of the islands rise only to a few hundred feet above sea level, although isolated peaks reach heights of well over 1,200 feet. The highest point is Mount Sage on Tortola, which is 1,780 feet (543 metres) high.

The principal islands

Of the 36 British islands, 16 are inhabited. Tortola (Turtle Dove), with an area of 21 square miles, is the largest and is the site of the group's capital and population centre, Road Town. Other larger islands in the British group are Anegada, with an area of 15 square miles; Virgin Gorda (the Fat Virgin), with an area of 8 square miles; and Jost Van Dyke, about 3 square miles. Lesser islands include Great Tobago, Salt, Peter, Cooper, Norman, Guana, Beef, Great Thatch, Little Thatch, and Marina Cay.

About 50 islands and cays constitute the U.S. group. Only three are of importance; several are uninhabited. The largest, St. Croix, is 28 miles long, 84 square miles in area, and lies about 40 miles south of the other islands. The island of St. Thomas, 32 square miles in area, is the site of the territory's capital, Charlotte Amalie. St. John has an area of 20 square miles. At the closest point, between Great Thatch Island and St. John, a distance of only half a mile separates the British and the U.S. groups.

The landscape of the islands offers scenes of dramatic contrast, varying from craggy cliffs and mountaintops to small lagoons with coral reefs and barrier beaches, from landlocked harbours to unprotected bays, and from small, level plains to elevated plateaus with rolling lands and junglelike regions. Individual islands have their own distinguishing characteristics.

In the British group, Tortola, of the same geologic formation as St. John, is hilly, with unbroken ranges running throughout its 15-mile length. Road Bay is Tortola's most important bay; it is exposed to the southeast but protected from all other sides by an amphitheatre of hills. Virgin Gorda, an island with several peninsulas, is rectangular in shape, about 2½ miles long, and 1¾ miles wide in the central part of the island. Its highest peak rises 1,359 feet. Anegada is the only flat island of the group. Its elevation is never more than 10 to 15 feet above sea level, and its coast, because of its many reefs, is dangerous to boats. Jost Van Dyke is a hilly, almost rugged island with two fine beaches on the south side.

St. Thomas

In the U.S. group, St. Thomas, composed primarily of a ridge of hills running east and west with branching spurs, has little level, tillable land. Crown Mountain (1,556 feet), northwest of the capital of Charlotte Amalie, is the island's highest elevation. Charlotte Amalie, facing a fine landlocked harbour, is built on five foothills. There are a number of springs on the island's northern side but only one small stream. Magens Bay, with 3,500 feet of white sandy beach, is reputed to be the finest beach in the West Indies. St. Thomas is surrounded by 17 islands and by cays and innumerable rocks.

St. Croix rises abruptly on the north to Mount Eagle (1,088 feet) and Blue Mountain (1,096 feet), but southward the land slopes to flatlands that near the coast are laced with lagoons. The island's only urban centres, Christiansted and Frederiksted, lie on the flat land. Since the coastal indentations are slight, there are few harbours and sheltered bays; dangerous reefs lie along the north and south coasts. While there are several rivulets on the island, it is generally poorly watered.

St. John—three miles east across Pillsbury Sound from St. Thomas and lying closest to the British Virgin Islands—has steep, lofty hills and valleys but little level, tillable land. Its highest elevations are Camelberg Peak (1,193 feet) and Bordeaux Mountain (1,277 feet). Its coastline is indented with forests and many fine, sheltered bays. Coral Bay, on the eastern end, whose steep shores allow large vessels to come close in, has been described as the best natural harbour in the Virgin Islands. A number of small streams on the south side of St. John, together with a multitude of springs, make it the best-watered island of the U.S. group. More than three-quarters of its area, about 14,700 acres (5,949 hectares; including Hassel Island in St. Thomas harbour), is preserved as Virgin Islands National Park.

Climate. The splendid climate is perhaps the Virgin Islands' chief asset. Although they are located in the tropics,

the heat is tempered by gentle trade winds that blow from the northeast most of the year. Humidity is low, and there is little pollen. The temperature rarely exceeds 90° F (32° C) or falls below 70° F (21° C). The average temperature is about 78° F (26° C). The dry season lasts from February to July, and the wet season from September to December. Hurricanes—averaging perhaps four in a century—occur usually between August and October, and there are occasional light earthquakes.

The water scarcity is so serious that nearly all buildings, both private and public, have their own water catchments.

Rainfall

Rainfall averages between 45 and 50 inches (1,143 and 1,270 millimetres) a year, but much of it runs off unused. In the driest sections of the large islands, rainfall usually averages a little less than 30 inches, with possibly as much as 80 inches on the upper slopes of Mount Sage on Tortola. But rainfall is erratic, varying widely from year to year.

Virgin Islanders have long depended almost entirely upon their own cisterns and wells and, in addition, have imported water in barges to meet their needs for fresh water—needs now rapidly increasing in proportion to population and industrial growth. Only Road Town in the British group has a piped supply. The first flash-type evaporator for the desalination of seawater in the Western Hemisphere was installed in St. Thomas, where the Virgin Islands Water and Power Authority, established in 1965, now operates several saltwater-distillation plants. A seawater desalination plant is located in Christiansted on St. Croix. Maximum use of salt water is made everywhere in the islands.

Plant and animal life. Vegetation is tropical. Supported in most places by thin soil, it includes royal poinciana (flamboyant) trees and other lush blooms, but the islands' generally sparse stands of shrubs and trees are not sufficient to be of commercial value. Among the tree species are mangoes, soursop (a small tropical tree with a large succulent fruit), coconut palms, and breadfruit. Cacao and wild orchids grow in the hills, while cactus, acacia, grass, and sugarcane flourish in the lowlands. Even though the woodlands are not dense, there are numerous species of birds and small game, such as deer. Sailfish, tarpon, marlin, kingfish, and wahoo abound in coastal and offshore waters.

The people. The population is overwhelmingly made up of blacks descended from African slaves. The number of Puerto Ricans and persons from the continental United States has increased in recent years. Less than half of the U.S. Virgin Islands population is native-born.

Despite a long tradition of racial equality and the absence of legal or de facto discrimination in housing, churches, theatres, and public facilities, there have been growing signs of discontent with a type of class discrimination that to a great extent corresponds with the difference between white and black. While there has been relatively little racial violence, there is an increasing demand by those of African descent for a greater role in determining economic and social decisions and in running the government.

The Chachas

The Chachas of St. Thomas form a distinct ethnic unit apart from the other islanders. They are descended from French Huguenots who a century ago came from Saint-Barthélémy—a West Indian island the French purchased from Sweden in 1877, after holding it themselves from 1648 to 1784. The Chachas maintain themselves as a clannish, aloof, industrious, fisher–farmer community.

The traditional language of the islands is English—much of it spoken in a dialect termed Calypso, which varies somewhat from island to island but is mutually intelligible among most West Indians. Some French is heard on St. Thomas, and there are numerous Spanish-speaking people on St. Croix, where many Puerto Ricans have settled.

Religious freedom has existed since the 1600s. The islands are predominantly Protestant, the largest denomination in the British Virgin Islands being Methodist and in the U.S. islands, Anglican. The second oldest Lutheran church in the Western Hemisphere held its first service in Charlotte Amalie in 1666, and the second oldest American synagogue is in St. Thomas. Roman Catholics form a large minority in the U.S. islands.

The British and U.S. components of the Virgin Islands differ demographically. The U.S. islands have about nine times the population of their British neighbours; the British islands have the lower birth rate, while the U.S. islands have a greater rate of natural population increase. The death rates for both entities are lower than the Caribbean mean. Both island groups have rates of total population growth that are affected substantially by immigration. The principle urban places are Road Town on Tortola and Charlotte Amalie on St. Thomas.

The economy (British Virgin Islands). Despite the colony's small size, British Virgin Islanders have a fairly high standard of living for the Caribbean region. The economy is based largely on tourism. The U.S. dollar is used as the official currency.

Agriculture. Despite the handicaps of difficult terrain and uncertain water supply, agriculture and stock raising (largely for export to the U.S. Virgin Islands) are undertaken. Farms are usually small holdings worked by owner-occupiers, many of whom are also part-time fishermen. The chief crops are fruits, vegetables, and coconuts for both domestic use and export; sugarcane is grown for the distillation of rum. Many of the younger people seek work in the U.S. islands, where wages are higher.

Industry. The tourist industry makes up almost half of the national economy. New roads have been built, harbours improved, and tourist hotels built. The number of tourists visiting the islands, attracted by opportunities for sport fishing and sailing, has continued to increase.

Burning wood for charcoal is a minor industry, and there are several boatyards for the construction and repair of vessels. Cement blocks and rum are manufactured.

Trade. Imports—mostly from the United States, Puerto Rico, and the United Kingdom—consist chiefly of foodstuffs, beverages, machinery, motor vehicles, building materials, and petroleum products. Exports, mainly to the U.S. Virgin Islands, include fresh fish, rum, sand and gravel, charcoal, fruit, and vegetables.

Transportation. Tortola has two main highways and numerous side routes; Virgin Gorda, Anegada, and Jost Van Dyke also have road networks. Small boats ply to and from the U.S. Virgin Islands. An airport, reconstructed in 1969, is located on Beef Island, which is connected by bridge to Tortola. Another airport, on Anegada, was opened in 1969, and there is plane service to the U.S. Virgin Islands, Puerto Rico, and Antigua. Road Harbour on Tortola is a deepwater port.

Sources of income

The economy (U.S. Virgin Islands). Until well into the 20th century, sugarcane and, to a lesser extent, cotton provided the main economic base. The harbour at St. Thomas also generated some income. When the United States acquired the islands, both social and economic conditions were poor. The first U.S. governor reported in 1917 that the islands were incapable of self-support. Since then, millions of dollars of U.S. aid have failed to make them self-supporting. U.S. aid and the development of tourism have nevertheless resulted in the territory's having one of the highest incomes per capita in the Caribbean.

Resources. Minerals of commercial value do not exist, although sand, rock, and gravel are present in quantities sufficient for construction purposes.

Fisheries. There is no large commercial fishing industry, but fish represent an important part of the islanders' diet, and there is a shellfish-farming operation on St. Croix. Tourism has encouraged sport fishing.

Industry. As in the British Virgin Islands, tourism is the most important economic sector. In the U.S. islands there are more than 200 miles of beaches, several historic 17th- and 18th-century buildings, pleasant vistas of mountain and sea, and numerous recreational facilities. The picturesque free port of Charlotte Amalie is also an attraction. Small manufacturing operations, including watch making, textile manufacturing, rum distilling, and pharmaceutical production, have been encouraged, as have some larger ones. One of the world's largest oil refineries is located on St. Croix, and petroleum products are the islands' leading export. Tax concessions were granted by the United States in1987 to encourage manufacturing.

Trade. The U.S. islands depend heavily on imports for

their survival—most importantly crude oil for the large refinery. Food is the next most important import. Other than refined petroleum, exports include chemical products, watches and watch movements, and rum. The main trading partner is the United States.

Transportation. The U.S. islands have a fairly good road network. Taxis and rental vehicles are available on all three islands, and regular passenger-bus services operate on St. Croix and St. Thomas. Interisland transport by small boat is available. Seagoing passenger and cargo vessels connect the ports of Charlotte Amalie, on St. Thomas, Frederiksted and Limetree Bay, on St. Croix, to ports abroad. International jet air services operate on St. Thomas and St. Croix.

Administration and social conditions (British Virgin Islands). *Government.* The islands are a crown colony with a governor, appointed by the British crown, who is responsible for defense and internal security, external affairs, and the civil service. He exercises his powers in consultation with the Executive Council, over which he presides. The council consists of three ministers plus the chief minister and the attorney general. The Legislative Council consists of one ex officio member (the attorney general), nine members elected by universal adult suffrage, a speaker, who is elected from outside the council by its members, and one member appointed by the governor.

The Executive Council

Justice. The law of the colony is made up of both the common law of England and statutory law, or locally enacted legislation. It is administered by the Eastern Caribbean Supreme Court, courts of summary jurisdiction, and magistrates' courts. The principal law officer is the attorney general.

Education and health. Education is compulsory and free, but there is a shortage of schools. The islands have several private schools, but there are no institutions of higher learning. Health conditions are fairly good. The administration carries out a regular immunization program, and there is a modernized hospital on Tortola. A program of social welfare has also been implemented.

Administration and social conditions (U.S. Virgin Islands). *Government.* Jurisdiction is exercised by the U.S. Department of the Interior. Limited legislative powers are held by a unicameral legislature consisting of 15 senators, each elected by popular vote for two-year terms. The governor and lieutenant governor are also elected by universal adult suffrage. There are 12 executive departments, of which 11 are headed by commissioners; the 12th, the Department of Law, is headed by the attorney general. Attempts to redraft the constitution to provide greater autonomy have been rejected, most recently in 1981. Residents of the Virgin Islands do not vote in U.S. presidential elections. They are represented in the U.S. House of Representatives by a nonvoting delegate.

Justice. Judicial power in the islands resides in municipal courts and in the Federal District Court of the Virgin Islands. The district judge and the district attorney are appointed by the president of the United States, with the advice and consent of the U.S. Senate. Municipal court judges are appointed by the governor, subject to confirmation by the legislature.

Education and health. Education is compulsory and free, but many students attend private schools. The University of the Virgin Islands (established in 1962) has campuses on both St. Thomas and St. Croix. Health services are more extensive in the U.S. than in the British islands. There is a large general hospital on St. Thomas, and hospital services are available on the other two large islands. Mobile units reach the outlying islands.

Health

Cultural life. The Virgin Islands have little in the way of a national culture. There is no typical music and little folklore. In 1965, however, the Virgin Islands Council of the Arts was created as an adjunct of the U.S. National Foundation on the Arts and Humanities. Community arts councils have been formed on all three American islands. They sponsor theatrical performances and provide training in the arts and crafts.

For statistical data on the land and people of the U.S. Virgin Islands, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

Pre-Columbian inhabitants of the islands probably included the Arawak, who were displaced by the Caribs; the latter had reached the stage of stone polishing and pottery making when Christopher Columbus arrived. On his second voyage, in 1493, Columbus dropped anchor at what is now known as Salt River Bay, St. Croix (which he called Santa Cruz), and sent a landing party ashore in search of fresh water and fruit. After a skirmish, the Caribs repulsed the Spanish. Columbus later encountered some of the other islands and named the group Santa Ursula y las Once Mil Virgenes (St. Ursula and the Eleven Thousand Virgins). In 1555 the Holy Roman emperor, Charles V (Charles I of Spain), sent forces that defeated the Caribs, claimed the territory, and ordered the annihilation of the natives. By 1596 most of the natives had been killed or had left.

Settlement and history of the British Virgin Islands. Tortola was first settled in 1648 by Dutch buccaneers who held the island until it was taken over in 1666 by a group of English planters. In 1672 Tortola was annexed to the British-administered Leeward Islands. In 1773 the planters were granted civil government, with an elected House of Assembly and a partly elected Legislative Council, and constitutional courts. The abolition of slavery in the first half of the 19th century dealt a heavy blow to the agricultural economy. In 1867 the constitution was surrendered and a legislative council was appointed that lasted until 1902, when sole legislative authority was vested in the governor-in-council. In 1950 a partly elected and partly nominated legislative council was reinstated. Following the defederation of the Leeward Islands colony in 1956 and the abolition of the office of governor in 1960, the islands became a crown colony. In 1958 the West Indies Federation was established, but the British Virgin Islands declined to join, in order to retain close economic ties with the U.S. islands. Under a constitutional order issued in 1967, the islands were given a ministerial form of government. The constitution was amended in 1977 to permit a greater degree of autonomy in internal affairs.

Settlement and history of the U.S. Virgin Islands. In 1666 St. Thomas was occupied by Denmark, which five years later founded a colony there to supply the mother country with sugar, cotton, indigo, and other products. Slaves from Africa were first introduced to St. Thomas in 1673 to work the cane fields, but the first regular consignment of slaves did not arrive until 1681. In 1684 the Danes claimed neighbouring St. John, which had been used primarily as a base by buccaneers. Denmark colonized the island with planters from St. Thomas in 1717. In 1733 they abandoned St. John after slaves rebelled, staged an uprising, and held the island for six months. They then purchased St. Croix, which had been in the possession of the French since 1651. Slaves continued to be imported from Africa; by 1742 there were 1,900 on St. Croix alone, and sugar production was bringing prosperity to the islands. The group came under the Danish crown in 1754, and the following year Charlotte Amalie was made a free port. The British occupied the islands from 1801 to 1802, and in the next year, 1803, the slave trade was abolished in the Danish West Indies. The British reoccupied the islands from 1807 to 1815, after which they reverted to Danish rule until 1917. Slavery itself was abolished in 1848 after a serious uprising in that year. Sugarcane production dropped, and there was a general decline in economic activity.

The Danish era

U.S. interest in the islands began in the Civil War period, but the U.S. Senate refused in 1870 to approve the purchase of St. Thomas and St. John for $7.5 million. The United States moved decisively only in World War I, when it was seen to be strategically important to control the main passage through the Caribbean to the Panama Canal, as well as routes along the eastern coasts of the American continent. Denmark at that time was willing to sell to avoid the jeopardy of seizure by the Allies or conquest by Germany, which then owned Hamburg–America Line docks, warehouses, steamers, and other property in St. Thomas. In 1917 the United States purchased the three islands for $25 million and the Virgin Islands became an

The U.S. purchase

unincorporated territory of the United States. The treaty of cession promised U.S. citizenship to the inhabitants, except where the choice of retaining Danish citizenship was exercised. An act in 1927 granted U.S. citizenship to most of the Virgin Islanders, and in 1932 another provided that all natives of the Virgin Islands who on the date of the act were residing in the continental United States or any of its insular possessions or territories should be U.S. citizens. The transition was accomplished smoothly by retaining the Danish government organization and legal code. All military, civil, and judicial power was invested in a governor appointed by the president of the United States. Administration was the responsibility of the U.S. Navy Department from 1917 until 1931, when jurisdiction was transferred to the Department of the Interior.

The Organic Act of 1936, enacted by the U.S. Congress for the establishment of congressional government, provided for two municipal councils, one for St. Thomas and St. John, the other for St. Croix, and a council for the whole territory. A Revised Organic Act adopted in 1954 created a central government and abolished the independent municipal councils, authorized distinct executive, legislative, and judicial branches, and provided for a substantial degree of self-government. In 1968 an act was approved, to take effect in 1970, legalizing the popular election of the islands' governor and lieutenant governor for four-year terms.

There has been little demand for autonomy from either the British or U.S. Virgin Islands, mostly because of fears of disrupting the lucrative tourist industry and of having increased taxes. Independence appears not to be the goal for either territory, and statehood continues to be only a remote possibility for the U.S. islands. The inability of either territory to develop a self-supporting economy continues to encourage their dependent status. (L.H.E./Ed.)

For later developments in the history of the U.S. and British Virgin Islands, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

OTHER ISLANDS

Aruba

The island of Aruba lies southwest of the Lesser Antilles in the Caribbean Sea, some 50 miles (80 kilometres) northwest of Curaçao and 18 miles north of the Venezuelan peninsula of Paraguaná. Politically a part of the Netherlands Antilles until 1986, Aruba in that year became a separate self-governing part of the Kingdom of The Netherlands. Its capital and main port is Oranjestad.

PHYSICAL AND HUMAN GEOGRAPHY

The land. Aruba, with an area of 75 square miles (193 square kilometres), is 19.6 miles long and 6 miles across at its widest point. The island is generally low in altitude and consists of igneous rocks, fringed with coral reefs. The 620-foot (189-metre) Mount Jamanota is the highest point and the 560-foot monadnock Mount Hooiberg (meaning Haystack) the most characteristic hill. In some places immense monolithic boulders of diarite are peculiarly piled on top of one another. Aruba has barren soil with little or no natural irrigation. Most drinking water is obtained by desalinating seawater. The temperature varies little from an annual average of 81° F (27° C), and the heat is tempered by northeasterly trade winds. Rainfall is low and variable, usually amounting to about 17 inches (430 millimetres) a year. The island lies outside the normal range of hurricanes. The natural vegetation consists of a variety of drought-resistant cacti, shrubs, and trees.

Topography

The people. Most of Aruba's population is racially mixed, including a considerable percentage of Amerindian stock, often in combination with Dutch, Spanish, and African strains. The black influence is minimal, however, few slaves having been imported to Aruba. The official language is Dutch, but the common language is Papiamento (also spoken in Curaçao and Bonaire), a creole that evolved mainly from Portuguese, Spanish, and Dutch. English and Spanish are also widely used. The major religion is Roman Catholicism. The birth and death rates are both relatively low, and the rate of natural increase is less than average for the West Indies.

The economy. Until the end of the 18th century, Aruba was used by the colonial authorities for horse breeding, the local and mainland Indians serving as herdsmen. Only from the early 19th century on was land sold to individual settlers. Agriculture remained of little importance despite efforts to grow aloe for pharmaceutical products. Gold mining began in 1824 but was discontinued by the early 20th century. Aruba's economy improved when oil refining started in the 1920s at the port of San Nicolas; crude oil was imported mainly from Venezuela. The closing of the refinery in 1985 provoked a serious economic crisis, which has been overcome by aggressive promotion and expansion of tourism, including the building of luxury hotels and casinos to complement the idyllic island setting. Attempts to diversify the economy include the development of a free-trade zone and plans to develop Aruba into an international offshore financial centre. Aruba's foreign trade is mainly with the United States, Venezuela, and The Netherlands. The island has an international airport and is further linked to the outside world by steamship and cruise ship services.

Foreign trade

Administration and social conditions. A governor, appointed by the Dutch crown, is the formal head of government and representative of the reigning monarch of The Netherlands. Executive authority is vested in a Council of Ministers, headed by a prime minister. The council is responsible to a unicameral legislature, the Staten (States), elected by universal adult suffrage. Most of the population is literate. All levels of education, including college, are available. Health standards on the island are high.

Cultural life. The Cultural Center at Oranjestad offers concerts, ballet, folkloric presentations, and art exhibits throughout the year. The city also includes historical, archaeological, and numismatic museums. Carnival time in February and the New Year's Day festivities are especially colourful celebrations.

For statistical data on the land and people of Aruba, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

Aruba was discovered and claimed by Spain in 1499. In 1636 it was taken by the Dutch and occupied by the Dutch West India Company. As part of the Netherlands Antilles, Aruba came briefly under British rule during the Napoleonic Wars but was returned to The Netherlands in 1816. The economy remained weak until the construction of an oil refinery in the 1920s that raised the standard of living dramatically. Immigration from the rest of the Caribbean, the United States, Venezuela, and Europe contributed to a substantial increase in population. Despite its new economic strength, Aruba remained politically subordinated to the main island of Curaçao. The autonomous status obtained in 1986 was the result of a popular movement to break away from Curaçaoan—rather than from Dutch—domination. Although full independence is scheduled for 1996, the Aruban government has scheduled a referendum for 1994, which it contends should establish the date for independence. (H.Ho.)

For later developments in the history of Aruba, see the *Britannica Book of the Year* section of the BRITANNICA WORLD DATA ANNUAL.

The Bahamas

The Commonwealth of The Bahamas, a former British colony, became an independent nation within the Commonwealth in 1973. The island group comprising The Bahamas (from Spanish *bajamar,* "shallow water") occupies an irregular submarine tableland that rises out of the

Atlantic depths and is separated from nearby lands to the south and west by deepwater channels. Lying to the north of Cuba and Hispaniola, the archipelago comprises nearly 700 islands and cays, only about 30 of which are inhabited, and more than 2,000 low, barren rock formations. It stretches more than 500 miles (800 kilometres) southeasterly from Grand Bahama Island, which lies about 60 miles off the southeastern coast of Florida, to Great Inagua Island, some 50 miles from the eastern tip of Cuba. The total land area is 5,382 square miles (13,939 square kilometres).

Strategic location

The strategic position of the Bahama Islands, which lie at the geographic centre of the New World landmass, commanding the gateway to the Gulf of Mexico, the Caribbean Sea, and the entire Central American region, has given the history of the islands a unique and often striking character. It was there that Christopher Columbus made his original landfall in the Americas. The subsequent fate of the peaceful original inhabitants remains one of the more tragic episodes in the development of the entire region, while the early attempts at European-dominated settlement were marked by intense national rivalries, interspersed with long periods of lawlessness and piracy. As a result, the society and culture that has evolved in The Bahamas is a distinctive blend of European and African heritages, the latter a legacy of the slave trade. The islands, lacking natural resources other than their magnificent climate and dazzling beaches, have become heavily dependent on the income generated by the extensive tourist facilities that have been developed, often as a result of the injection of foreign capital. The continued popularity of the islands, largely with North American tourists, has maintained a relatively high standard of living among the population, most of whom are black.

The capital city, Nassau, is situated on the small (80 square miles) but important New Providence Island. Other islands, known collectively as the Family (or Out) Islands, include Grand Bahama (530 square miles), which contains the major settlements of Freeport and West End; Andros (2,300 square miles), the largest island; Great Abaco (372 square miles); and Eleuthera, site of one of the early attempts at colonization. In spite of the concentration of the population in urban centres (especially Nassau and Freeport) devoted to tourism, the traditional pattern of small farming and fishing prevails in many villages, notably in the southeastern islands.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief and soils.* Extensive areas of flatland, generally a few feet in elevation, are the dominant topographic features of the major islands; Bimini, for example, has an elevation of only 20 feet (six metres). A number of islands fronting the Atlantic have a range or series of ranges of hills on their northeastern side and parallel to the longer axis of the island. These are formed of sand washed ashore and blown inland by the trade winds. The newer hills adjacent to the seashore are normally sand dunes. Solidity increases toward the interior, where the particles become cemented to form Bahama limestone. Eleuthera and Long Island have the greatest number of hills exceeding 100 feet, and the highest point, 206 feet (63 metres), is on Cat Island. Beneath the soil, the islands are composed of the skeletal remains of coral and other marine organisms. There are no rivers, but several islands—particularly New Providence, San Salvador, and Great Inagua—have large lakes.

Temperature and rainfall

Climate. The Bahamian climate, mild throughout the year, is one of the great attractions of the area. The average temperature varies from 70° F (21° C) during the winter to 81° F (27° C) during the summer, and extremes seldom fall below 60° F (16° C) or rise above 90° F (32° C). The average annual rainfall is about 44 inches (1,120 millimetres), occurring mostly during the summer months. Prevailing winds, coming from the northeast in winter and from the southeast in summer, lend a cooling influence to a generally humid atmosphere. Hurricanes present something of a threat during the period from mid-July to mid-November and have occasionally caused great destruction.

Plant and animal life. Extensive and beautiful forests of Caribbean pine are found on Grand Bahama, Great Abaco, Andros, and New Providence. Hardwood forests, known locally as "coppices," also occur on some of the islands. Elsewhere, the woody vegetation consists mostly of shrubs and low trees. Animal life is dominated by frogs, lizards, and snakes, all of them nonpoisonous, and several species of bats are found in caves along the more rocky coasts. Larger animals include the agouti, a rodent; the raccoon; the iguana; and the elegant flamingo, the national bird. All of these have been much reduced in numbers and in distribution. In addition, several animals—notably sheep, horses, and other livestock—have been introduced from Europe. The surrounding waters abound with fish and other edible marine animals, such as conch and crayfish.

Settlement patterns. The centres of population are widely distributed on each island. Some are located to the leeward, where it is calm and sheltered—for example, Cat Island—and others face the north and northeastern sides, where they are exposed to the northeast trade winds—as in the case of the Abaco Cays. Main settlements usually occur where there is a natural harbour or at least accessibility for shipping. There has been a marked shift of population from fishing and farming villages to the centres of tourist and commercial activity. Most of the population movement has been to New Providence, Grand Bahama, and Great Abaco.

Rural to urban migration

The people. A minority of the population is descended from English pioneer settlers and loyalist refugees. Most of the population is of African descent, many with varying amounts of Caucasian blood. There are also minorities of Greeks, Syrians, Haitians, and other West Indians. English is the only language native to Bahamians, although since the influx of Haitian immigrants, French or its creole dialect is spoken.

A high percentage of Bahamians are members of Christian churches. Most are Protestants, the largest denominations being Baptist, Church of God, Anglicans, and Methodists. Roman Catholics form a large minority.

About two-thirds of the Bahamian population is concentrated on New Providence Island, which, with Grand Bahama and Great Abaco, has received the most internal migration. The Bahamas' rate of population increase is much higher than the Caribbean average, primarily because of immigration from the United States and other West Indian islands. The rate of natural population increase is about equal to the Caribbean mean, but both the birth and death rates are less than the average for the West Indies as a whole.

The economy. The major economic activities that sustained The Bahamas until the mid-20th century were farming and fishing. During the latter part of the 19th century, sponge fishing and the production of pineapple and sisal were important. Sponge fishing was the major industry between the 1860s and 1938, but it declined and has been replaced by tourism. Banking has become the second most important industry, and The Bahamas is growing as an international financial centre. Crayfishing contributes significantly to the economy.

Manufacturing industries include food-processing plants (mainly tomato and pineapple canning and crayfish freezing), a petrochemical refinery and a pharmaceutical factory (mainly producing hormones). Freeport has become the most important centre of industrial development and is second only to Nassau in tourist and commercial activity. Salt is produced by solar evaporation at Inagua. Aragonite, a hard, granular limestone used in the manufacture of cement, steel, petrochemicals, glass, and other products, is mined near Sandy Cay. The government has encouraged the development of small industries, and various forms of food production continue at numerous places for domestic consumption. The Bahamas has an open shipping registry, and the merchant fleet sailing under its flag is one of the world's largest. There is no direct taxation, and customs duties produce the major share of revenue.

Transportation. Nassau and Freeport and their environs have modern paved roads, as do most of the inhabited islands. A fleet of small motor vessels carries passengers, freight, and mail weekly between Nassau and the Family

Islands. The deepwater harbours of Nassau and Freeport are dredged to depths of more than 30 feet. Numerous foreign passenger and freight ships visit the two ports each year. The airplane has become of increasing importance to the Bahamian economy. Throughout the islands there are some 60 airports, with varying accommodations and facilities. Most of these serve only interinsular aircraft; but international airports are located at Nassau and Freeport, and international flights also connect with the islands of Great Exuma, Great Abaco, and Eleuthera.

Administration and social conditions. *Government.* The government is patterned after that of the United Kingdom. Since 1973 the British monarch has appointed the governor-general, who in turn appoints a prime minister. The latter must be a member of the House of Assembly and must be able to command a majority of Assembly votes. The parliament consists of two houses. The House of Assembly is elected by the people; the Senate, which has severely restricted powers, is appointed by the governor, who selects the majority of the members on the advice of the prime minister. The life of the parliament is limited to a maximum of five years, but, if the prime minister is unable to control the Assembly effectively or if he considers it expedient, both bodies are dissolved and reconstituted. Judicial power on the islands resides in the Court of Appeal, the Supreme Court, and magistrates' courts.

Executive leadership

Education and health and welfare. There is little illiteracy in The Bahamas. Schooling is compulsory from five to 14 years of age. Public secondary and technical schools have been expanding. The College of The Bahamas, established in 1974 in Nassau, is a community college that offers programs in conjunction with other universities, including the University of the West Indies, Florida International University, and the University of Miami. Bahamians are relatively free of malnutrition and debilitating diseases, and medical problems among children are largely those involving common infections. Increasing alcohol and drug abuse has become a concern, and care for the aged is a mounting problem. Extensive government-sponsored housing developments have alleviated some of the housing problems.

Cultural life. Bahamian culture reflects the origins of its people and has also been influenced by the peoples of nearby islands. Outstanding among traditional group activities is the Junkanoo parade on Boxing Day and New Year's Day in Nassau. The main thoroughfare is given over to hundreds of gaily bedecked celebrants who, with clanging cowbells and beating drums, march and dance to a goombay rhythm of African origin. Island folklore includes stories of a three-toed, human-faced creature called the Chickcharney and the workings of Obeah, a brand of witchcraft. In Nassau, Bahamian drama and art have come into their own. There is a repertory season at the Dundas Centre for the Performing Arts that includes serious drama, musicals, and dancing. Various drama groups also present musicals and drama in the off-season. Works of art can be seen at the various galleries. The National Trust is concerned with the preservation of wildlife and historic buildings, and the Department of Archives preserves public and private records, making them accessible to the public. The Historical Society promotes local history.

Traditional folk activities

For statistical data on the land and people of The Bahamas, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

It is widely held that on Oct. 12, 1492, Christopher Columbus first landed on an island called by its native inhabitants Guanahani, which he renamed San Salvador. Its actual identity is still in dispute, but some scholars believe it is the place known today as Watling Island, while some others claim the first landfall to have been at Samana Cay, or Cat Island. Columbus explored the island and others nearby and then sailed to Cuba and Hispaniola. The natives of the Bahamas, whom Columbus called Lucayans, were Arawak Indians. They also inhabited the Greater Antilles and were peaceful. Between 1492 and 1508, Spanish raiders carried off about 40,000 natives to work in the mines of Hispaniola (Haiti and the Dominican Republic), and the islands remained depopulated more than a century before the first English settlement took place.

Although Columbus took formal possession of the islands with pomp and ceremony in the name of Spain, and under the Treaty of Tordesillas between Spain and Portugal in 1494 the islands were held to come within the Spanish sphere, the Spanish made little attempt to settle them.

British colonization. British interest began in 1629 when Charles I granted Sir Robert Heath, attorney general of England, territories in America including "Bahama and all other Isles and Islands lying southerly there or neare upon the foresayd continent." Heath made no effort to settle the Bahamas. In the 1640s Bermuda was troubled by religious disputes. In 1647 Captain William Sayle, who had twice been governor of Bermuda, took the leadership of an enterprise to seek an island upon which dissidents could worship as they pleased. In July of that year the Company of Eleutherian Adventurers was formed in London "for the Plantation of the Islands of Eleutheria, formerly called Buhama in America, and the Adjacent Islands." Sayle and about 70 prospective settlers, consisting of Bermuda Independents and some persons who had come from England, sailed from Bermuda for the Bahamas some time before October 1648. The place of their landing is uncertain, but modern belief is that they settled on Eleuthera, then known as Cigatoo. They had envisioned establishing a flourishing plantation colony, but unproductive soil, internal discord, and Spanish interference dashed their hopes. Some of the settlers, including Sayle, returned to Bermuda.

Colonization by Captain Sayle

New Providence was first settled in 1656 by a new group of Bermudans. In 1663 South Carolina was granted by Charles II to eight of his friends as lords proprietors, and they appointed Sayle as the first governor. Both Sayle and certain of those who had interested themselves in the settlement of New Providence independently drew the attention of the lords proprietors to the possibilities of the Bahama Islands, and in consequence the Duke of Albemarle and five others acquired a grant of the islands from Charles II in 1670, and they accepted nominal responsibility for the civil government. New Providence, with the largest population, became the seat of government.

The proprietors did not take a very active interest in the settlement or development of the islands, however, and they soon became a haven for pirates, whose depredations against Spanish ships provoked frequent and savage retaliatory raids. In 1671 they appointed John Wentworth as the first governor. Although elaborate instructions for the government of the colony were issued and a parliamentary system of government was instituted, the lot of both governors and settlers was far from easy. New Providence was often overrun by the Spaniards alone or in combination with the French, while any governor attempting to institute a semblance of law and order received short shrift from the settlers, who had found piracy the most lucrative profession. In 1684 the king himself intervened and required that a law be passed against the pirates, but apparently it had little effect.

Early in the 18th century, official representations were being made for direct crown control. The lords proprietors surrendered the civil and military government to the king in 1717 and leased the islands to Captain Woodes Rogers, whom the king commissioned as the first royal governor and charged with the responsibility of exterminating pirates and establishing more stable conditions. When he arrived in 1718, armed with a disciplined troop of soldiers, about 1,000 pirates surrendered and received the king's pardon, while eight of the unrepentant were hanged. His measures were so effective that in 1728 the colony was able to adopt the motto, "Expulsis piratis restituta commercia."

Restoration of law and order

In 1660 the present site of the capital was known as Charles Towne in honour of Charles II, but these early settlers saw fit to change the name to Nassau when William and Mary came to the throne; the German region Nassau was a holding of William's family. With the restoration of order following the establishment of the royal government, the settlers demanded an assembly. In 1729 Woodes Rogers, acting under authority from the crown, issued a proclamation summoning a representative assembly and

from then on, apart from the brief interruptions caused by foreign invasion, the government of the colony carried on in an orderly manner.

In 1776 Nassau was captured by the U.S. Navy, which was seeking supplies during the Revolutionary War; they evacuated after a few days. In May 1782 the colony surrendered to Spain and, though it was restored to Britain by the Peace of Versailles in January 1783, it was brilliantly recaptured in April by Colonel Andrew Devaux, a loyalist commander, before news of the treaty had been received. On the conclusion of the American Revolution many loyalists emigrated from the United States to the Bahamas under very favourable terms offered by the crown. Among the newcomers was Lord Dunmore, formerly governor of New York and of Virginia, who served as governor of the Bahamas from 1786 to 1797. The loyalists fled with their slaves to the islands, doubling the white population and trebling the black. The cotton plantations that they developed yielded well for a few years, but exhausted soil, insect pests, and, finally, abolition of slavery led to their ultimate collapse. In 1787 the proprietors surrendered their remaining rights for £12,000.

Early 19th-century efforts of the assembly to thwart the attempts of the executive to ameliorate the conditions of the slaves continued until the United Kingdom Abolition Act came into force in the colony on Aug. 1, 1834; full emancipation came in 1838. A legislative council was created by royal letters patent in 1841.

Considerable wealth poured into the islands as the result of blockade-running during the American Civil War and the handling of liquor during Prohibition in the 1920s in the United States. This activity made no lasting contribution to the islands, however, nor did it establish any firm economic base. Before and after these periods, many attempts were made to grow pineapples, citrus fruits, tobacco, tomatoes, and sisal for export, but despite initial promise, all failed. Sponge fishing also collapsed in 1938.

Establishment of tourism

After World War II, strenuous efforts to establish tourism as the basis of the economy were strikingly successful, transforming the economic and social structure of the islands.

Independence. Politically, Bahamians have had considerable control over their affairs since the first assembly in 1729. In May 1963 a conference was held in London to consider a new constitution for the islands. It was then agreed that the colony should have full internal self-government, the governor retaining reserved powers only for foreign affairs, defense, and internal security. The new constitution came into force on Jan. 7, 1964, and constitutional advances in 1969 brought the country to the verge of complete self-government.

Party politics emerged in 1953, when the Progressive Liberal Party (PLP) was formed by blacks to oppose the group in power, who in 1958 responded with a party of their own, the white-controlled United Bahamian Party (UBP). As the political battle progressed, the PLP raised the cry for majority rule. The climax came after the general elections of 1967, when the PLP under the leadership of Lynden Pindling was able to form a government with a slight majority. Elections that were held the following year gave that party 29 of the 38 Assembly seats. It is widely believed that the UBP was defeated mainly on the racial issue, but the accusation that that party had introduced criminal-controlled gambling and that some of its members had profited thereby had a telling effect.

In general the PLP advocated stricter government control of the economy, increasing Bahamian ownership of business enterprises and the replacement of foreign workers by Bahamians. Although the move toward self-government received bipartisan support, some factions advocated that total independence should come later than 1973, the target year of the PLP government. In 1969 the name of

Dominance of the PLP

Commonwealth of the Bahama Islands was adopted, but upon independence, on July 10, 1973, the official form became The Commonwealth of The Bahamas. The PLP maintained its position as the majority party after independence. The main opposition was formed by the Free National Movement established in 1972 through a merger of the UBP and alienated anti-independence PLP members. The government embarked on programs to improve economic development, increase the standard of living, and halt the rising unemployment rate. The Bahamas was admitted to the Caribbean Community (Caricom) in July 1983. Alleged drug trafficking in government became a major issue and threatened PLP power in the late 1980s.

(E.P.A./D.R.Ha./D.G.Ss.)

For later developments in the history of The Bahamas, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Cayman Islands

Cayman Islands is a British colony in the Caribbean Sea, comprising the islands of Grand Cayman, Little Cayman, and Cayman Brac, situated about 180 miles (290 kilometres) northwest of Jamaica. The islands are the outcroppings of a submarine mountain range that extends northeastward from Belize to Cuba. The total area is 102 square miles (264 square kilometres), and the capital is George Town, on Grand Cayman.

PHYSICAL AND HUMAN GEOGRAPHY

The land. The limestone islands are low-lying, though Cayman Brac has a central bluff that constitutes 90 percent of its landmass. The coasts are ironshore (limestone fringes with numerous marine fossils) interspersed with sandy beaches and enclosed by coral reefs. The largest island is Grand Cayman (76 square miles), which has a magnificent seven-mile beach and a 36-square-mile sound that is a breeding ground for much marine life. There are no rivers. The Caymans are cool from November to

Climate

March, with the temperatures ranging from 65° to 75° F (18° to 24° C). Rainfall at George Town averages 60 inches (1,524 millimetres) annually, although the eastern districts of Grand Cayman and the other islands are drier. The Caymans lie in the hurricane zone, and hurricanes can occur from June through November. Quite a few bird species inhabit the islands.

The people. About one-quarter of Caymanians are of European, mainly British, stock; another quarter are blacks, the descendants of African slaves; and about one-half of the population is of mixed ancestry. English is the official language. Most of the major Christian denominations are represented in the islands.

The economy. The physical beauty and superb climate of the islands have made them a haven for tourists. Airports on Grand Cayman and Cayman Brac and a private airstrip on Little Cayman facilitate tourism. International finance has become a major industry, owing to the absence of direct taxes and to liberal company and banking laws that generally ensure confidential transactions. More than 500 banks and trust companies, including most of the world's 50 largest banks, are registered in the Caymans; revenue paid by registered businesses contributes considerably to the government budget. There is a shortage of labour, and much of the work force is made up of immigrants on tightly controlled work permits.

Administration and social conditions. The 1972 constitution provides for internal autonomy under a governor, an executive council, and a legislative assembly. The governor is responsible for foreign affairs, defense, internal security, the police, and the civil service. Education is compulsory for all children between the ages of four years, nine months and 16.

Cultural life. The F.J. Harquail Theatre on Grand Cayman is the main venue for local and visiting companies. There is a daily newspaper, a government-owned radio station, and a monthly news magazine.

For statistical data on the land and people of the Cayman Islands, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

The Caymans were sighted by Christopher Columbus on May 10, 1503, during his last voyage to the West Indies. At first the Spaniards named them Las Tortugas because of the many turtles in the surrounding waters, but by 1530 they were known as the Caimanas or Caymanes

after the alligators (*caimánes*) reported to be native there. After the Treaty of Madrid that ceded Jamaica to Great Britain, the first permanent settlement was established on Grand Cayman. Most of the settlers were British mariners and privateers and shipwrecked passengers and African slaves, as well as land-grant holders from Jamaica. The remoteness of the islands and integration following the emancipation of slaves in 1835, resulted in a socially homogeneous society.

First permanent settlement

By the end of the 18th century uncontrolled fishing eliminated the native turtle population, virtually the only resource of the island. Cayman Islanders searched farther and farther away for new turtle grounds, but as international restrictions grew, turtle fishing was greatly reduced.

For some time the Cayman Islands were a dependency of Jamaica, becoming internally self-governing in July 1959. When Jamaica declared its independence (1962) the Caymans reverted to direct British rule. A new constitution providing for autonomy on most domestic issues was approved in 1972. The Caymans had by then developed offshore banking and tourism, enabling the colony to relinquish aid from Britain. (C.A.Wi./E.P.E./Ed.)

For later developments in the history of the Cayman Islands, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Netherlands Antilles

The Netherlands Antilles (Dutch: Nederlandse Antillen) are composed of two widely separated groups of islands, about 500 miles (800 kilometres) apart in the Caribbean Sea. Until 1986 Aruba was politically a part of the Netherlands Antilles, but in that year it obtained separate political status. The area of the islands is 308 square miles (800 square kilometres). They form an integral part of the Kingdom of The Netherlands and are fully autonomous in internal affairs. The southern group of islands, of which Aruba is geographically a part, includes Curaçao and Bonaire; they lie about 60 miles off the Venezuelan coast. The northern group includes Sint Eustatius, Saba, and the southern part of Saint Martin (called Sint Maarten) the other part being administered by the French territory of Guadeloupe; geographically this group lies within the Leeward Islands of the Lesser Antilles. The capital and largest city of the Netherlands Antilles is Willemstad on Curaçao.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The southern islands are generally low in altitude, though hills range from the 787-foot (240-metre) Mount Brandaris in Bonaire to the 1,230-foot Mount Saint Christoffel in Curaçao. The islands consist mainly of igneous rocks and are fringed with coral reefs. The northern islands consist of volcanic rocks rising to the 1,119-foot Sentry Hill in the Dutch part of Saint Martin, the 1,968-foot Quill, an extinct volcano on Sint Eustatius, with a large forested crater, and the 2,821-foot Mount Scenery on Saba. Curaçao, the largest island of the Netherlands Antilles (171 square miles), is indented in the south by deep bays, the largest of which, the Schottegat, provides a magnificent harbour for Willemstad. Bonaire, with an area of 111 square miles, lies about 20 miles east of Curaçao. Sint Eustatius (with an area of eight square miles), along with Saba, forms the northwestern termination of the inner volcanic arc of the Lesser Antilles. Saba (five square miles) is actually the peak of an extinct volcano, Mount Scenery, surrounded by sea cliffs. The villages of The Bottom and Windwardside, occupying an old crater, are approached up a steep road from a rocky landing place on the south coast.

Geologic formations

Drainage and soils. For the most part, the islands of the Netherlands Antilles have barren soil and little or no natural irrigation. On Curaçao and Bonaire there is much bare, eroded soil, the result of centuries of overgrazing. Drinking water on these islands is obtained mainly from distilled seawater.

Climate. The temperature of the southern islands varies little from an annual average of 81° F (27° C), and the heat is tempered by the easterly trade winds. The islands lie west of the usual hurricane zone. Rainfall is low and variable, often less than 20 inches (488 millimetres) a year. The climate is similar in the northern islands, but rainfall is greater and hurricanes occur. The annual rainfall is greatest on Sint Eustatius and Saba (42 inches and 47 inches, respectively) and falls mainly between May and November, occasionally in association with hurricanes.

Plant and animal life. The vegetation of the southern Netherlands Antilles, much overgrazed by animals, is sparse. Cacti and other drought-resistant plants abound. The island of Bonaire is known for its flamingos.

Settlement patterns. More than 90 percent of the population are urban dwellers. The rural population of the islands is generally dispersed, and villages are scarce except on Saba. Characteristic of Curaçao are its *landhuizen*, large 18th- and 19th-century rural mansions located on hills. The capital city of Willemstad has some splendid sections of Dutch-style colonial architecture with tropical adaptations and painted in white and pastel colours. More than 85 percent of the inhabitants of the islands reside on Curaçao; the next two most populous islands are Saint Martin and Bonaire. The most sparsely populated islands are Sint Eustatius and Saba.

Urban-rural distribution

The people. The islands have mostly black and mulatto populations except for Saba, which is about evenly divided between white and black. Most of the islands have small white minorities. Migration to Curaçao from other Caribbean islands, Venezuela, and Europe increased after the opening of the oil refinery in 1918.

Dutch is the official language, but Papiamento—a local Spanish-based creole that includes Portuguese, Dutch, and some African words—is widely used in the southern islands and is taught in elementary schools. English is the principal language of the northern islands, and Spanish is also spoken in the south. The major religion of the islands is Roman Catholicism, with small Lutheran (Reformed tradition), Methodist, Seventh-day Adventist, and Jewish minorities. For more than three centuries, a Sephardic Jewish community has lived in Curaçao.

The birth and death rates are relatively low, and the rate of natural increase is lower than on most other islands of the Caribbean. Migration to The Netherlands has increased.

The economy. Unlike most other Caribbean islands, the economy of the Netherlands Antilles never has been based on the export of sugar or other plantation crops, which the climatic conditions on the larger islands made impossible. Instead, Curaçao (and during the 18th century, Sint Eustatius) developed into a regional trading and financial centre, activities that, together with oil-refining and tourism, have become the basis of the islands' economy.

Resources. The southern islands have a greater abundance of exploitable minerals than the northern ones. Curaçao has some calcium phosphate mining; salt is processed on Bonaire.

Agriculture. Agriculture plays a minor role in the economy of the islands, although sugarcane and cotton plantations were once established on Saint Martin and Sint Eustatius. Curaçao was at one time used mainly for livestock raising, but, after the overgrazing of land, new small-scale agricultural ventures were begun, such as the cultivation of aloes for pharmaceutical products and oranges for Curaçao liqueur. Aloes are also grown on Bonaire. A fishery, established by the Japanese, is important to the economy of Sint Maarten. Saba is engaged chiefly in raising livestock and the cultivation of vegetables, particularly potatoes, which are exported to neighbouring islands.

Industry. The main industry of Curaçao is oil refining, which started with the opening up of the Venezuelan oil fields in 1914. After the oil refinery opened in 1918 on Curaçao, the industry became the economic mainstay of the islands. Bonaire has a textile factory, and Sint Maarten a rum distillery. Willemstad has become an important Caribbean banking centre. For all of the islands tourism and service industries have become increasingly important.

Oil refining

Trade. The main exports of the Netherlands Antilles are petroleum and petroleum products, all of which are produced on Curaçao. The entrepôt trade in the free ports of Curaçao is also significant. Curaçao's foreign trade is mainly with Venezuela, the United States, and The

Netherlands. Most of the islands' requirements of food and commercial goods are met by imports.

Transportation. Curaçao has an extensive road system and is linked to the outside world by Dutch, U.S., and Venezuelan airlines as well as by numerous steamship services. Sint Maarten also has an international airport.

Administration and social conditions. The Netherlands Antilles are a self-governing part of the Kingdom of The Netherlands. A governor, nominated by the local government and appointed by the crown, is the formal head of government and representative of the monarch. Executive authority is vested in a Council of Ministers headed by a prime minister. The council is responsible to the unicameral legislature (Staten), which is elected by universal suffrage. Although education in the Netherland Antilles is not compulsory, most of the population is literate. The main language of instruction is Dutch, and there is a university on Curaçao. The general standard of health on the islands is high.

Cultural life. Carnival time in February and the New Year's festivities are colourful celebrations. The Bonaire International Sailing Regatta is held every October, attracting boating enthusiasts from around the world. The 3,500-acre national park on Curaçao provides a broad cross section of the island's natural flora and fauna. Radio broadcasts are received throughout the islands, and there are television stations on Saint Martin and Curaçao.

For statistical data on the land and people of the Netherlands Antilles, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

The islands known as the Netherlands Antilles were visited in the early 16th century by the Spanish, whose activities caused the decimation of the native Indian population. The Dutch, attracted by salt deposits, occupied the islands in the early part of the 17th century, and, except for brief periods of British occupation, the islands have remained Dutch possessions. Through much of the 17th and 18th centuries the islands prospered from Dutch trade in slaves, plantation products, and contraband, but the economy declined from 1816 until 1914.

Colonial rule. *Curaçao.* Curaçao was discovered by Alonso de Ojeda and Amerigo Vespucci in 1499 and settled in 1527 by the Spanish, who used it mainly for livestock raising. In 1634 Johannes van Walbeeck of the Dutch West India Company occupied and fortified the island, which became the base for a rich entrepôt trade flourishing through the 18th century. During the colonial period Curaçao was a major centre of Caribbean slave trade.

Centre of Caribbean slave trade

There were two short periods during the Napoleonic Wars when Curaçao was held by the British, but it was returned to The Netherlands by the Treaty of Paris in 1815. The 19th century was a period of economic decline alleviated by cultivation of aloes and oranges. Only after the construction of the Schottegat oil refinery, however, did economic conditions greatly improve.

Bonaire. Bonaire was discovered in 1499 by Alonso de Ojeda and Amerigo Vespucci, settled by the Spanish in 1501, and claimed by the Dutch in 1634. It became part of the Dutch West India Company in 1636 and remained a government plantation until 1863. From 1807 to 1814 it was under British control.

Saint Martin. Discovered by Christopher Columbus on Nov. 11, 1493 (St. Martin's Day), the island was taken by French pirates in 1638, though the Spaniards settled there in 1640. In 1648, French and Dutch prisoners of war allegedly met after the Spanish departure and amicably divided the island. The Dutch obtained the smaller but more valuable section containing salt deposits.

Sint Eustatius. Sint Eustatius, first colonized by the French and English in 1625, was taken by the Dutch in 1632. It became the main centre of slave trade in the eastern Caribbean and by 1780 had a population of 2,500. In 1781 the British sacked Oranjestad (after the U.S. flag had officially been saluted there for the first time), and the island never regained its trade. In the 17th and 18th centuries most of the land was under sugar cultivation.

Saba. Saba was settled by the Dutch in 1632 but, because of its inaccessibility and ruggedness, never achieved any economic importance.

Political developments since World War II. After World War II, negotiations began with the aim of conferring a greater measure of self-government on the islands. On Dec. 15, 1954, a charter was signed making the islands an autonomous part of The Netherlands. In 1969, Curaçao was torn by labour conflicts leading to riots and arson. Since then discussions on complete independence have moved unevenly.

Politics in the Netherlands Antilles are now dominated by three issues: economic problems, the coming of independence, and the degree of autonomy each island should have within the federation. By the mid-1970s it was clear that most of the Netherlands Antilles feared the economic aftereffects of independence. The Dutch government pressed for independence but insisted on preserving a federated structure embracing all islands. In an unofficial referendum in 1977, Aruba voted to secede from the Antilles federation, yet to remain—as a separate entity—within the kingdom. By 1978 all six islands accepted the concept of insular self-determination. In 1983 the Dutch government agreed that by 1986 Aruba would achieve its separate status on condition that it would accept full independence by 1996.

Independence movement

In 1989 the political leadership of Sint Maarten announced its wish to achieve full independence in the shortest possible term. In the case of Sint Maarten, as in that of Aruba, secessionist feelings were fueled by animosity toward the central administration in Curaçao. Efforts to placate such feelings by increasing insular autonomy have not had the desired results. (D.R.Ha./H.Ho.)

For later developments in the history of the Netherlands Antilles, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Trinidad and Tobago

Trinidad and Tobago are two Caribbean islands that constitute an independent republic. Forming the two southernmost links in the Caribbean chain of islands, they lie close to the South American continent, northeast of Venezuela and northwest of Guyana. The country has a total area of 1,978 square miles (5,123 square kilometres). The capital of the republic, Port-of-Spain, is situated in the northwest of Trinidad.

Trinidad, the larger island, comprises 1,841 square miles. It is seven miles (11 kilometres) at the nearest point from the Venezuelan coast, from which it is separated by the Gulf of Paria and two narrow channels where there are several small islands and rocks. Tobago, the smaller island, with an area of about 115 square miles, lies in the Atlantic, 19 miles to the northeast of Trinidad. Lying diagonally from southwest to northeast, Tobago is about 32 miles long and is more than 11 miles across at its widest point. Little Tobago, also called Bird of Paradise Island, is a mile off Tobago's northeastern coast; it is noted as the only wild habitat outside of New Guinea for the greater bird of paradise.

Little Tobago

Trinidad and Tobago achieved independence from the United Kingdom in 1962 and subsequently obtained membership in the British Commonwealth.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief and drainage.* Physiographically, the islands represent an extension of the South American mainland. The outstanding physical feature of Trinidad is its Northern Range (a continuation of the coastal ranges of the Andes Mountains in Venezuela), which runs from west to east at an average elevation of 1,500 feet, rising to 3,084 feet (940 metres) at the Aripo Mountain, the nation's highest peak. The Northern Range is the site of many waterfalls, the most spectacular of which are the Blue Basin and the Maracas Falls, both 298 feet high. On the southern side of the range, foothills with an elevation of about 500 feet descend to the Northern Plain.

Running across the centre of the island, from southwest to northeast, is the Central Range, the highest point of

which is Mount Tamana (1,009 feet). A third row of mainly low hills, the Southern Range, adds further variety to the mostly flat or undulating surface of Trinidad.

These three mountain ranges determine the island's drainage pattern. Rivers are numerous but short, the longest being the Ortoire in the south (31 miles) and the Caroni in the north (25 miles). Swamps occupy parts of the low-lying areas, among them the Caroni Swamp in the northwest and clusters along the east and south coasts.

An oil-bearing belt occupies the southern quarter of the island, extending west into the Gulf of Paria and east into the Atlantic Ocean. Gas and water seepages give rise to mud volcanoes of various types, the best-known of which is the Devil's Woodyard. In the southwest of the island is the sedimentary volcano known as the Pitch Lake, which contains reserves of asphalt.

Topography

Tobago is physiographically a further extension of the Venezuelan coastal range and the Northern Range of Trinidad. Its dominant feature is the Main Ridge, which runs from northeast to southwest, rising to heights of about 1,800 feet. The ridge slopes more gently to the southwest onto a coral plain. The coral formation has given rise to a number of reefs, one of which, Buccoo Reef, is known for its marine life and is popular for scuba diving and snorkeling. The island has only a few short streams.

Climate. The climate of Trinidad and Tobago is tropical, with a high relative humidity. The coolest months are January and February, when the average minimum temperature is about 68° F (20° C). The warmest months are April, May, and October, which have an average maximum temperature of about 89° F (32° C). In general, mean temperatures range between 77° F (25° C) in February and 85° F (29° C) in April. Temperatures vary significantly between day and night, and the climate along the coast is tempered by sea breezes.

There is a main dry season from January to May and a lesser dry season (Petite Carême, or Indian Summer) in September and October. The prevailing winds are the northeast trades. The islands are outside the main hurricane zone, but Tobago has twice within 100 years—in 1867 and 1963—been struck by disastrous hurricanes.

Plant and animal life. Vegetation zones are well marked on both islands. In general, the highest areas coincide with the most luxuriant tropical rain forest vegetation. Cultivated estates or small settlements are established in clearings on the hills. In the dry season the hills are decorated with the orange flowers of the mountain immortelle, a large flowering tree that grows to a height of about 80 feet, and the gold flowers of the poui. Sugarcane, the main agricultural crop, is grown on Trinidad's Central Plain.

The Caroni Swamp, a bird sanctuary, is frequented by flocks of white flamingo and egret as well as the scarlet ibis—a national bird. The greater bird of paradise has been introduced to the island of Little Tobago. The forests on both islands are hunting grounds for small game, the most sought after being the paca, or *lappe*. Other animals include the agouti (a short-haired, short-eared, rabbitlike rodent), collared peccary, or *quenck* (a wild hog), tattoo (an armadillo), prehensile porcupine, and iguana (a large lizard). There are four main groups of reptiles: snakes, lizards, turtles, and crocodiles (one kind, the caiman, related to the alligator).

Settlement patterns. Soils, climate, and vegetation all have influenced the pattern of local settlement. Villages stretch ribbonlike along the major roadways. In Trinidad, though not in Tobago, villages are so diverse in plan that it would be difficult to call any typical.

Even in the sugar belt of the Central Plain, with its mainly, though not exclusively, East Indian population, patterns vary. Kinship tends to be the important structural element in the life of the East Indian village in Trinidad; caste may also have a localized influence. Religious festivals, such as Dīwālī (Festival of Lights) and various forms of puja (ceremonial offering), are important events. Houses vary in size and architecture from the simple thatched hut to the well-built two-storied dwelling, brightly painted and roofed with corrugated iron.

The East Indian culture

A somewhat different life-style prevails in villages inhabited by people predominantly of African descent, though many villages have both East Indian and African characteristics. Because of the conditioning of the slave system, traditional African culture has undergone considerable mutation or reinterpretation. The family unit is nuclear rather than extended and may be based upon marriage or upon a stable extralegal relationship.

These different rural cultural streams converge on the capital, Port-of-Spain. This city, with its mixed population, its Spanish influence (particularly in architecture), and French creole flavour, is one of the most cosmopolitan in the world. San Fernando, with its large East Indian population, is the second largest town and is located south of Port-of-Spain along the west coast. Arima is the oldest municipality in Trinidad. Scarborough, the chief town in Tobago, is an administrative centre and market town.

The people. The original inhabitants of Trinidad were chiefly Arawak. Although there are inhabitants of the town of Arima who claim descent from Carib royalty, there is doubt that the land was settled by Caribs. Tobago was frequently visited by Amerindians, probably both Arawak and Carib, but was not settled before the arrival of Columbus. Spanish, French, African, English, East Indian, and Chinese have all contributed to the ethnic composition of the islands' population. The various immigrant groups brought with them their languages, culture, and religion. Although English is the official language, four creole languages (Trinidad English, a French creole, a Spanish dialect creole, and Manzanillan—a melange of military English, Spanish, and several African tongues) and some East Indian languages also are spoken.

Under the Spanish, Roman Catholicism was the official religion, and it was strengthened by French immigration during the French and Haitian revolutions. Protestantism gained a foothold in various forms (Anglican, Methodist, Moravian, and Baptist) with the advent of the British. East Indians brought with them their languages and their Hindu and Muslim religions. Both Sunnite and Shī'ite Muslim groups are present. Further diversification followed with the immigration of Syrians and Lebanese. African-influenced religious groups include the Shango and Shouter cults.

Population growth

The first census of Trinidad and Tobago, in 1861, recorded a population of almost 100,000. By 1921 the population had more than tripled to some 360,000. Both the birth and death rates have remained fairly stable since the 1960s, and the rate of natural increase has been high. Emigration from the islands, however, has moderated the total population growth rate.

The economy. Principal features of the economy are a dependence on the petroleum industry and a fairly high level of unemployment. Petroleum continues to dominate the economy despite the oil glut of the 1980s. Tourism and manufacturing are significantly growing sectors of the economy.

Resources. Oil production is both land- and sea-based. Trinidad has extensive oil and gas fields, as well as deposits of asphalt, coal, gypsum, limestone, sand and gravel, iron ore, argillite, and fluorspar. Natural gas has been found off the coasts of Trinidad and Tobago.

Agriculture. Agriculture is a relatively small sector of the economy. The major agricultural export commodities are sugar, cocoa, and coffee. Other agricultural products include coconuts, citrus fruits, rice, copra, poultry, and vegetables.

Industry. Oil production and refining is the major industry. Government policy has encouraged economic diversification to reduce dependence on imports and on petroleum production. Industrial plants are engaged in the assembly of consumer durables, including motor vehicles and radio and television receivers, and in the manufacture of fertilizers, cement, paper products, furniture, garments, and processed foods. Government enterprises include chemical and fertilizer plants and a steel mill.

Tourism is potentially one of the fastest-growing industries. It is based particularly on Tobago and on Trinidad's northwest peninsula.

Transportation. The islands are served by a fairly well-developed network of main and local roads, but there is heavy congestion in urban areas. Several small shipping

lines and a domestic airline connect Tobago to Trinidad. Piarco International Airport on Trinidad and Crown Point Airport on Tobago have interisland connections and links with New York City, North and South America, western Europe, and the Caribbean. Port-of-Spain is the chief commercial port; petroleum exports are handled in ports in the south, such as Point Fortin, Pointe-à-Pierre, and Brighton. There are extensive port facilities at Point Lisas.

Administration and social conditions. *Government.* The first constitution, promulgated as a British Order in Council (1962), provided for a governor-general appointed by the British monarch, a Cabinet, and a bicameral legislature. Under the constitution adopted in 1976, Trinidad and Tobago is a republic headed by an elected president.

The 1976 constitution

The legislature consists of a 36-member House of Representatives, elected by universal adult suffrage every five years, and a 31-member Senate, whose members are appointed by the president on the advice of the prime minister and the minority party leader. The voting age is 18. Legislation passed in 1980 provided for a separate Tobago House of Assembly consisting of 12 members, elected at a primary election from electoral districts, who in turn elect three additional members as well as a chairman and deputy chairman. In January 1987 Tobago was granted full internal self-government. The legislation provides for a measure of devolution of executive powers in areas such as revenue raising and collection, agriculture, industry, tourism, environmental conservation, and social services. Trinidad is divided into 11 local government areas, including eight rural counties, which have county councils, and the three municipalities of Port-of-Spain, San Fernando, and Arima, which have municipal councils.

Education. Education is free at primary and secondary levels and compulsory between the ages of six and 12. There has, however, been a growing disparity between the number of school places and the growing number of school-age children, particularly at the secondary-school level. Faculties of the University of the West Indies—engineering, law, medicine, social science, natural science, education, agriculture, and liberal arts—are situated on the university campus at St. Augustine, about eight miles east of Port-of-Spain. There are technical institutes in Port-of-Spain, Centeno, and San Fernando, as well as several teacher-training colleges.

Health and welfare. The demand for additional housing in the urban areas has increased, and construction has been hampered by population movement, higher construction costs, shortage of land, and inadequate long-term financing. State provision for social security takes the form of noncontributory old-age pensions, noncontributory government employee pension schemes provided out of public revenues, and workers' compensation compulsorily paid by employers. A national insurance scheme has been established. The proportion of doctors in the population has remained low, principally as a result of the emigration of doctors, but the general health of the population has improved.

Cultural life. Trinidad has many writers of international stature, including C.L.R. James, Samuel Selvon, and V.S. Naipaul. The island is known for its steel band and calypso music and for the limbo. Derived from African music and dance forms, these are an important feature of the annual carnival celebration. Cricket is the most popular sport.

African musical influence

(A.N.R.R./Da.W.)

For statistical data on the land and people of Trinidad and Tobago, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

Christopher Columbus reached Trinidad in 1498 on his third voyage. The island had been inhabited since long before the Christian era by Amerindian peoples who came originally from the Orinoco delta region. At the time of Columbus' voyage, tribal groups speaking Arawakan languages were in the majority, but there were probably Carib speakers also. In the 16th century many of these Trinidadian Amerindians were captured by Spanish slavers and sent to work in other Spanish possessions, but there was no effective possession of the island by Spain until 1592. In that year, Antonio de Berrio took official possession of the island and founded St. Joseph, the capital and, for nearly 200 years, the only European settlement. Even after 1592 the development of the island proceeded very slowly. Few Spaniards emigrated to Trinidad, only a handful of African slaves were imported, and there was little production or export. In the 17th and early 18th centuries, tobacco and, later, cacao were cultivated using Amerindian labour, but after a disastrous cacao failure in the 1720s the industry declined. Lacking a real productive base, there was little trade or shipping, and the island remained undeveloped until the late 18th century.

Starting in 1776, the Spanish government encouraged Roman Catholics from the other Caribbean islands to settle in Trinidad with their slaves. This immigration, which became significant after the Cedula (decree) of 1783, which offered generous land and tax incentives to settlers, transformed Trinidad's population, economy, and society. Most of the settlers were French, and French influence became dominant. Many slaves were brought in from the other colonies and from Africa. Plantations were established, production of cotton and sugar began, and trade increased markedly. By 1797, when Britain seized the island from Spain, Trinidad had begun its development as a plantation economy and a slave society.

Under British rule (formal cession came in 1802), Trinidad's development as a sugar colony continued, although in 1806–07 the slave trade was completely prohibited. Slavery was abolished in two stages between 1834 and 1838, and the sugar planters were unable to secure the kind of steady, tractable, and cheap labour they wanted. In 1845 the immigration of indentured workers from India began; it continued until 1917. As early as 1870, about a quarter of the total population consisted of Indians. The original Amerindian inhabitants had by then virtually disappeared. Other immigrants came to Trinidad after 1838 from the smaller British Caribbean colonies, Africa (as free settlers rescued from foreign slave ships), Madeira, China, Syria, Lebanon, Venezuela, and the United Kingdom. Trinidad's population became the most heterogeneous in the Caribbean.

Decimation of Amerindians

Tobago, also sighted by Columbus in 1498, was virtually untouched by European settlement until the 18th century. Development as a sugar colony began when it was ceded to Great Britain in 1763 and continued throughout the period from 1763 to 1814, when Tobago changed hands between Great Britain and France several times. Tobago's sugar production peaked in the 1790s but began an irreversible decline after 1807. Great Britain was ceded Tobago for the last time in 1814, but by then its importance as a sugar-exporting colony had already begun to wane. Unlike Trinidad, Tobago had its own bicameral legislature until 1874. In 1889, with the island's economy in shambles as a result of the collapse of its sugar industry, Tobago was amalgamated with Trinidad, while retaining a subordinate legislature and separate taxes. In 1899 it became a ward (administrative district) of Trinidad and Tobago.

In 1925 a constitutional reform added seven elected members to Trinidad and Tobago's Legislative Council. Further agitation, and especially an islandwide series of strikes and riots in 1937 under Uriah Butler, led to the grant of universal suffrage in 1945 and other constitutional reforms granting a measure of self-government. For about 10 years after universal suffrage, politics in the colony were characterized by individualism and confusion, but in 1956 the People's National Movement (PNM) won a victory at the polls and formed the first party-based cabinet government under Eric Williams as chief minister. In 1962 Trinidad and Tobago attained independence and in 1976 became a republic within the Commonwealth.

The PNM won six consecutive elections and held power from 1956 to 1986, under the leadership of Williams (prime minister from 1962) until his death in 1981 and then under George M. Chambers. This continuity and stability in government were accompanied by economic problems and social unrest, culminating in widespread disturbances in 1970–71. But the oil boom in 1973–81 brought sudden prosperity to most sections of the population, and Trinidad and Tobago entered a period

of rapid development and industrialization. A substantial state sector and fairly comprehensive social welfare programs were created from the petroleum profits, while the private sector expanded rapidly. The collapse of oil prices, along with the PNM's failure to win support from most Indo-Trinidadians and deep-seated corruption, led to a marked decline in the party's popularity, especially after Williams' death.

In December 1986 the PNM was decisively defeated. The National Alliance for Reconstruction (NAR), a coalition party, won 33 of 36 seats on a program calling for divestment of most state-owned companies, reorganization of the civil service, and structural readjustment of the economy in the light of shrinking oil revenues. Arthur N.R. Robinson, NAR's political leader, became prime minister. (B.M.B.)

For later developments in the history of Trinidad and Tobago, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Turks and Caicos Islands

The Turks and Caicos Islands, a British colony in the West Indies, comprises two groups of islands lying on the southeastern periphery of The Bahamas, of which they form a physical part, and north of the island of Hispaniola. The islands include eight large cays (keys) and many smaller cays, islets, reefs, banks, and rocks that have a combined area of 166 square miles (430 square kilometres). The Turks group is composed of Grand Turk, Salt Cay, and lesser cays. The Caicos group, lying northwest of the Turks and separated from them by a 22-mile (35-kilometre) deepwater channel, the Turks Island Passage, consists of six principal islands—South Caicos, East Caicos, Middle (or Grand) Caicos, North Caicos, Providenciales, and West Caicos—and several cays. Only six of the larger cays and two of the smaller cays are inhabited. More than 80 percent of the population lives on Grand Turk (which includes Cockburn Town—the seat of government and main commercial centre), South Caicos, and Providenciales. The name Turks allegedly derives from a species of cactus whose scarlet flowers resemble a Turkish fez; that of Caicos perhaps from *cay icoco* (a coco plum tree).

Name derivation

PHYSICAL AND HUMAN GEOGRAPHY

The land. The islands are low-lying coralline limestone precipitated from the ocean and have numerous Karst features, including banana holes (small sinkholes containing rich soil), caves, caverns, and sea cliffs. Arable soil acreage is limited. Aragonite exists on the shallow banks off West Caicos. The highest elevation is 163 feet (50 metres), on Providenciales.

The climate is tropical savanna. Winter temperatures average between 75° and 80° F (24° and 27° C), summer temperatures between 85° and 90° F (29° and 32° C). The easterly trade winds moderate the climate. Rainfall averages about 29 inches (736 millimetres) annually at Grand Turk, and drinking water is in short supply. These are the driest islands in the Bahamas chain. The hurricane season is May through November. Hurricanes strike about every 10 years.

Types of vegetation include xerophytic shrub-scrub, coppice, savanna, and marsh-swamp. Mangroves, cacti, and Caribbean pines are found, and beefwood trees (*Casuarina*) have been planted as a windbreak. Terrestrial animal life consists mostly of insects (especially butterflies and mosquitoes), lizards (notably iguanas), and birds (notably flamingos); the islands are on migratory bird routes.

The people. More than 90 percent of the population are blacks. Native-born individuals are called "Belongers." About 14,000 Turks and Caicos citizens live in The Bahamas, where there is greater economic opportunity. The main religious denominations are Baptist, Methodist, and Anglican. English is the official language.

The economy. Lack of arable land restricts agriculture on the islands, though corn (maize), beans, cassava, and subsistence crops (staples and fruits) are grown on the western Caicos Islands. Rough grazing of beef cattle is the dominant land use, although much land is unused. Seafood is the major source of protein. Fishing—for lobster, conch, jacks, snapper, and others—and boatbuilding are traditional industries.

The islands have a food deficit, though lobster and conch are exported. Offshore financial services doing business with more than 7,000 registered companies also contribute to the economy.

Offshore banking

Three international airports and several smaller airstrips accommodate tourism, upon which the economy is heavily reliant. Grand Turk and Cockburn Harbour on South Caicos are main ports. Newer port facilities have opened on Providenciales, the main island for tourism.

Administration and social conditions. The constitution of 1976 provides for internal democratic self-government with an appointed executive council and an elected legislative assembly, over which a governor presides. Education is free and compulsory for children ages five to 15. Grand Turk has a hospital, and there are health clinics on several of the islands.

Cultural life. Water sports—sailing, big game fishing, and, especially, scuba diving among the coral reefs—are popular and attract tourists to the islands. Radio and satellite-cable television are available. Publications include the *Turks and Caicos News* (weekly) and *Conch News* (weekly).

For statistical data on the land and people of the Turks and Caicos Islands, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

A pre-Colombian Indian culture (Arawak perhaps) existed on the islands, but they were probably uninhabited when discovered by Juan Ponce de León, the Spanish explorer, in 1512. They remained mostly unsettled by Europeans until 1678, when Bermudans arrived and established a solar salt industry. The Caicos Islands were settled by royalist sympathizers from the United States after the War of Independence; they established cotton plantations employing slaves.

In 1799 the islands were annexed by the Bahama Islands government, but in 1848 they were granted a separate charter. In the meantime slavery had been abolished (1833–43), and the plantation owners left the islands, their former slaves remaining in possession.

Annexation by the Bahamas

After a period of financial difficulties, the colony was placed under the authority of the British governor-general at Kingston, Jam. (1874–1959), because ships voyaging between England and Jamaica passed the Turks and Caicos and made communication much easier than with Nassau in The Bahamas. The islands became a crown colony in 1962 when Jamaica became independent. For a time in the 1960s and '70s the islands were under the control of the Bahama Islands, but with Bahamian independence (1973) the Turks and Caicos were placed under a British governor at Grand Turk. As preparation for independence a commission was appointed in the 1980s to make recommendations on a new constitution and to consider the future economic direction of the islands. (J.H.Bo./Ed.)

For later developments in the history of the Turks and Caicos Islands, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

BIBLIOGRAPHY

The region: General information on the West Indies and on the countries of the region is found in *South America, Central America, and the Caribbean* (biennial), a reference work from the Europa Publications series; *The South American Handbook*, ed. by JOHN BROOKS (annual); and *The Caribbean Handbook*, ed. by JEREMY TAYLOR (annual). DAVID WATTS, *The West Indies: Patterns of Development, Culture, and Environmental Change Since 1492* (1987), offers an authoritative historical geography that focuses on the human impact on the natural environment. Detailed geographic analysis is provided in ROBERT C. WEST and JOHN P. AUGELLI, *Middle America: Its Lands and Peoples*, 3rd ed. (1989). Guides to the West Indies include *Birnbaum's The Caribbean, Bermuda, and the Bahamas* (annual); and EUGENE H. KAPLAN, *A Field Guide to Southeastern and Caribbean Seashores* (1988), and *A Field Guide to Coral Reefs of the Caribbean and Florida* (1982), the last two focusing on the fauna of the region. JAMES BOND, *Birds of the West Indies*, 5th ed. (1985), is a well-illustrated reference work.

Studies of the human population of the region include SIDNEY W. MINTZ and SALLY PRICE (eds.), *Focus—Caribbean,* 11 issues (1984), a collection of essays on topics ranging from race and colour to language and emigration; DAVID LOWENTHAL, *West Indian Societies* (1972), focusing on social structure and social conditions; AARON LEE SEGAL (ed.), *Population Policies in the Caribbean* (1975); and H. HOETINK, *The Two Variants in Caribbean Race Relations: A Contribution to the Sociology of Segmented Societies* (1967; originally published in Dutch, 1962; reissued as *Caribbean Race Relations: A Study of Two Variants,* 1971). The society of the region is explored in M.G. SMITH, *The Plural Society in the British West Indies* (1965, reprinted 1974), and *Culture, Race, and Class in the Commonwealth Caribbean* (1984). See also FRANKLIN W. KNIGHT, *The Caribbean, the Genesis of a Fragmented Nationalism* (1978); FRANKLIN W. KNIGHT and COLIN PALMER (eds.), *The Modern Caribbean* (1989); Race, class, education, and population growth are considered in MALCOLM CROSS, *Urbanization and Urban Growth in the Caribbean: An Essay on Social Change in Dependent Societies* (1979). The roots of Caribbean culture are explored in MARGARET E. CRAHAN and FRANKLIN W. KNIGHT (eds.), *Africa and the Caribbean: The Legacies of a Link* (1979). PETER A. ROBERTS, *West Indians and Their Language* (1988), is an informed synthesis of research findings.

Economic analyses that address the history of the one-crop economy and the changes in attitude toward land since the emancipation of the slaves include JEAN BESSON and JANET MOMSEN (eds.), *Land and Development in the Caribbean* (1987); JOHN S. BRIERLEY and HYMIE RUBENSTEIN (eds.), *Small Farming and Peasant Resources in the Caribbean* (1988); and MALCOLM CROSS and ARNAUD MARKS (eds.), *Peasants, Plantations, and Rural Communities in the Caribbean* (1979). Works on the modern economies include ANTHONY PAYNE, *The Politics of the Caribbean Community, 1961–79: Regional Integration Among New States* (1980), on the efforts at economic integration; CARIBBEAN COMMON MARKET, *The Caribbean Community in the 1980s* (1981), a review report on economic progress; DELISLE WORRELL, *Small Island Economies: Structure and Performance in the English-Speaking Caribbean Since 1970* (1987), a study of postindependence developments; CLIVE Y. THOMAS, *The Poor and the Powerless: Economic Policy and Change in the Caribbean* (1988), focusing on the analysis of nationalist economic models; and CAREY R. D'AVINO, *Caribbean Tax Havens* (1986), a specially commissioned report on the tax advantages for foreign investments, a specialty of the Caribbean economy.

Many histories of the region explore slavery and its origins and consequences: CARL BRIDENBAUGH and ROBERTA BRIDENBAUGH, *No Peace Beyond the Line: The English in the Caribbean, 1624–1690* (1972); RICHARD S. DUNN, *Sugar and Slaves: The Rise of the Planter Class in the English West Indies, 1624–1713* (1972); RICHARD B. SHERIDAN, *Sugar and Slavery: An Economic History of the British West Indies, 1623–1775* (1974); B.W. HIGMAN, *Slave Populations of the British Caribbean, 1807–1834* (1984); MICHAEL CRATON, *Testing the Chains: Resistance to Slavery in the British West Indies* (1982); WILLIAM A. GREEN, *British Slave Emancipation: The Sugar Colonies and the Great Experiment, 1830–1865* (1976); MANUEL MORENO FRAGINALS, FRANK MOYA PONS, and STANLEY L. ENGERMAN (eds.), *Between Slavery and Free Labor: The Spanish-Speaking Caribbean in the Nineteenth Century* (1985); R.W. BEACHEY, *The British West Indies Sugar Industry in the Late 19th Century* (1957, reprinted 1978); and MALCOLM CROSS and GAD HEUMAN (eds.), *Labour in the Caribbean: From Emancipation to Independence* (1988). The rise of the post-emancipation peasantry is explored in SIDNEY W. MINTZ, *Caribbean Transformations* (1974, reissued 1984); GORDON K. LEWIS, *The Growth of the Modern West Indies* (1968), is a general history. A path-breaking work relating to dependency is ERIC WILLIAMS, *Capitalism & Slavery* (1944, reissued 1983); economic dependency on foreign countries is developed in ANTHONY PAYNE and PAUL SUTTON (eds.), *Dependency Under Challenge: The Political Economy of the Commonwealth Caribbean* (1984). The history of political and ideological conflict in the contemporary West Indies is told in ANTHONY PAYNE, *The International Crisis in the Caribbean* (1984); PAUL SUTTON, *Dual Legacies in the Contemporary Caribbean: Continuing Aspects of British and French Dominion* (1986); ROBERT D. CRASSWELLER, *The Caribbean Community: Changing Societies and U.S. Policy* (1972); and VIRGINIA R. DOMÍNGUEZ and JORGE I. DOMÍNGUEZ, *The Caribbean, Its Implications for the United States* (1981).

Greater Antilles: (Cuba): Basic information is presented in *Cuba, a Country Study,* 3rd ed., edited by JAMES D. RUDOLPH (1985), which includes an extensive bibliography. ANTONIO NÚÑEZ JIMÉNEZ, *Geografía de Cuba,* 3 vol. (1972), provides a survey. SERGIO G. ROCA (ed.), *Socialist Cuba: Past Interpretations and Future Challenges* (1988), explores, among other topics, demography, religion, and race relations. Problems of population migration and employment are studied in JOSÉ LUIS LUZÓN, *Economía, población y territorio en Cuba, 1899–1983* (1987). Historical aspects of economic development are explored in LEVÍ MARRERO, *Cuba, economía y sociedad* (1972–), with 14 volumes published by 1989. Studies on contemporary economic conditions include CARMELO MESA-LAGO, *The Economy of Socialist Cuba: A Two-Decade Appraisal* (1981); CLAES BRUNDENIUS, *Revolutionary Cuba, the Challenge of Economic Growth with Equity* (1984); JORGE F. PÉREZ-LÓPEZ, *Measuring Cuban Economic Performance* (1987); and ANDREW ZIMBALIST (ed.), *Cuba's Socialist Economy Toward the 1990s* (1987). Cultural developments and policies are surveyed in ARMANDO HART DÁVALOS and LUIS BÁEZ, *Changing the Rules of the Game* (1983; originally published in Spanish, 1983). ALEJO CARPENTIER, *Ese músico que llevo dentro 3: la música en Cuba* (1987), explores Cuban music; and MICHAEL CHANAN, *The Cuban Image: Cinema and Cultural Politics in Cuba* (1985), discusses the political aspects of filmmaking in the country.

JAIME SUCHLICKI, *Historical Dictionary of Cuba* (1988), provides comprehensive information on persons, places, and events from 1492 to 1985 and includes a bibliography. HUGH THOMAS, *Cuba: The Pursuit of Freedom* (1971), is a massive history. Detailed examinations of the historical role of sugar growing and slavery include MANUEL MORENO FRAGINALS, *The Sugarmill: The Socioeconomic Complex of Sugar in Cuba, 1760–1860* (1976; originally published in Spanish, 1964); FRANKLIN W. KNIGHT, *Slave Society in Cuba During the Nineteenth Century* (1970); and REBECCA J. SCOTT, *Slave Emancipation in Cuba: The Transition to Free Labor, 1860–1899* (1985). Excellent treatments embracing the achievements of modern historiography include LOUIS A. PÉREZ, JR., *Cuba Between Empires, 1878–1902* (1983), *Cuba Under the Platt Amendment, 1902–1934* (1986), and *Cuba: Between Reform and Revolution* (1988). Studies of the 20th century, with special emphasis on Castro and the revolution, include CARMELO MESA-LAGO (ed.), *Revolutionary Change in Cuba* (1971); JORGE I. DOMÍNGUEZ, *Cuba: Order and Revolution* (1978); CARMELO MESA-LAGO, *Cuba in the 1970s: Pragmatism and Institutionalization,* rev. ed. (1978); TAD SZULC, *Fidel: A Critical Portrait* (1986); and MAX AZICRI, *Cuba: Politics, Economics, and Society* (1988). The role of revolutionary Cuba in the world is assessed in WAYNE S. SMITH, *The Closest of Enemies: A Personal and Diplomatic Account of U.S.-Cuban Relations Since 1957* (1987); BARRY B. LEVINE (ed.), *The New Cuban Presence in the Caribbean* (1983); COLE BLASIER and CARMELO MESA-LAGO (ed.), *Cuba in the World* (1979); and JORGE I. DOMÍNGUEZ (ed.), *Cuba: Internal and International Affairs* (1982).

(Dominican Republic): Comprehensive treatment is found in THOMAS E. WEIL *et al., Area Handbook for the Dominican Republic,* 2nd ed. (1973, reissued as *Dominican Republic, a Country Study,* 1982). Other general surveys include IAN BELL, *The Dominican Republic* (1981); and HOWARD J. WIARDA and MICHAEL J. KRYZANEK, *The Dominican Republic, a Caribbean Crucible* (1982), a discussion of the land, people, economy, and politics in a contemporary and historical setting. A more detailed study of the population and its social and economic situation is found in H. HOETINK, *The Dominican People, 1850–1900: Notes for a Historical Sociology* (1982; originally published in Spanish, 1971). Later social and economic developments are studied in KENNETH EVAN SHARPE, *Peasant Politics: Struggle in a Dominican Village* (1977), which discusses local-level politics and the international economic system; and JOSÉ A. MORENO, *Barrios in Arms: Revolution in Santo Domingo* (1970), a sociologist's eyewitness account of the impact of the 1965 revolution on inner-city Santo Domingo. An analysis of the impact of population migration is presented in GLENN L. HENDRICKS, *The Dominican Diaspora: From the Dominican Republic to New York City—Villagers in Transition* (1974).

ROBERT D. CRASSWELLER, *Trujillo: The Life and Times of a Caribbean Dictator* (1966), though a biography, serves as a good introduction to 20th-century history. JOHN BARTLOW MARTIN, *Overtaken by Events: The Dominican Crisis from the Fall of Trujillo to the Civil War* (1966), is the story of the 1965 revolution written by the former U.S. ambassador; while ABRAHAM E. LOWENTHAL, *The Dominican Intervention* (1972), is a historian's analysis of the same events. Interpretive studies of postrevolutionary history include HOWARD J. WIARDA, *Dictatorship, Development, and Disintegration: Politics and Social Change in the Dominican Republic,* 3 vol. (1975); and MICHAEL J. KRYZANEK and HOWARD J. WIARDA, *The Politics of External Influence in the Dominican Republic* (1988).

(Haiti): THOMAS E. WEIL *et al., Area Handbook for Haiti* (1973, reissued as *Haiti, a Country Study,* 1986), provides a geographic summary. GUY LASSERRE, PAUL MORAL, and PIERRE USSELMANN (eds.), *Atlas d'Haiti* (1985), is a collection of maps prepared by an international team of specialists. An important source on the geography of the colony of Saint-Domingue is MÉDÉRIC-LOIS-ELIE MOREAU DE SAINT-MÉRY, *A Civilization That Perished: The Last Years of White Colonial Rule in Haiti,*

ed. by IVOR D. SPENCER (1985; originally published in French, 2 vol., 1797–98). For views of Haitian society, see JAMES G. LEYBURN, *The Haitian People,* rev. ed. (1966, reprinted 1980); PAUL MORAL, *Le Paysan haïtien: étude sur la vie rurale en Haïti* (1961, reprinted 1978); MATS LUNDAHL, *Peasants and Poverty: A Study of Haiti* (1979); ANDRÉ-MARCEL D'ANS, *Haiti, paysage et société* (1987); CHARLES R. FOSTER and ALBERT VALDMAN, *Haiti Today and Tomorrow: An Interdisciplinary Study* (1984); and DAVID NICHOLLS, *From Dessalines to Duvalier: Race, Colour, and National Independence in Haiti* (1979, reissued 1988). Economic conditions in Haiti are analyzed in MATS LUNDAHL, *The Haitian Economy: Man, Land, and Markets* (1983); SIMON M. FASS, *Political Economy in Haiti: The Drama of Survival* (1988); and CHRISTIAN A. GIRAULT, *Le Commerce du café en Haïti: habitants, spéculateurs et exportateurs* (1981), a discussion of the coffee trade. For culture, see HAROLD COURLANDER, *The Drum and the Hoe: Life and Lore of the Haitian People* (1960, reprinted 1985); and J. MICHAEL DASH, *Literature and Ideology in Haiti, 1915–1961* (1981).

RAYFORD W. LOGAN, *Haiti and the Dominican Republic* (1968), is a standard history; BRIAN WEINSTEIN and AARON SEGAL, *Haiti: Political Failures, Cultural Successes* (1984), covers the developments of the 1970s; and ALEX DUPUY, *Haiti in the World Economy: Class, Race, and Underdevelopment Since 1700* (1989), is a perceptive political history with an economic focus.

(*Jamaica*): A general work is IRVING KAPLAN *et al.*, *Area Handbook for Jamaica* (1976). For physical geography, see R.M. BENT and ENID L. BENT-GOLDING, *A Complete Geography of Jamaica* (1966), an illustrated survey; and BARRY FLOYD, *Jamaica: An Island Microcosm* (1979). Reference information is found in COLIN G. CLARKE, *Jamaica in Maps: Graphic Perspectives of a Developing Country* (1974). A discussion of Jamaican geology is included in ALAN G. FINCHAM, *Jamaica Underground: A Register of Data Regarding the Caves, Sinkholes, and Underground Rivers of the Island* (1977). A brief review of both physical and social geography is MIKE MORRISSEY, *Our Island, Jamaica* (1983). See also L. ALAN EYRE, *Geographic Aspects of Population Dynamics in Jamaica* (1972). MERVYN C. ALLEYNE, *Roots of Jamaican Culture* (1988), focuses on African influences in Jamaican culture.

CLINTON V. BLACK, *The History of Jamaica,* new ed. (1983), offers a general outline. The material on the Jamaican Arawak is scarce, but much of the literature is summarized in the opening chapters of FRANCIS J. OSBORNE, *History of the Catholic Church in Jamaica* (1977, reissued 1988). On the Spanish period, FRANCISCO MORALES PADRÓN, *Jamaica española* (1952), remains the most substantial work. S.A.G. TAYLOR, *The Western Design: An Account of Cromwell's Expedition to the Caribbean* (1965), discusses the English invasion and early settlement. The plantation as an institution is explored in MICHAEL CRATON, *Searching for the Invisible Man: Slaves and Plantation Life in Jamaica* (1978). MAVIS C. CAMPBELL, *The Maroons of Jamaica, 1655–1796: A History of Resistance, Collaboration & Betrayal* (1988), focuses on activities and communities in the 18th century. For later periods, see EDWARD BRATHWAITE, *The Development of Creole Society in Jamaica, 1770–1820* (1971); B.W. HIGMAN, *Slave Population and Economy in Jamaica, 1807–1834* (1976); DOUGLAS HALL, *Free Jamaica, 1838–1865: An Economic History* (1959); PHILIP CURTIN, *Two Jamaicas* (1955, reissued 1970); MAVIS C. CAMPBELL, *The Dynamics of Change in Slave Society* (1976); and GAD HEUMAN, *Between Black and White: Race, Politics, and the Free Coloreds in Jamaica, 1792–1865* (1981), the last two focusing on the transition from slavery to emancipation. SYDNEY H. OLIVIER (LORD OLIVIER), *Jamaica, the Blessed Island* (1936, reissued 1971), provides a view of the penultimate stage of British colonial rule. Sociopolitical and economic changes of the 20th century are analyzed in EVELYNE HUBER STEPHENS and JOHN D. STEPHENS, *Democratic Socialism in Jamaica: The Political Movement and Social Transformation in Dependent Capitalism* (1986); and ANTHONY PAYNE, *Politics in Jamaica* (1988).

(*Puerto Rico*): RAFAEL PICÓ, *The Geography of Puerto Rico* (1974), offers a comprehensive survey. Anthropological, ethnological, demographic, and cultural features are discussed in JOSEPH P. FITZPATRICK, "Puerto Ricans," pp. 858–867 in STEPHAN THERNSTROM (ed.), *Harvard Encyclopedia of American Ethnic Groups* (1980); JULIAN H. STEWARD *et al.*, *The People of Puerto Rico: A Study in Social Anthropology* (1956); MARÍA TERESA BABÍN, *The Puerto Ricans' Spirit: Their History, Life, and Culture* (1971; originally published in Spanish, 1970); and STAN STEINER, *The Islands: The Worlds of the Puerto Ricans* (1974). For the country's economy, see ROBERT J. TATA and DAVID R. LEE, "Puerto Rico: Dilemmas of Growth," *Focus* 28(2):1–10 (November–December 1977); UNITED STATES. DEPARTMENT OF COMMERCE, *Economic Study of Puerto Rico,* 2 vol. (1979); and JAMES L. DIETZ, *Economic History of Puerto Rico: Institutional Change and Capitalist Development* (1986).

Comprehensive histories are provided by KAL WAGENHEIM, *Puerto Rico: A Profile,* 2nd ed. (1975); and KAL WAGENHEIM and OLGA JIMENEZ DE WAGENHEIM (eds.), *The Puerto Ricans: A Documentary History* (1973). Other historical studies include ARTURO MORALES CARRIÓN, *Puerto Rico and the Non-Hispanic Caribbean: A Study in the Decline of Spanish Exclusivism,* 2nd ed. (1971), with an analysis of the development of trade and commerce, and *Puerto Rico, a Political and Cultural History* (1983); HENRY WELLS, *The Modernization of Puerto Rico: A Political Study of Changing Values and Institutions* (1969); THOMAS G. MATHEWS, *Puerto Rican Politics and the New Deal* (1960, reprinted 1976); and, on the contemporary situation, RAYMOND CARR, *Puerto Rico, a Colonial Experiment* (1984).

Lesser Antilles: There are few works that treat all the islands of the Lesser Antilles or describe a particular island comprehensively, although a number of broad overviews are listed in the earlier section on the region. An informative geologic survey, covering locations from both the Lesser Antilles and the Netherlands Antilles, is offered in J.H. WESTERMANN and H. KIEL, *The Geology of Saba and St. Eustatius, with Notes on the Geology of St. Kitts, Nevis, and Montserrat, Lesser Antilles* (1961). GUY LASSERRE, *La Guadeloupe: Étude géographique,* 3 vol. (1978), is a detailed geography. Studies of flora include CLARISSA THÉRÈSE KIMBER, *Martinique Revisited: The Changing Plant Geographies of a West Indian Island* (1988); and DAVID WATTS, *Man's Influence on the Vegetation of Barbados, 1627 to 1800* (1966).

The people of Barbados are discussed in JILL SHEPPARD, *The "Redlegs" of Barbados, Their Origins and History* (1977), which explores the history of indentured servants; FARLEY BRATHWAITE (ed.), *The Elderly in Barbados* (1986), a survey of social and economic conditions of the elderly; and GRAHAM M.S. DANN (ed.), *Everyday in Barbados: A Sociological Perspective* (1976), which discusses social structures and recreational activity. JEAN BENOIST (ed.), *L'Archipel inachevé: culture et société aux Antilles françaises* (1972), is an anthropological study of the French islands of Martinique, Guadeloupe, La Désirade, Marie-Galante, and Saint-Barthélemy. STUART B. PHILPOTT, *West Indian Migration: The Montserrat Case* (1973), explores the impact of migration on village population. BONHAM C. RICHARDSON, *Caribbean Migrants: Environment and Human Survival on St. Kitts and Nevis* (1983), focuses on migration as a response to degradation of environment. KAREN FOG OLWIG, *Cultural Adaptation and Resistance on St. John: Three Centuries of Afro-Caribbean Life* (1985), examines the society of one of the Virgin Islands.

Analyses of economic conditions include DELISLE WORRELL (ed.), *The Economy of Barbados, 1946–1980* (1982), a study of the trends of the major sectors; BONHAM C. RICHARDSON, *Panama Money in Barbados, 1900–1920* (1985), which discusses the impact of remittances on a wide range of economic activities and social attitudes; MICHEL-ROLPH TROUILLOT, *Peasants and Capital: Dominica in the World Economy* (1988), which explores patterns of land ownership and agricultural production; and C. BOURNE, E.R. LEFRANC, and F. NUNES (compilers), *Small Farming in the Less Developed Countries of the Commonwealth Caribbean* (1980), which provides information on Grenada, Saint Vincent and the Grenadines, Saint Lucia, Dominica, Saint Kitts and Nevis, Montserrat, and Antigua. Studies of individual island-state economies include JOHN S. BRIERLEY, *Small Farming in Grenada, West Indies* (1974); DEIRDRE M. KELLY, *Hard Work, Hard Choices: A Survey of Women in St. Lucia's Export-Oriented Electronics Factories* (1987); and HYMIE RUBENSTEIN, *Coping With Poverty: Adaptive Strategies in a Caribbean Village* (1987).

Historical works which concentrate mostly on slavery and plantation life include the following: VINCENT T. HARLOW, *A History of Barbados, 1625–1685* (1926, reprinted 1969), an examination of the period of the early British colonies; GARY A. PUCKREIN, *Little England: Plantation Society and Anglo-Barbadian Politics, 1627–1700* (1984), a revisionist economic history, particularly strong on the creolizing process; HILARY BECKLES, *Black Rebellion in Barbados: The Struggle Against Slavery, 1627–1838* (1984), a provocative interpretation of slave resistance. JEROME S. HANDLER, *The Unappropriated People: Freedmen in the Slave Society of Barbados* (1974), which fills a gap in historiography; KARL WATSON, *The Civilised Island, Barbados: A Social History, 1750–1816* (1979), a study of the mature slave society; CLAUDE LEVY, *Emancipation, Sugar, and Federalism: Barbados and the West Indies, 1833–1876* (1980), on postslavery adjustments; GORDON C. MERRILL, *The Historical Geography of St. Kitts and Nevis, the West Indies* (1958), which discusses the colonial period on the islands; LENNOX HONYCHURCH, *The Dominica Story: A History of the Island,* 2nd ed. (1984), a well-illustrated study covering developments up to the 1980s and benefiting from the author's personal involvement in the constitutional changes leading to independence; and GEORGE BRIZAN, *Grenada, Island of Conflict: From*

Amerindians to People's Revolution, 1498–1979 (1984), the work of a Grenadian historian and politician.

Other islands: (Aruba and Netherlands Antilles): Although independent, Aruba is often treated in the literature with the Netherlands Antilles. Geographic information is given in J.H. WESTERMANN and J.I.S. ZONNEVELD, *Photo-Geological Observations and Land Capability & Land Use Survey of the Island of Bonaire, Netherlands Antilles* (1956); RENÉ RÖMER, *Curaçao* (1981; originally published in Dutch, 1977); JOHN Y. KEUR and DOROTHY L. KEUR, *Windward Children: A Study in Human Ecology of the Three Dutch Windward Islands in the Caribbean* (1960); H. HOETINK, *Het patroon van de oude Curaçaose samenleving* (1958), a sociological study, including a discussion of race; FOLKERT STEENMEIJER, *Food and Nutrition of Arubans* (1957); VERA M. GREEN, *Migrants in Aruba: Interethnic Integration* (1974); JULIA G. CRANE, *Educated to Emigrate: The Social Organization of Saba* (1971); and ANKE KLOMP, *Politics on Bonaire* (1987; originally published in Dutch, 1983).

A comprehensive history of the Netherlands Antilles, including Aruba, is JOHANNES HARTOG, *Geschiedenis van de Nederlandse Antillen,* 5 vol. (1954–81). One-volume treatments are CORNELIS C. GOSLINGA, *The Dutch in the Caribbean and on the Wild Coast, 1580–1680* (1971), *The Dutch in the Caribbean and in the Guianas, 1680–1791* (1985), and *A Short History of the Netherlands Antilles and Surinam* (1979), a concise survey. An overview of historical sources is given in M.A.P. MEILINK-ROELOFSZ (ed.), *Dutch Authors on West Indian History: A Historiographical Selection,* trans. from Dutch (1982). A special history is ISAAC S. EMMANUEL and SUZANNE A. EMMANUEL, *History of the Jews of the Netherlands Antilles,* 2 vol. (1970).

(The Bahamas): Geographic information is available in GAIL SAUNDERS, *The Bahamas: A Family of Islands* (1988), a general guidebook; MARY MOSELEY, *The Bahamas Handbook* (1926), a classic study of landscape, flora, fauna, history, economy, and tourism that covers the islands from Grand Bahama to Grand Turk; MICHAEL CRATON, *Sun 'n Sixpence: A Guide to Nassau and the Bahama Out Islands* (1964), a descriptive handbook; and MICHAEL HALKITIS, KAREN RIGG, and STEVEN SMITH, *The Climate of the Bahamas* (1980), including an examination of Grand Turk. NATHANIEL LORD BRITTON and CHARLES FREDERICK MILLSPAUGH, *The Bahama Flora* (1962), provides an examination of the botanical species on the islands from Grand Bahama to Grand Turk. Historic architecture is discussed in GAIL SAUNDERS and DONALD CARTWRIGHT, *Historic Nassau* (1979). Works on the Family Islands include STEVE DODGE, *Abaco, the History of an Out Island and Its Cays* (1983); and MARGERY O. ERICKSON, *Great Inagua* (1987).

The most comprehensive general history is MICHAEL CRATON, *A History of the Bahamas,* 3rd ed. (1986). PAUL ALBURY, *The Story of the Bahamas* (1975), expands on the growth of tourism, political parties, and nationalism and gives special attention to the Family Islands; see also his *Paradise Island Story* (1984). The loyalists and the slavery period are examined in GAIL SAUNDERS, *Bahamian Loyalists and Their Slaves* (1983), and *Slavery in the Bahamas, 1648–1838* (1985). Political histories include RANDOL FAWKES, *The Faith that Moved the Mountain* (1979); DORIS L. JOHNSON, *The Quiet Revolution in the Bahamas* (1972); and COLIN A. HUGHES, *Race and Politics in the Bahamas* (1981).

(Cayman Islands): Pictorial works include PAUL HUMANN, *Beautiful Isles, Cayman* (1986); and PAUL HUMANN and FEODOR PITCAIRN, *Cayman, Underwater Paradise* (1979). Information on the people is available from government publications such as *Statistical Abstracts of the Cayman Islands* (annual); and CAYMAN ISLANDS, *Annual Reports.* See also *Cayman Islands Government, Economic & Financial Review* (1982).

The most useful source on history remains GEORGE S.S. HIRST, *Notes on the History of the Cayman Islands,* in parts (1909–10, reprinted 1967, 5 vol. in 3). NEVILLE WILLIAMS, *A History of the Cayman Islands* (1970), is a short survey. See also ULF HANNERZ, *Caymanian Politics: Structure and Style in a Changing Island Society* (1974).

(Trinidad and Tobago): Geography is discussed in HELMUT BLUME, *The Caribbean Islands* (1974; originally published in German, 1968); and JEREMY TAYLOR, *Masquerade: The Visitor's Introduction to Trinidad and Tobago* (1986).

For history, see ERIC E. WILLIAMS, *History of the People of Trinidad and Tobago* (1962); ARIE BOOMERT, "The Arawak Indians of Trinidad and Coastal Guiana, ca. 1500–1650," *Journal of Caribbean History* 19(2):123–188 (1984); LINDA A. NEWSON, *Aboriginal and Spanish Colonial Trinidad: A Study in Culture Contact* (1976); BRIDGET BRERETON, *Race Relations in Colonial Trinidad, 1870–1900* (1979), which explores multiethnic population characteristic of Trinidad; and *A History of Modern Trinidad, 1783–1962* (1981); and DONALD WOOD, *Trinidad in Transition* (1968; reissued 1986).

(Turks and Caicos): Information on these islands is found in works cited above in *The Bahamas* section. For works specifically on the Turks and Caicos, see *Turks and Caicos Islands* (annual), a standard reference paper of the Foreign and Commonwealth Office of the British Government that provides basic facts; DONOVAN S. CORRELL and HELEN B. CORRELL, *Flora of the Bahama Archipelago: Including the Turks and Caicos Islands* (1982); W. JOHN RODMAN, "Turks and Caicos Islands: The Caribbean Secret," *The Caribbean & West Indian Chronicle,* special issue (June–July 1983), on economic conditions; J. HARRISON *et al.,* "Tourism and Development: The Problems of Small Islands in the Caribbean," *Environments* 16(3):125–130 (1984), a study of the Turks and Caicos economic development; C.O. HESSE and K. HESSE, "Conch History in the Turks and Caicos Islands," *Underwater Naturalist* 10(3):4–9 (1977), which discusses ecology and commercial fishing; and JAMES H. STARK, *Starks's History and Guide to the Bahama Islands: Including Their History, Inhabitants, Climate, Agriculture, Geology, Government, and Resources* (1891), a classic background study that provides a history.

(C.G.C./C.M.-L./S.H.L./F.W.Kn./N.L.G./H.J.Wi./C.A.G./M.J.-MacL./C.V.B./D.J.Bu./R.J.Ta./T.G.Ms./J.D.Mo./W.K.M./H.Ho./D.G.Ss./C.A.W./E.P.E./Da.W./B.M.B./J.H.Bo.)

Western Africa

South of the Sahara and east and north of the Atlantic Ocean, western Africa is latitudinally divided into two parallel belts of land: the western portion of the Sudan, a geographic area that stretches across the entire width of Africa, and the coastal region, or Guinea Coast. Each belt has its own geography, cultures, and history.

The nations of the western Sudan include Burkina Faso (formerly Upper Volta), Cape Verde, Chad, The Gambia, Mali, Mauritania, Niger, and Senegal. The nations of the Guinea Coast are Benin, Cameroon, Côte d'Ivoire (Ivory Coast), Equatorial Guinea, Ghana, Guinea, Guinea-Bissau, Liberia, Nigeria, Sierra Leone, and Togo.

Western Africa is a term used in the *Encyclopædia Britannica* to designate a geographic region within the continent of Africa. The term West Africa is also often used to refer to this part of the continent. As conventionally understood, however, West Africa comprises all of the areas considered here except Cameroon, Chad, Equatorial Guinea, and the Saharan parts of Mali, Mauritania, and Niger. West Africa is primarily a political and economic designation; these countries joined to establish the Economic Community of West African States (ECOWAS) in 1975.

The article is divided into the following sections:

The region 790
  Physical and human geography 790
    The land
    Traditional cultures
  History 806
    Muslims in western Africa
    The states of the Sudan
    The beginnings of European activity
    The Islāmic revolution in the western Sudan
    The Guinea coastlands and the Europeans (1807–79)
    Colonization
    Decolonization and the regaining of independence
    Western Africa since independence
Countries of the western Sudan 831
  Burkina Faso 831
  Cape Verde 833
  Chad 835
  The Gambia 839
  Mali 841
  Mauritania 846
  Niger 850
  Senegal 854
Countries of the Guinea Coast 860
  Benin 860
  Cameroon 864
  Côte d'Ivoire 868
  Equatorial Guinea 872
  Ghana 876
  Guinea 884
  Guinea-Bissau 888
  Liberia 890
  Nigeria 894
  Sierra Leone 905
  Togo 909
Bibliography 912

THE REGION

Physical and human geography

THE LAND

Geology. Western Africa is underlain by crystalline rocks that outcrop over about 55 percent of the subcontinent, elsewhere being buried under sedimentary rocks. Volcanic rocks constitute a third, small group of surface rocks.

The West African shield

The crystalline rocks, collectively referred to as the West African shield or craton, comprise three main types of rock assemblage. The basement complexes are highly deformed and contorted gneisses, migmatites (metamorphosed and banded mixed rocks), quartzites, and amphibolites. The supracrustal formations of phyllites, schists, banded ironstones, quartzites, and greenstones were originally laid down upon preexisting basement complexes as sedimentary and volcanic formations, but they have been folded, faulted, and metamorphosed during one or more episodes of orogenic deformation. The granitic intrusions, varying from several to hundreds of square miles in area, were intruded into basement complexes and supracrustals at the end of major tectonic events.

The West African shield consists of three age provinces. The oldest part, whose assemblages are Archean with reactivation ages older than 2.5 billion years, lies in Sierra Leone, Liberia, and Guinea and is called the Liberian Craton. The central part—in Côte d'Ivoire, Ghana, and Burkina Faso—is dominated by the Birimian supracrustals, which were deposited during the Proterozoic era and tectonized in the Eburnian event of 1.8 to two billion years ago. In the east, beneath Benin, Togo, Nigeria, Chad, and Niger, the shield contains Archean-age basement complexes and Proterozoic supracrustals, which were deformed and the basement reactivated between 650 and 500 million years ago during the Pan-African thermotectonic event (see below). In the far west, the Rokelide and Mauritanide metamorphosed and deformed rocks of Guinea and Sierra Leone show ages of 550 to 350 million years. As elsewhere in Africa, the shield rocks contain abundant and diverse mineral resources including iron ore, gold, rutile, bauxite, chromite, manganese, diamonds, copper, lead, zinc, and uranium, though many occurrences are small and low-grade.

The sedimentary rocks lie on the shield in broad, shallow (a maximum of three miles [five kilometres] thick) basins to the north and in narrow, deeper (a maximum of 7.5 miles thick) basins along the coast. The rock types include shales, sandstones, conglomerates, and limestones, which were originally deposited in vast lakes, deltas, and shallow seas. There were three long periods of sedimentation, although each period contained phases of erosion.

The oldest formations span the later Proterozoic to Paleozoic eras, from about one billion to about 350 million years ago, and include sediments of two glaciations, for during parts of this time the western African region lay close to the South Pole. These older formations were laid down in the Volta and Taoudeni basins (the latter one of the largest sedimentary basins in the world), and parts were involved later in the Pan-African orogenesis. The sandstones and conglomerates are now very hard and resistant and form prominent escarpments, such as the Volta and the Bandiagara, and plateaus, such as the Fouta Djallon and the Manding.

The second major period of sedimentation was during the Mesozoic and the early Cenozoic eras, from 200 to 65 million years ago, when the vast inland basins of Iullmedan and Chad and the narrower Benue basin developed. During this period the Atlantic Ocean began to open, and the sedimentary coastal basins of Senegal, Sierra Leone, Côte d'Ivoire, and southern Nigeria formed along

the continental margins. Sedimentation has continued to the present—albeit with interruptions due to vertical tectonic movements and sea level changes—on the Niger delta, in the Chad basin, and in the coastal basins. These vast sedimentary basins contain a wide range of mineral resources, including petroleum and natural gas (the Niger delta being among the largest fields in the world), coal, phosphates, gypsum, uranium, and zinc.

Generally less resistant or thinner than the older sedimentary rocks, the Mesozoic and Cenozoic rocks do not form striking topographic features or underlie vast plains. Some of the sandstones, however, produce extensive scarps, such as the Tegama and Awka.

Since the Pan-African event, igneous activity has been spatially limited. Four occurrences of volcanic rocks are worthy of note, however. First, intrusions of basic and ultrabasic magma occurred as sills, plugs, and dikes in Sierra Leone, Liberia, Guinea, and Côte d'Ivoire during the early stages in the breakup of the ancient supercontinent of Gondwanaland (see below) and the opening of the Atlantic Ocean (200 to 175 million years ago). Today the thicker intrusions produce bold cliffs and small plateaus. Second, approximately 92 million years ago kimberlite dikes and plugs were intruded from the upper mantle, also in Sierra Leone, Guinea, and Liberia. Now deeply eroded, they were the source for the diamond placer deposits in those countries. Third, between 340 and 145 million years ago large caldera volcanoes were built in Niger and Nigeria along a distinct north–south axis. Most of the volcanic carapaces have since been eroded, and the distinctive circular and ring-shaped granite plutons that fed the volcanoes have been exposed as mountain and hill massifs. They are rich in cassiterite, and their erosion has generated tin placer deposits that are mined in Nigeria. Fourth, extensive volcanicity produced lava plateaus, volcanoes, plugs, craters, and fumaroles along the mountain belt extending northeast from Mount Cameroon, itself a complex of active and extinct volcanoes, through the beautiful Bambouto, Adamaoua, and Mambila Songola highlands, toward the Chad basin. Although starting about 25 million years ago, this volcanicity has been particularly active during the past one million years.

The Pan-African event

The last in the series of orogenic and crustal reactivation (deformation and metamorphism of preexisting rocks) phases of the West African shield, the enigmatic Pan-African event, involved convergence between the West African Plate and the Central Saharan Plate (which included the shield rocks of Togo, Benin, Niger, and Nigeria), with rock deformation, granitization, and mountain building. Since then the subcontinent has undergone only slow vertical tectonic movements; these have caused profound erosion and the gradual exposure of the deeper parts of the shield assemblages. Sediments from this erosion accumulated in the sedimentary covers described above. After the Pan-African event, Africa was part of a supercontinent called Gondwanaland, which also included South America, peninsular India, Antarctica, and Australia. About 175 million years ago Gondwanaland began to break up along the lines of the present continental coasts. Sedimentary basins and rift valleys developed along future separation zones. There were several extensive marine transgressions across the low-lying eastern parts of the West African region during the Cretaceous period, at the same time slow uplift dominated the western parts of the shield. Gradually the Atlantic Ocean widened and during the past 65 million years intensive weathering and erosion processes dominated most of the subcontinent.

Relief and drainage. Virtually the whole of western Africa lies below 5,000 feet (1,500 metres); most of the region lies below 1,500 feet and is dominated by plains. Isolated high plateaus and mountains are found in Nigeria, Togo, Guinea, Liberia, Côte d'Ivoire, Cameroon, and Sierra Leone. Smaller hills and ridges abound, especially on the crystalline craton rocks.

The river system is dominated by the Niger, which, rising barely 310 miles from the Atlantic coast in Guinea, flows northeast toward the desert before looping southeast and south through Niger and Nigeria. Other large systems include the Sénégal, Volta, and Benue. Numerous short rivers flow to the Atlantic, and the Chad basin is the focus for inland rivers. River discharges are highly seasonal, especially in the northern areas.

Effects of weathering and erosion

The extensive erosional plains (planation surfaces), usually underlain by completely weathered rock (saprolite) 15 to 300 feet thick, are the result of long periods of weathering and erosion following pulses of tectonic uplift and subsidence. Four main planation surfaces have been recognized, each separated by rocky escarpments or hilly dissected country. The oldest and highest plains, well preserved in northern Nigeria, began forming at the beginning of the Cenozoic era and are underlain by the thickest saprolites, which often include a layer of iron and sometimes aluminum-rich duricrust. Most of western Africa, however, is made of younger planation surfaces that have formed during the past 25 million years. Depositional plains in areas of tectonic subsidence characterize the western and southwestern coastlands, the Niger delta and its lower valley, the Chad basin, and the subhumid lands bordering the Sahara desert.

Particularly resistant rock types give rise to isolated hills (inselbergs) whose prominence is enhanced by the monotony of the surrounding plains. Certain granite intrusions produce the striking dome-shaped bare hills called bornhardts; others are the more blocky koppies. Schists, banded ironstones, and quartzites are carved into steep-sided linear ridges, and hard sandstones and duricrusted saprolites produce escarpments, plateaus, and mesas.

In the humid rain forest belt, the plains are dominated by low, rounded hills and narrow, swampy valley floors eroded from thick porous saprolites. Landslides and deep, narrow ravines characterize the saprolite-blanketed higher hillslopes. In the drier savanna lands, valley floors tend to be wider and flatter and occasionally broken by ledges and low mesas capped by duricrust.

Two types of coasts dominate western Africa's ocean shoreline. Low, muddy coasts, with mangrove swamps and interconnecting tidal creeks, are found along major river deltas and along coasts where the offshore current is weak. Elsewhere are long, smooth, sandy beaches often backed by older, sandy barrier ridges and lagoons.

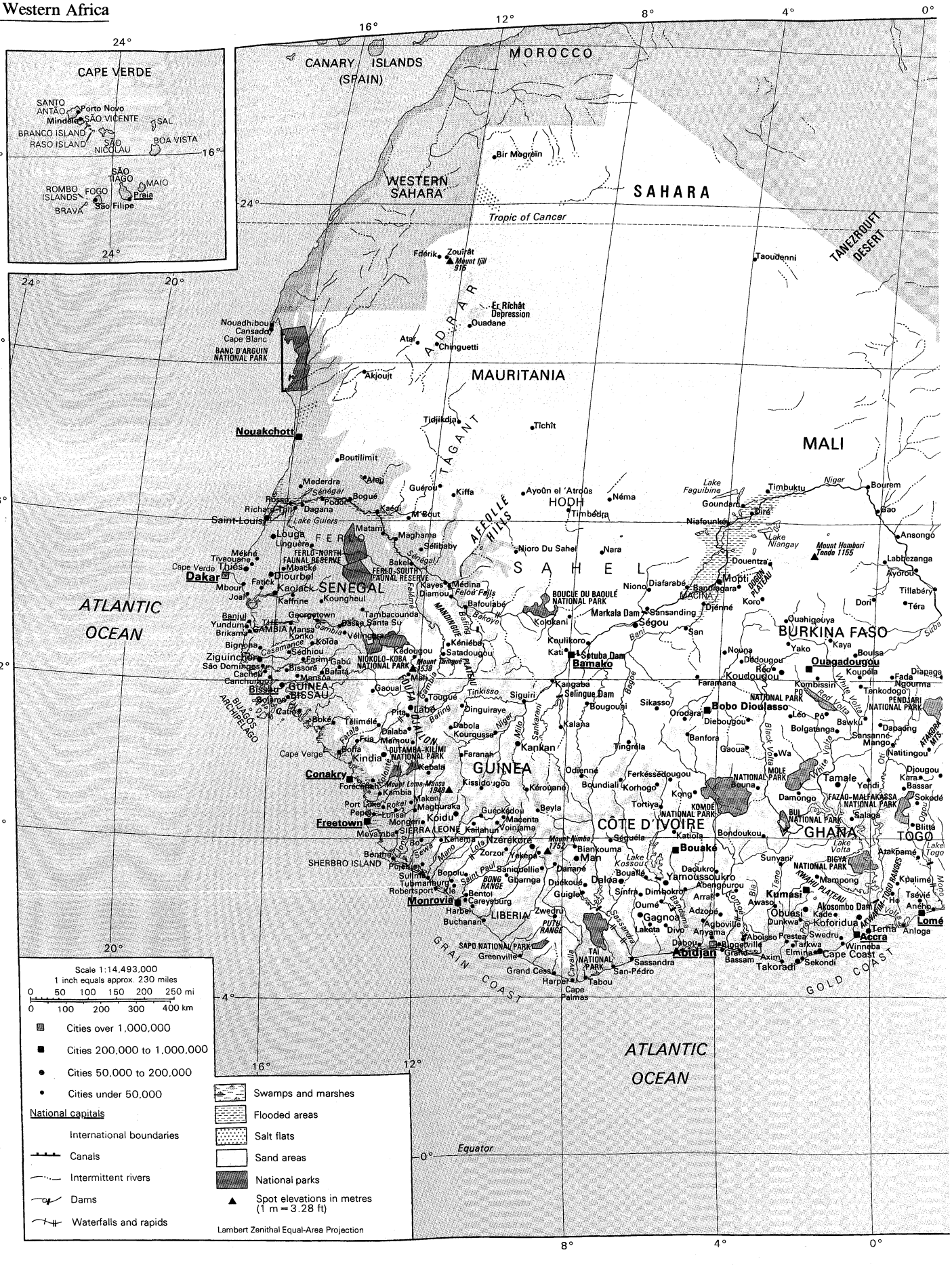
During the past 1.6 million years, the Quaternary period, global climates have changed frequently. In the colder periods western Africa was characterized by reduced rainfall, longer dry seasons, expansion of arid zones, and reduction in the extent of forested land. In the far north of the subcontinent are fossil sand dunes, relicts of these drier periods when desert aridity extended south by up to 300 miles. The last of these arid periods was from 22,000 to 14,000 years ago. Virtually all the rain forests were then replaced by open savannas. The alluvial floodplains in the region were built when climates returned to present-day levels of humidity during the past 10,000 years.

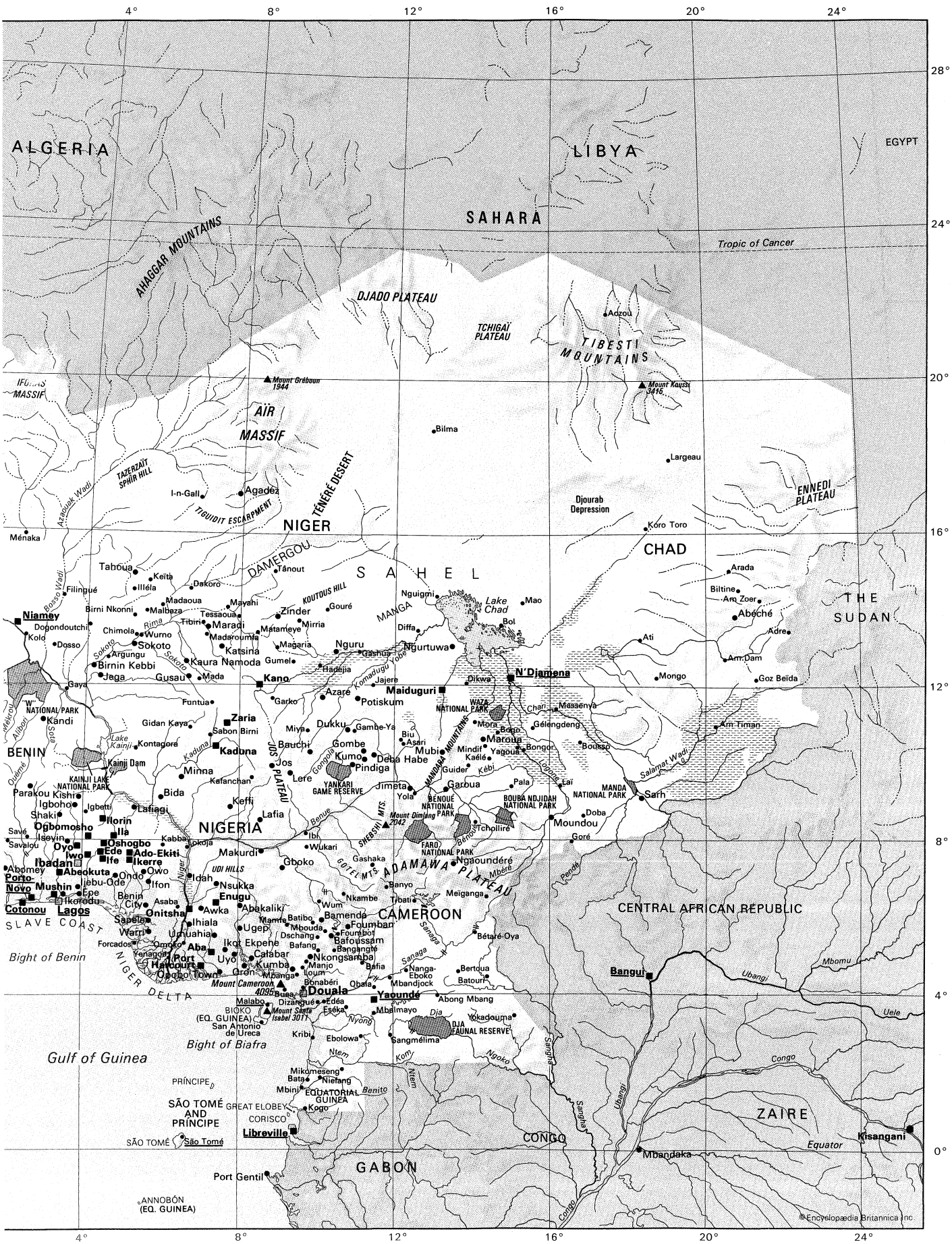
Soils. Western African soil types are strongly controlled by their parent materials, topographic position, and climate. Consequently, the spatial distribution of the major soil types reflects the north–south climatic zonation, the distribution of the major geologic formations, and, at the local scale, their topographic conditions.

Primary soil classes

There are four primary regional soil classes. In the rain forests, where rainfall exceeds 60 inches annually, are deep, red and yellow-red ferrallitic soils. They show little horizoning (distinct layers of soil) and are friable and very porous. Composed of kaolinitic clays and dispersed iron oxides, they contain no weatherable minerals, have a low cation exchange capacity (CEC), or ability to hold and exchange cations (important to plant growth), and are inherently infertile. Any fertility comes from the organic matter content, and this is almost totally leached within two years of clearance for agriculture.

Associated with the savanna woodland zone and the zones with about 20 to 50 inches of rainfall, and occupying the greater part of the region, are ferruginous soils. These are usually less than six feet deep, and horizons are well developed with prominent iron oxide mottles and concretions and clay textures in the B horizons below the organic-rich topsoil. They contain a moderate-to-high reserve of weatherable minerals and have textural properties that vary with the parent material from which they devel-


CAPE VERDE
SANTO ANTÃO
Porto Novo
Mindelo
SÃO VICENTE
SAL
BRANCO ISLAND
RASO ISLAND
SÃO NICOLAU
BOA VISTA
SÃO TIAGO
MAIO
ROMBO ISLANDS
FOGO
Praia
São Filipe
BRAVA
CANARY ISLANDS (SPAIN)
MOROCCO
WESTERN SAHARA
SAHARA
Tropic of Cancer
TANEZROUFT DESERT
ADRAR
Er Richât Depression
Ouadane
Chinguetti
Atar
Akjoujt
BANC D'ARGUIN NATIONAL PARK
Nouadhibou
Cansado
Cape Blanc
MAURITANIA
TAGANT
Nouakchott
Tidjikdja
Tichît
MALI
Boutilimit
Aleg
Mederdra
Guérou
Kiffa
Ayoûn el 'Atroûs
Néma
HODH
Timbédra
Lake Faguibine
Timbuktu
Niger
Bourem
Gao
Goundam
Diré
Niafounké
Lake Niangay
Ansongo
Mount Hombori Tondo 1155
Labbezanga
Ayorou
Tillabéry
Téra
Dori
Douentza
Mopti
DIOUDIN PLATEAU
Koro
Bandiagara
MACINA
Djénné
Diafarabé
Niono
Nara
Nioro Du Sahel
AFFOLLÉ HILLS
SAHEL
Rosso
Richard-Toll
Dagana
Podor
Bogué
Kaédi
M'Bout
Saint-Louis
Lake Guiers
Matam
Maghama
Sélibaby
Louga
Linguère
FERLO
FERLO-NORTH FAUNAL RESERVE
FERLO-SOUTH FAUNAL RESERVE
Mékhé
Tivaouane
Mbacké
Bakel
Sénégal
Cape Verde
Thiès
Dakar
Diourbel
Fatick
Kaolack
SENEGAL
Mbour
Joal
Kaffrine
Koungheul
Médina Feloo Falls
Kayes
Diamou
Faleme
Bafoulabé
BOUCLE DU BAOULÉ NATIONAL PARK
Markala Dam
Ségou
Sansanding
Koulikoro
Kolokani
Bani
Niger
Mali
Bakoye
Falémé
Banjul
Yundum
GAMBIA
Brikama
Georgetown
Basse Santa Su
Tambacounda
Gambia
Casamance
Korko
Kolda
Vélingara
Kédougou
Kénièba
Satadougou
MANDINGUE PLATEAU
Bamako
Sotuba Dam
Kati
San
BURKINA FAUSO
Ouahigouya
Yako
Kaya
Nouna
Dédougou
Boulsa
Ouagadougou
Koudougou
Réo
Koupéla
Fada
Diapaga
Kombissiri
Ngourma
Tankodogo
Ziguinchor
São Domingos
Cachéu
Canchungo
Bissora
Farim
Gabú
Batata
Mansôa
NIOKOLO-KOBA NATIONAL PARK
Mount Tamgué 1538
Gaoual
Mali
Tougué
Tinkisso
Siguiri
Kangaba
Selingue Dam
Bougouni
Faramana
Sikasso
Bobo Dioulasso
Orodara
Diebougou
Léo
Pô
Bawku
Dabaang
Sansanné-Mango
Bolgatanga
PENDJARI NATIONAL PARK
ATAKORA MTS.
Red Volta
White Volta
Black Volta
Bissau
GUINEA-BISSAU
Bolama
Bijagós Archipelago
Canal du Geba
Farim
Boké
Télimélé
Pita
Labé
Dinguiraye
Bafing
Niger
Milo
Sankarani
Bagoé
Kalana
Banfora
Gaoua
Wa
Natitingou
Djougou
Kara
Bassar
Tamale
Yendi
Damongo
MOLE NATIONAL PARK
Bouna
FAZAO-MALFAKASSA NATIONAL PARK
Sokodé
Ogou
Salaga
Blitta
Atakpamé
TOGO
Lake Togo
Lake Volta
BUI NATIONAL PARK
BIGYA NATIONAL PARK
Sunyani
Bondoukou
Katiola
Séguéla
Odienné
Ferkessédougou
Boundiali
Korhogo
Kong
Tortiya
KOMOÉ NATIONAL PARK
Cape Verge
Fatala
Fria
Mamou
Dalaba
Pita
Kindia
Conakry
Forécariah
Mount Loma Mansa 1940
Kambia
Port Loko
Pepel
Lunsar
Rokel
Makeni
Magburaka
Freetown
Mongeri
SIERRA LEONE
Moyamba
Bo
Kenema
Koidu
Kailahun
Voinjama
Guéckédou
Macenta
Beyla
Kérouané
Kissidougou
Faranah
Kankan
Tingréla
GUINEA
Kolenté
Kolenté
Kabala
Kamalo
Kouroussa
Dabola
Niger
Milo
Sankarani
Kalana
Bougouni
Sikasso
Faramana
Kangaba
Selingue Dam
Bagoé
Siguiri
Tinkisso
Tougué
Labé
Dinguiraye
Bafing
Falémé
Gaoual
Mali
Gabú
Batata
Mansôa
Farim
Bissora
Canchungo
Cachéu
São Domingos
Ziguinchor
Bissau
GUINEA-BISSAU
Bolama
Bijagós Archipelago
Canal du Geba
Boké
Télimélé
Pita
Labé
Dinguiraye
Bafing
Niger
Milo
Sankarani



© Encyclopædia Britannica, Inc.

MAP INDEX

Cities and towns

Aba ..... 5 07 N 7 22 E
Abakaliki ..... 6 20 N 8 06 E
Abéché ..... 13 49 N 20 49 E
Abengourou ..... 6 44 N 3 29 W
Abeokuta ..... 7 09 N 3 21 E
Abidjan ..... 5 19 N 4 02 W
Aboisso ..... 5 28 N 3 12 W
Abomey ..... 7 11 N 1 59 E
Abong Mbang .... 3 59 N 13 11 E
Accra ..... 5 33 N 0 13 W
Ado-Ekiti ..... 7 38 N 5 13 E
Adre ..... 13 28 N 22 12 E
Adzopé ..... 6 06 N 3 52 W
Agadez ..... 16 58 N 7 59 E
Agboville ..... 5 56 N 4 13 W
Akjoujt ..... 19 45 N 14 23 W
Aleg ..... 17 03 N 13 55 W
Am Dam ..... 12 46 N 20 29 E
Am Timan ..... 11 02 N 20 17 E
Am Zoer ..... 14 13 N 21 23 E
Aného ..... 6 14 N 1 36 E
Anloga ..... 5 48 N 0 54 E
Ansongo ..... 15 40 N 0 30 E
Anyama ..... 5 30 N 4 03 W
Aozou ..... 21 49 N 17 25 E
Arada ..... 15 01 N 20 40 E
Argungu ..... 12 45 N 4 31 E
Arrah ..... 6 40 N 3 58 W
Asaba ..... 6 11 N 6 45 E
Asari ..... 10 31 N 12 18 E
Atakpamé ..... 7 32 N 1 08 E
Atar ..... 20 31 N 13 03 W
Ati ..... 13 13 N 18 20 E
Awaso ..... 6 14 N 2 16 W
Awka ..... 6 13 N 7 05 E
Axim ..... 4 52 N 2 14 W
Ayorou ..... 14 44 N 0 55 E
Ayoûn el 'Atroûs .. 16 40 N 9 37 W
Azare ..... 11 41 N 10 12 E
Bafang ..... 5 09 N 10 11 E
Bafatá ..... 12 10 N 14 40 W
Bafia ..... 4 45 N 11 14 E
Bafoulabé ..... 13 48 N 10 50 W
Bafoussam ..... 5 28 N 10 25 E
Bakel ..... 14 54 N 12 27 W
Bamako ..... 12 39 N 8 00 W
Bamenda ..... 5 56 N 10 10 E
Bandiagara ..... 14 21 N 3 37 W
Banfora ..... 10 38 N 4 46 W
Bangangté ..... 5 09 N 10 31 E
Banjul ..... 13 27 N 16 35 W
Banyo ..... 6 45 N 11 49 E
Bassar ..... 9 15 N 0 47 E
Basse Santa Su .. 13 19 N 14 13 W
Bata ..... 1 51 N 9 45 E
Batibo ..... 5 50 N 9 52 E
Batouri ..... 4 26 N 14 22 E
Bauchi ..... 10 19 N 9 50 E
Bawku ..... 11 03 N 0 15 W
Benin City ..... 6 20 N 5 38 E
Bentol ..... 6 26 N 10 36 W
Bertoua ..... 4 35 N 13 41 E
Bétaré-Oya ..... 5 36 N 14 05 E
Beyla ..... 8 41 N 8 38 W
Biankouma ..... 7 44 N 7 37 W
Bida ..... 9 05 N 6 01 E
Bignona ..... 12 49 N 16 14 W
Bilma ..... 18 41 N 12 56 E
Biltine ..... 14 32 N 20 55 E
Bingerville ..... 5 21 N 3 54 W
Bir Mogreïn ..... 25 14 N 11 35 W
Birni Nkonni ..... 13 48 N 5 15 E
Birnin Kebbi ..... 12 28 N 4 12 E
Bissau ..... 11 51 N 15 35 W
Bissorã ..... 12 13 N 15 27 W
Biu ..... 10 37 N 12 12 E
Blitta ..... 8 19 N 0 59 E
Bo ..... 7 58 N 11 45 W
Bobo Dioulasso ... 11 12 N 4 18 W
Boffa ..... 10 10 N 14 02 W
Bogo ..... 10 44 N 14 36 E
Bogué (Boghé) ... 16 35 N 14 16 W
Boké ..... 10 56 N 14 18 W
Bol ..... 13 28 N 14 43 E
Bolama ..... 11 35 N 15 28 W
Bolgatanga ..... 10 47 N 0 51 W
Bonabéri ..... 4 04 N 9 41 E
Bondoukou ..... 8 02 N 2 48 W
Bongor ..... 10 17 N 15 22 E
Bonthe ..... 7 32 N 12 30 W
Bopolu ..... 6 54 N 10 46 W
Bouaflé ..... 6 59 N 5 45 W
Bouaké ..... 7 41 N 5 02 W
Bougouni ..... 11 25 N 7 29 W
Boulsa ..... 12 39 N 0 34 W
Bouna ..... 9 16 N 3 00 W
Boundiali ..... 9 31 N 6 29 W
Bourem ..... 16 57 N 0 21 W
Bousso ..... 10 29 N 16 43 E
Boutilimit ..... 17 33 N 14 42 W
Brikama ..... 13 16 N 16 39 W
Buchanan (Grand Bassa) ..... 5 53 N 10 03 W
Buea ..... 4 09 N 9 14 E
Cacheu ..... 12 16 N 16 10 W
Calabar ..... 4 57 N 8 19 E
Canchungo (Teixeira Pinto) .. 12 26 N 16 05 W
Cansado ..... 20 51 N 17 02 W
Cape Coast ..... 5 06 N 1 15 W
Careysburg ..... 6 24 N 10 33 W
Catió ..... 11 17 N 15 15 W
Chimola ..... 13 18 N 5 22 E
Chinguetti ..... 20 27 N 12 22 W
Conakry ..... 9 31 N 13 43 W
Cotonou ..... 6 21 N 2 26 E
Dabola ..... 10 45 N 11 07 W
Dabou ..... 5 19 N 4 23 W
Dagana ..... 16 31 N 15 30 W
Dakar ..... 14 40 N 17 26 W
Dakoro ..... 14 31 N 6 46 E
Dalaba ..... 10 42 N 12 15 W
Daloa ..... 6 53 N 6 27 W
Damongo ..... 9 05 N 1 49 W
Danané ..... 7 16 N 8 09 W
Daoukro ..... 7 03 N 3 58 W
Dapaong ..... 10 52 N 0 12 E
Deba Habe ..... 10 13 N 11 23 E
Dédougou ..... 12 28 N 3 28 W
Diafarabé ..... 14 09 N 5 01 W
Diamou ..... 14 05 N 11 16 W
Diapaga ..... 12 04 N 1 47 E
Diébougou ..... 10 58 N 3 15 W
Diffa ..... 13 19 N 12 37 E
Dikwa ..... 12 02 N 13 55 E
Dimbokro ..... 6 39 N 4 42 W
Dinguiraye ..... 11 18 N 10 43 W
Diourbel ..... 14 40 N 16 15 W
Diré ..... 16 16 N 3 24 W
Divo ..... 5 50 N 5 22 W
Dizangué ..... 3 46 N 9 59 E
Djénné ..... 13 54 N 4 33 W
Djougou ..... 9 42 N 1 40 E
Doba ..... 8 39 N 16 51 E
Dogondoutchi .... 13 38 N 4 02 E
Dori ..... 14 02 N 0 02 W
Dosso ..... 13 03 N 3 12 E
Douala ..... 4 03 N 9 42 E
Douentza ..... 15 00 N 2 57 W
Dschang ..... 5 27 N 10 04 E
Duékoué ..... 6 45 N 7 21 W
Dukku ..... 10 49 N 10 46 E
Dunkwa ..... 5 58 N 1 47 W
Ebolowa ..... 2 54 N 11 09 E
Ede ..... 7 44 N 4 26 E
Edéa ..... 3 48 N 10 08 E
Elmina ..... 5 05 N 1 21 W
Enugu ..... 6 26 N 7 29 E
Epe ..... 6 35 N 3 59 E
Eséka ..... 3 39 N 10 46 E
Fada Ngourma ... 12 04 N 0 21 E
Faramana ..... 12 03 N 4 40 W
Faranah ..... 10 02 N 10 44 W
Farim ..... 12 29 N 15 13 W
Fatick ..... 14 20 N 16 25 W
Fdérik ..... 22 41 N 12 43 W
Ferkéssédougou .. 9 36 N 5 12 W
Filingué ..... 14 21 N 3 19 E
Forcados ..... 5 22 N 5 26 E
Forécariah ..... 9 26 N 13 06 W
Fort-Archambault, see Sarh
Fort Lamy, see N'Djamena
Foumban ..... 5 43 N 10 55 E
Foumbot ..... 5 30 N 10 38 E
Freetown ..... 8 30 N 13 15 W
Fria ..... 10 27 N 13 32 W
Funtua ..... 11 32 N 7 19 E
Gabú (Nova Lamego) ..... 12 17 N 14 13 W
Gagnoa ..... 6 08 N 5 56 W
Gambe-Ya ..... 10 48 N 10 57 E
Gao ..... 16 16 N 0 03 W
Gaoua ..... 10 20 N 3 11 W
Gaoual ..... 11 45 N 13 12 W
Garko ..... 11 39 N 8 48 E
Garoua ..... 9 18 N 13 24 E
Gashaka ..... 7 22 N 11 29 E
Gashua ..... 12 52 N 11 03 E
Gaya ..... 11 53 N 3 27 E
Gbarnga ..... 7 00 N 9 29 W
Gboko ..... 7 19 N 9 00 E
Gélengdeng ..... 10 56 N 15 32 E
Georgetown ..... 13 32 N 14 46 W
Gidan Kaya ..... 10 51 N 6 45 E
Gombe ..... 10 17 N 11 10 E
Goré ..... 7 55 N 16 38 E
Goundam ..... 16 25 N 3 40 W
Gouré ..... 13 58 N 10 18 E
Goz Beïda ..... 12 13 N 21 25 E
Grand Bassa, see Buchanan
Grand-Bassam ... 5 12 N 3 44 W
Grand Cess (Grand Sesters) .. 4 34 N 8 13 W
Greenville (Sino) .. 5 00 N 9 02 W
Guéckédou ..... 8 33 N 10 09 W
Guérou ..... 16 48 N 11 50 W
Guider ..... 9 56 N 13 57 E
Guiglo ..... 6 33 N 7 29 W
Gumel ..... 12 38 N 9 23 E
Gusau ..... 12 10 N 6 40 E
Hadejia ..... 12 27 N 10 03 E
Harbel ..... 6 16 N 10 21 W
Harper ..... 4 22 N 7 43 W
Ho ..... 6 36 N 0 28 E
Ibadan ..... 7 23 N 3 54 E
Ibi ..... 8 11 N 9 45 E
Idah ..... 7 06 N 6 44 E
Ife ..... 7 28 N 4 34 E
Ifon ..... 6 55 N 5 46 E
Igbetti ..... 8 45 N 4 08 E
Igboho ..... 8 50 N 3 45 E
Ihiala ..... 5 51 N 6 51 E
Ijebu-Ode ..... 6 49 N 3 56 E
Ikerre ..... 7 30 N 5 14 E
Ikorodu ..... 6 37 N 3 31 E
Ikot Ekpene ..... 5 10 N 7 43 E
Ila ..... 8 01 N 4 54 E
Illéla ..... 14 28 N 5 15 E
Ilorin ..... 8 30 N 4 33 E
I-n-Gall ..... 16 47 N 6 56 E
Iseyin ..... 7 58 N 3 36 E
Iwo ..... 7 38 N 4 11 E
Jajere ..... 11 59 N 11 26 E
Jega ..... 12 13 N 4 23 E
Jimeta ..... 9 17 N 12 28 E
Joal ..... 14 10 N 16 51 W
Jos ..... 9 55 N 8 54 E
Kabala ..... 9 35 N 11 33 W
Kabba ..... 7 50 N 6 04 E
Kade ..... 6 05 N 0 50 W
Kaduna ..... 10 31 N 7 26 E
Kaédi ..... 16 09 N 13 30 W
Kaélé ..... 10 07 N 14 27 E
Kafanchan ..... 9 35 N 8 18 E
Kaffrine ..... 14 06 N 15 33 W
Kailahun ..... 8 17 N 10 34 W
Kalana ..... 10 47 N 8 12 W
Kambia ..... 9 07 N 12 55 W
Kandi ..... 11 08 N 2 56 E
Kangaba ..... 11 56 N 8 25 W
Kankan ..... 10 23 N 9 18 W
Kano ..... 12 00 N 8 31 E
Kaolack ..... 14 09 N 16 04 W
Kara (Lama Kara) . 9 33 N 1 12 E
Kati ..... 12 44 N 8 04 W
Katiola ..... 8 08 N 5 06 W
Katsina ..... 13 00 N 7 36 E
Kaura Namoda ... 12 36 N 6 35 E
Kaya ..... 13 05 N 1 05 W
Kayes ..... 14 27 N 11 26 W
Kédougou ..... 12 33 N 12 11 W
Keffi ..... 8 51 N 7 52 E
Keïta ..... 14 46 N 5 46 E
Kenema ..... 7 52 N 11 12 W
Kéniéba ..... 12 50 N 11 14 W
Kérouané ..... 9 16 N 9 01 W
Kiffa ..... 16 37 N 11 24 W
Kindia ..... 10 04 N 12 51 W
Kishi ..... 9 05 N 3 51 E
Kissidougou ..... 9 11 N 10 06 W
Kle ..... 6 42 N 10 53 W
Koforidua ..... 5 14 N 1 20 W
Kogo ..... 1 05 N 9 42 E
Koidu ..... 8 38 N 10 59 W
Kolda ..... 12 53 N 14 57 W
Kolo ..... 13 19 N 2 20 E
Kolokani ..... 13 35 N 8 02 W
Kombissiri ..... 12 04 N 1 20 W
Kong ..... 9 09 N 4 37 W
Kontagora ..... 10 24 N 5 29 E
Korhogo ..... 9 27 N 5 38 W
Koro ..... 14 04 N 3 05 W
Koro Toro ..... 16 05 N 18 30 E
Koudougou ..... 12 15 N 2 22 W
Koulikoro ..... 12 53 N 7 33 W
Koungheul ..... 13 59 N 14 48 W
Koupéla ..... 12 11 N 0 21 W
Kouroussa ..... 10 39 N 9 53 W
Kpalimé (Palimé) .. 6 54 N 0 38 E
Kribi ..... 2 57 N 9 55 E
Kumasi ..... 6 41 N 1 37 W
Kumba ..... 4 38 N 9 25 E
Kumo ..... 10 03 N 11 13 E
Labbezanga ..... 14 57 N 0 42 E
Labé ..... 11 19 N 12 17 W
Lafia ..... 8 29 N 8 31 E
Lafiagi ..... 8 52 N 5 25 E
Lagos ..... 6 27 N 3 23 E
Laï ..... 9 24 N 16 18 E
Lakota ..... 5 51 N 5 41 W
Lama Kara, see Kara
Largeau ..... 17 55 N 19 07 E
Léo ..... 11 06 N 2 06 W
Lere ..... 9 43 N 9 21 E
Linguère ..... 15 24 N 15 07 W
Lokoja ..... 7 48 N 6 44 E
Lomé ..... 6 08 N 1 13 E
Louga ..... 15 37 N 16 13 W
Loum ..... 4 43 N 9 44 E
Lunsar ..... 8 41 N 12 32 W
Macenta ..... 8 33 N 9 28 W
Mada ..... 12 09 N 6 56 E
Madaoua ..... 14 06 N 6 26 E
Madaroumfa ..... 13 18 N 7 09 E
Magaria ..... 13 00 N 8 54 E
Magburaka ..... 8 43 N 11 57 W
Maghama ..... 15 31 N 12 51 W
Maiduguri ..... 11 51 N 13 09 E
Makeni ..... 8 53 N 12 03 W
Makurdi ..... 7 44 N 8 32 E
Malabo (Santa Isabel) ..... 3 21 N 8 40 E
Malbaza ..... 13 56 N 5 31 E
Mali ..... 12 05 N 12 18 W
Mamfe ..... 5 46 N 9 17 E
Mamou ..... 10 23 N 12 05 W
Mampong ..... 7 04 N 1 24 W
Man ..... 7 24 N 7 33 W
Manjo ..... 4 51 N 9 49 E
Mansa Konko .... 13 28 N 15 33 W
Mansôa ..... 12 04 N 15 19 W
Mao ..... 14 07 N 15 19 E
Maradi ..... 13 29 N 7 06 E
Maroua ..... 10 36 N 14 20 E
Massenya ..... 11 24 N 16 10 E
Matam ..... 15 40 N 13 15 W
Matameye ..... 13 26 N 8 28 E
Mayahi ..... 13 58 N 7 40 E
Mbacké ..... 14 48 N 15 55 W
Mbalmayo ..... 3 31 N 11 30 E
Mbandjock ..... 4 27 N 11 55 E
Mbanga ..... 4 30 N 9 34 E
Mbini ..... 1 34 N 9 37 E
Mbouda ..... 5 38 N 10 15 E
Mbour ..... 14 24 N 16 58 W
M'Bout ..... 16 02 N 12 35 W
Mederdra ..... 16 55 N 15 39 W
Médina ..... 14 24 N 11 43 W
Meïganga ..... 6 31 N 14 18 E
Mékhé ..... 15 07 N 16 38 W
Ménaka ..... 15 55 N 2 24 E
Mikomeseng ..... 2 08 N 10 37 E
Mindelo ..... 16 53 N 25 00 W
Mindif ..... 10 24 N 14 26 E
Minna ..... 9 37 N 6 33 E
Mirria ..... 13 43 N 9 07 E
Miya ..... 10 54 N 9 50 E
Mongeri ..... 8 19 N 11 44 W
Mongo ..... 12 11 N 18 42 E
Monrovia ..... 6 19 N 10 48 E
Mopti ..... 14 30 N 4 12 W
Mora ..... 11 03 N 14 09 E
Moundou ..... 8 34 N 16 05 E
Moyamba ..... 8 10 N 12 26 W
Mubi ..... 10 16 N 13 16 E
Mushin ..... 6 32 N 3 22 E
Nanga-Eboko .... 4 41 N 12 22 E
Nara ..... 15 10 N 7 17 W
Natitingou ..... 10 19 N 1 22 E

N'Djamena (Fort Lamy) ....12 07 N 15 03 E
Néma .....16 37 N 7 15 W
Ngaoundéré ..... 7 19 N 13 35 E
Nguigmi .....14 15 N 13 07 E
Ngurtuwa .....13 05 N 13 34 E
Nguru .....12 53 N 10 28 E
Niafounké .....15 56 N 4 00 W
Niamey .....13 31 N 2 07 E
Niefang ..... 1 51 N 10 15 E
Niono .....14 15 N 6 00 W
Nioro Du Sahel ...15 14 N 9 35 W
Nkambe ..... 6 38 N 10 40 E
Nkongsamba ..... 4 57 N 9 56 E
Nouadhibou .....20 54 N 17 04 W
Nouakchott .....18 06 N 15 57 W
Nouna .....12 44 N 3 52 W
Nova Lamego, see Gabú
Nsukka ..... 6 52 N 7 23 E
Nzérékoré ..... 7 45 N 8 49 W
Obala ..... 4 10 N 11 32 E
Obuasi ..... 6 12 N 1 40 W
Odienné ..... 9 30 N 7 34 W
Ogbomosho ..... 8 08 N 4 16 E
Omoko ..... 5 21 N 6 39 E
Ondo ..... 7 06 N 4 50 E
Onitsha ..... 6 10 N 6 47 E
Opobo Town ..... 4 31 N 7 32 E
Orodara .....10 59 N 4 55 W
Oron ..... 4 50 N 8 14 E
Oshogbo ..... 7 46 N 4 34 E
Ouadane .....20 56 N 11 37 W
Ouagadougou ....12 22 N 1 31 W
Ouahigouya .....13 35 N 2 25 W
Oumé ..... 6 23 N 5 25 W
Owo ..... 7 11 N 5 35 E
Oyo ..... 7 51 N 3 56 E
Pala ..... 9 22 N 14 54 E
Palimé, see Kpalimé
Parakou ..... 9 21 N 2 37 E
Pepel ..... 8 35 N 13 03 W
Pindiga ..... 9 59 N 10 54 E
Pita .....11 05 N 12 24 W
Pô .....11 10 N 1 09 W
Podor .....16 40 N 14 57 W
Port Harcourt .... 4 46 N 7 01 E
Port Loko ..... 8 46 N 12 47 W
Porto-Novo ..... 6 29 N 2 37 E
Porto Novo .....17 01 N 25 04 W
Potiskum .....11 43 N 11 04 E
Praia .....14 55 N 23 31 W
Prestea ..... 5 26 N 2 09 W
Pujehun ..... 7 21 N 11 42 W
Réo .....12 19 N 2 28 W
Richard-Toll .....16 28 N 15 41 W
Robertsport ..... 6 45 N 11 22 W
Rosso .....16 30 N 15 49 W
Sabon Birni .....10 49 N 7 18 E
Saint-Louis .....16 02 N 16 30 W
Salaga ..... 8 33 N 0 31 W
San .....13 18 N 4 54 W
San Antonio de Ureca ..... 3 16 N 8 32 E
San-Pédro ..... 4 44 N 6 37 W
Sangmélima ..... 2 56 N 11 59 E
Saniquellie (Sangbui) ..... 7 22 N 8 43 W
Sansanding .....13 44 N 6 00 W
Sansanné-Mango ..10 21 N 0 28 E
Santa Isabel, see Malabo
São Domingos ...12 24 N 16 12 W
São Filipe .....14 54 N 24 31 W
São Tomé ..... 0 20 N 6 44 E
Sapele ..... 5 55 N 5 42 E
Sarh (Fort-Archambault) ... 9 09 N 18 23 E
Sassandra ..... 4 57 N 6 05 W
Satadougou .....12 37 N 11 25 W
Savalou ..... 7 56 N 1 58 E
Savé ..... 8 02 N 2 29 E
Sédhiou .....12 44 N 15 33 W
Ségou .....13 27 N 6 16 W
Séguéla ..... 7 57 N 6 40 W
Sekondi ..... 4 56 N 1 42 W
Sélibaby .....15 10 N 12 11 W
Shaki ..... 8 40 N 3 23 E
Siguiri .....11 25 N 9 10 W
Sikasso .....11 19 N 5 40 W
Sinfra ..... 6 37 N 5 55 W
Sino, see Greenville
Sokodé ..... 8 59 N 1 08 E
Sokoto .....13 04 N 5 15 E
Sulima ..... 6 58 N 11 35 W
Sunyani ..... 7 20 N 2 20 W
Swedru ..... 5 32 N 0 42 W
Tabou ..... 4 25 N 7 21 W
Tahoua .....14 54 N 5 16 E
Takoradi ..... 4 53 N 1 45 W
Tamale ..... 9 24 N 0 50 W
Tambacounda ....13 47 N 13 40 W
Tânout .....14 58 N 8 53 E
Taoudenni .....22 40 N 3 59 W
Tarkwa ..... 5 18 N 1 59 W
Tchien, see Zwedru
Tcholliré ..... 8 24 N 14 10 E
Teixeira Pinto, see Canchungo
Télimélé .....10 54 N 13 02 W
Tema ..... 5 37 N 0 01 W
Tenkodogo .....11 47 N 0 22 W
Téra .....14 01 N 0 45 E
Tessaoua .....13 45 N 7 59 E
Thiès .....14 48 N 16 56 W
Tibati ..... 6 28 N 12 38 E
Tibiri .....13 34 N 7 04 E
Tichît .....18 28 N 9 30 W
Tidjikdja .....18 33 N 11 25 W
Tillabéry .....14 13 N 1 27 E
Timbédra .....16 15 N 8 10 W
Timbuktu .....16 46 N 3 01 W
Tingréla .....10 29 N 6 24 W
Tivaouane .....14 57 N 16 49 W
Tortiya ..... 8 46 N 5 41 W
Tougué .....11 27 N 11 41 W
Tsévié ..... 6 25 N 1 13 E
Tubmanburg (Vaitown) ..... 6 52 N 10 49 W
Ugep ..... 5 48 N 8 05 E
Umuahia ..... 5 32 N 7 29 E
Uyo ..... 5 03 N 7 56 E
Vélingara .....13 09 N 14 07 W
Voinjama ..... 8 25 N 9 45 W
Wa .....10 03 N 2 29 W
Warri ..... 5 31 N 5 45 E
Winneba ..... 5 20 N 0 37 W
Wukari ..... 7 51 N 9 47 E
Wum ..... 6 23 N 10 04 E
Wurno .....13 18 N 5 26 E
Yagoua .....10 20 N 15 14 E
Yako .....12 58 N 2 16 W
Yamoussoukro ... 6 49 N 5 17 W
Yaoundé ..... 3 52 N 11 31 E
Yekepa ..... 7 35 N 8 32 W
Yenagoa ..... 4 55 N 6 16 E
Yendi ..... 9 26 N 0 01 W
Yokadouma ..... 3 31 N 15 03 E
Yola ..... 9 12 N 12 29 E
Yundum .....13 20 N 16 41 W
Zaria .....11 04 N 7 42 E
Ziguinchor .....12 35 N 16 16 W
Zinder .....13 48 N 8 59 E
Zorzor ..... 7 47 N 9 26 W
Zouîrât .....22 42 N 12 30 W
Zwedru (Tchien) .. 6 04 N 8 08 W

Physical features and points of interest

Adamawa Plateau ..... 7 00 N 13 00 E
Adrar, *region* ..... 20 30 N 13 30 W
Affollé Hills .....16 55 N 10 25 W
Air Massif .....18 00 N 8 30 E
Akosombo (Volta) Dam ..... 6 16 N 0 03 E
Akwapim–Togo Ranges ..... 7 25 N 0 40 E
Alibori, *river* .....11 56 N 3 17 E
Annobón (Pagalu), *island* .. 1 25 S 5 36 E
Atakora Mountains .....10 45 N 1 30 E
Atlantic Ocean ... 8 00 N 16 00 E
Azaouak Wadi ....15 30 N 3 18 E
Bafing, *river* .....13 49 N 10 50 W
Bagoé, *river* .....12 36 N 6 34 W
Bakoye, *river* .....13 49 N 10 50 W
Banc d'Arguin National Park ....20 00 N 16 30 W
Bandama, *river* ... 5 10 N 5 00 W
Bani, *river* .....14 30 N 4 12 W
Benin, Bight of ... 5 30 N 4 00 E
Benito (Woleu), *river* ..... 1 36 N 9 37 E
Bénoué National Park ..... 8 30 N 14 00 E
Benue (Bénoué), *river* ..... 7 48 N 6 46 E
Bewa, see Mano
Bia, *river* ..... 5 21 N 3 11 W
Biafra, Bight of ... 3 20 N 9 20 E
Bijagós Archipelago .....11 15 N 16 05 W
Bintimani, see Loma Mansa
Bioko, *island* ..... 3 30 N 8 42 E
Black Volta (Volta Noire), *river* ..... 8 41 N 1 33 W
Blanc, Cape .....21 00 N 17 04 W
Boa Vista, *island* .....16 05 N 22 50 W
Bong Range ..... 6 52 N 10 10 W
Bosso Wadi .....12 25 N 2 50 E
Bouba Ndjidah National Park .... 8 30 N 14 30 E
Boucle du Baoulé National Park .....14 00 N 9 00 W
Branco Island ....16 39 N 24 41 W
Brava, *island* .....14 52 N 24 43 W
Bui National Park .. 8 30 N 2 20 W
Cameroon, Mount ..... 4 12 N 9 11 E
Campo, see Ntem
Casamance, *river* .....12 33 N 16 46 W
Cavalla (Cavally), *river* ..... 4 22 N 7 32 W
Chad, Lake .....13 20 N 14 00 E
Chari, *river* .....12 58 N 14 31 E
Corisco, *island* ... 0 55 N 9 19 E
Damergou, *region* .....15 00 N 8 55 E
Digya National Park ..... 7 00 N 0 30 W
Dimlang (Vogel Peak), Mount .... 8 24 N 11 47 E
Dja, *river* ..... 5 18 N 12 14 E
Dja Faunal Reserve ..... 3 20 N 13 00 E
Djado Plateau ....21 45 N 12 50 E
Djourab Depression .....16 30 N 18 00 E
Dogon Plateau ...14 30 N 3 30 W
Elobey, Great, *island* ..... 0 59 N 9 30 E
Ennedi Plateau ...17 15 N 22 00 E
Er Richât Depression .....21 07 N 11 24 W
Faguibine, Lake .....16 45 N 3 54 W
Falémé, *river* .....14 46 N 12 14 W
Faro National Park ..... 8 30 N 13 00 E
Fatala, *river* .....10 13 N 14 00 W
Fazao–Malfakassa National Park .... 8 30 N 1 00 E
Feloú Falls .....14 21 N 11 21 W
Ferlo, *region* .....15 00 N 14 00 W
Ferlo-North Faunal Reserve ..15 00 N 14 00 W
Ferlo-South Faunal Reserve ..14 30 N 14 00 W
Fogo, *island* .....14 55 N 24 25 W
Fouta Djallon, *region* .....11 30 N 12 30 W
Gambia, *river* ....13 28 N 16 34 W
Gold Coast ..... 5 20 N 0 45 W
Gongola, *river* .... 9 30 N 12 04 E
Gotel Mountains .. 7 00 N 11 40 E
Grain Coast ..... 5 00 N 9 00 W
Great Scarcies, see Kolenté
Gréboun, Mount ..20 00 N 8 35 E
Guiers Lake .....16 12 N 15 50 W
Guinea, Gulf of ... 2 00 N 2 30 E
Hodh, *area* .....16 10 N 8 40 W
Hombori Tondo, Mount .....15 16 N 1 40 W
Iforas Massif .....20 00 N 2 00 E
Ijill, Mount .....22 38 N 12 33 W
Jong, *river* ..... 7 32 N 12 23 W
Jos Plateau .....10 00 N 9 30 E
Kaduna, *river* .... 8 45 N 5 48 E
Kainji Dam ..... 9 52 N 4 36 E
Kainji Lake, *reservoir* .....10 30 N 4 35 E
Kainji Lake National Park ....10 00 N 4 00 E
Kebbi, see Sokoto
Kébi, *river* ..... 9 18 N 13 33 E
Kolenté (Great Scarcies), *river* .. 8 55 N 13 08 W
Kom, *river* ..... 2 59 N 11 34 E
Komadugu Yobe (Komadougou Yobé), *river* .....13 42 N 13 20 E
Komoé, *river* ..... 5 12 N 3 44 W
Komoé National Park ..... 9 06 N 3 30 W
Kossou, Lake .... 7 00 N 5 30 W
Koussi, Mount ....19 50 N 18 30 E
Koutous Hill .....14 30 N 10 00 E
Kwahu Plateau ... 6 30 N 0 30 W
Lofa (Loffa), *river* ..... 6 36 N 11 05 W
Logone, *river* .....12 06 N 15 02 E
Loma Mansa (Bintimani), Mount ..... 9 13 N 11 07 W
Macina, *region* ...14 30 N 5 00 W
Maio, *island* .....15 15 N 23 10 W
Manda National Park ..... 9 35 N 18 12 E
Mandara Mountains .....10 45 N 13 40 E
Mandingue Plateau .....12 35 N 8 10 W
Manga, *region* ....15 00 N 14 00 E
Mano (Bewa), *river* ..... 6 56 N 11 31 W
Markala Dam .....13 42 N 6 05 W
Mbéré, *river* ..... 7 45 N 15 36 E
Mékrou, *river* .....12 24 N 2 49 E
Milo, *river* .....11 04 N 9 14 W
Mole National Park .....10 00 N 1 40 W
Mono, *river* ..... 6 17 N 1 51 E
Ngoko, *river* ..... 1 40 N 16 03 E
Niangay, Lake ....15 50 N 3 00 W
Niger, *river* ..... 5 33 N 6 33 E
Niger Delta ..... 4 50 N 6 00 E
Nimba, Mount .... 7 37 N 8 25 W
Niokolo–Koba National Park ....13 00 N 13 00 W
Ntem (Campo), *river* ..... 2 21 N 9 49 E
Nyong, *river* ..... 3 17 N 9 54 E
Ogou, *river* ..... 7 50 N 1 19 E
Oti, *river* ..... 7 37 N 0 01 E
Ouémé (Weme), *river* ..... 6 29 N 2 32 E
Outamba-Kilimi National Park .... 9 50 N 12 15 W
Pagalu, see Annobón
Palmas, Cape .... 4 22 N 7 44 W
Pendjari National Park .....11 00 N 1 30 E
Po National Park ..11 30 N 1 00 W
Pra, *river* ..... 5 01 N 1 37 W
Príncipe, *island* ... 1 37 N 7 25 E
Putu Range ..... 5 38 N 8 12 W
Raso Island .....16 37 N 24 36 W
Red Volta (Volta Rouge), *river* ....10 34 N 0 30 W
Rima, *river* .....13 04 N 5 10 E
Rokel (Seli), *river* .. 8 33 N 12 48 W
Rombo Islands ...14 58 N 24 40 W
Sahara, *desert* ....24 00 N 0 04 E
Sahel, *region* .....15 00 N 5 00 W
Saint Paul, *river* ... 6 23 N 10 48 W
Sal, *island* .....16 45 N 22 55 W
Salamat Wadi .... 9 27 N 18 06 E
Sanaga, *river* ..... 3 35 N 9 38 E
Sangha, *river* ..... 1 13 S 16 49 E
Sankarani, *river* ...12 01 N 8 19 W
Santa Isabel, Mount ..... 3 35 N 8 46 E
Santo Antão, *island* .....17 05 N 25 10 W
São Nicolau, *island* .....16 35 N 24 15 W
São Tiago, *island* .....15 05 N 23 40 W
São Tomé, *island* .. 0 12 N 6 39 E
São Vicente, *island* .....16 50 N 25 00 W

Sapo National Park ..... 5 20 N 8 30 W
Sassandra, *river* .. 4 58 N 6 05 W
Seli, see Rokel
Selingue Dam .... 11 34 N 8 10 W
Sénégal, *river* .... 15 48 N 16 32 W
Sewa, *river* ..... 7 18 N 12 08 W
Shebshi Mountains ..... 8 30 N 11 45 E
Sherbro Island .... 7 33 N 12 42 W
Sirba, *river* ..... 13 46 N 1 40 E
Slave Coast ..... 6 25 N 3 00 E
Sokoto (Kebbi), *river* ..... 11 24 N 4 07 E
Sota, *river* ..... 11 52 N 3 24 E
Sotuba Dam ..... 12 39 N 7 58 W
Tagant, *area* ..... 17 31 N 12 07 W
Taï National Park .. 5 30 N 7 25 W
Tamgué, Mount ... 12 00 N 12 18 W
Tanezrouft Desert ..... 24 00 N 0 45 W
Tano (Tanoé), *river* ..... 5 07 N 2 56 W
Tazerzaït Srhir Hill ..... 18 39 N 5 15 E
Tchigaï Plateau ... 21 30 N 14 50 E
Ténéré Desert .... 19 00 N 10 30 E
Tibesti Mountains ..... 21 30 N 17 30 E
Tiguidit Escarpment ..... 16 22 N 7 45 E
Tinkisso, *river* .... 11 21 N 9 10 W
Togo, Lake ..... 6 15 N 1 25 E
Udi Hills ..... 7 10 N 7 30 E
Verde, Cape ..... 14 43 N 17 30 W
Verga, Cape ..... 10 12 N 14 27 W
Vogel Peak, see Dimlang
Volta, *river* ..... 5 46 N 0 41 E
Volta, Lake, *reservoir* ..... 7 30 N 0 15 E
Volta Blanche, see White Volta
Volta Dam, see Akosombo Dam
Volta Noire, see Black Volta
Volta Rouge, see Red Volta
"W" National Park ..... 12 00 N 2 00 E
Waza National Park ..... 11 20 N 14 40 E
Weme, see Ouémé
White Volta (Volta Blanche), *river* ..... 9 10 N 1 15 W
Woleu, see Benito
Yankari Game Reserve ..... 9 45 N 10 30 E

oped, being generally sandy and free-draining over most of the crystalline and sedimentary formations.

Between the ferrallitic and ferruginous soils are a broad belt of ferrisols. These are less leached and slightly more fertile than the ferrallitics but less fertile than the ferruginous. They developed in association with the drier margins of the rain forests.

In areas with less than 20 inches of rainfall and associated with the northern Sudan savanna and the Sahel (Arabic *sāḥil:* "shore," referring to the region bordering the Sahara) savannas are the brown and reddish brown semiarid soils. Little leached, they contain free carbonates and chemically active clays and have a moderate to high CEC. Their fertility has supported many centuries of cultivation when supplemented by organic manure.

In addition to these four, several less extensive soils may be found. Eutrophic brown soils are developed on basaltic lavas and limestone rocks in the savanna regions. Their surface area is not great, but they are capable of sustained high yields when protected from erosion and supplemented by manure. The broad floodplains and coastal swamps contain immature hydromorphic soils. Although often high in their CEC, they are usually gleyed and seasonally waterlogged. In the Sahel within large depressions and formed from clay-rich materials are black vertisols. These are among the most fertile in western Africa, but they are difficult to till as they dry to a concretelike hardness in the dry season and turn into sticky, heavy muds in the wet seasons. Iron oxide crusts and pisolitic parent materials are also found in the savannas, and the soils found above them are thin, gravelly, and infertile.

Rainfall and Temperature Patterns in Western Africa

	Freetown	Lagos	Niamey
Annual rainfall (inches)	138	71	22
Rainfall in wettest month (inches)	37	17	7
Length of wet season (months with more than four inches)	7	6	2
Mean annual temperature*	81° (27°)	79° (26°)	84° (29°)
Mean daily temperature* in coolest month	77° (25°)	75° (24°)	75° (24°)
Mean diurnal temperature range* in coolest month	9° (5°)	13° (7°)	38° (21°)
Mean daily temperature* in hottest month	82° (28°)	82° (28°)	93° (34°)
Mean diurnal temperature range* in hottest month	9° (5°)	18° (10°)	34° (19°)

*Fahrenheit (Celsius)

Climate. Lying entirely within the tropics and sandwiched between the Sahara to the north and the equatorial Atlantic to the south, western Africa displays a gradual change in climate from hot, wet, and humid in the south to very hot and dry in the north. The climate is dominated by two air masses. One is derived from the quasi-permanent high-pressure cell generally located over the Sahara, from which are derived the hot, dusty northeast trades, or harmattans. In the winter months these winds occasionally extend over all of western Africa south to the coastlands. The other air mass is derived from a similar cell located over the tropical Atlantic. It is characteristically very humid and brings cloud cover, heavy rainfall, and high humidity to western Africa, which it progressively dominates from January to September. Separating these two air masses is the Intertropical Convergence Zone (ITCZ), and along it the denser humid southerly air penetrates beneath the dry easterly air.

Five climate zones

Using these two air mass systems, western African weather can be described in terms of five north–south zones. The northernmost zone, zone A, lies beneath the dry easterlies and north of the ITCZ. Temperatures, insolation, and evaporation are high; humidity is low; skies are clear, but dusty. Next south is zone B, immediately south of the ITCZ, with high humidity, clear skies, high temperatures, and small, puffy cumulus clouds but no rain. Zone C has a thicker moist air system, and in it develop thunderstorms and line squalls caused by air pressure variations in the dry easterlies aloft and by excessive convective heating over uplands. The thunderstorms and line squalls move westward, covering hundreds of square miles and bringing high winds, intense rainfall and, occasionally, hailstones, and exciting lightning displays. Zone D, in the thickest part of the moist southwesterly air system, exhibits deep cloud cover, persistent and often heavy rainfall, high humidity, and lower temperatures. Zone E is characterized by high humidity and cumulus clouds but little rain. Subsiding dry air above the moist surface winds suppresses convective rainfall, and at the height of the wet season the southern coast of western Africa experiences a "little dry season." This weather zonation moves northward from January to September and retreats southward from September to December. The climate of any spot depends upon this seasonal migration of the weather zones.

Most areas in western Africa experience a single wet season, whose duration decreases northward from 11 to 12 months along the southwest coast to three months in the north. Annual total rainfall similarly decreases northward from more than 120 inches (3,000 millimetres) between Monrovia (Liberia) and Conakry (Guinea) on the southwest coast to about 20 inches between Dakar (Senegal), Mopti (Mali), and Niamey (Niger) in the north. Mount Cameroon exhibits the greatest total annual rainfall (greater than 150 inches) because of its orographic effect on the moisture-laden southwesterly winds. Along the southern coastlands there is a double wet season separated by the "little dry season" of July to September, and annual rainfall ranges from about 30 inches at Accra to more than 120 inches along the coast of Cameroon.

Largely because of the decreasing extent and duration of cloud cover, regional patterns of daily temperature maxima and diurnal ranges increase northward. At specific locations temperature characteristics vary with the seasonal passage of the ITCZ and its associated weather zones. Zones A and B bring higher and more extreme ranges, whereas zones C, D, and E are accompanied by lower and more even temperatures. These rainfall and temperature contrasts are illustrated in the Table using data for three towns—Freetown, Sierra Leone, on the southwest coast, with a single long and exceptionally wet season; Lagos, Nigeria, on the south coast, with a double wet season; and Niamey, Niger, in the drier, hotter north.

Plant and animal life. Vegetation in western Africa has been so strongly modified during centuries of cultivation, grazing, and burning that there are now few areas of untouched primary vegetation. Traditionally, most western African agriculture has involved abandoning cultivated ground after several years to enable soil fertility to be rebuilt by naturally regenerating vegetation. Fires—to clear land for cultivation, to bring on a new grass flush for cattle, to drive animals during hunts, or started by lightning—also have affected the composition of virtually

Effects of cultivation on vegetation

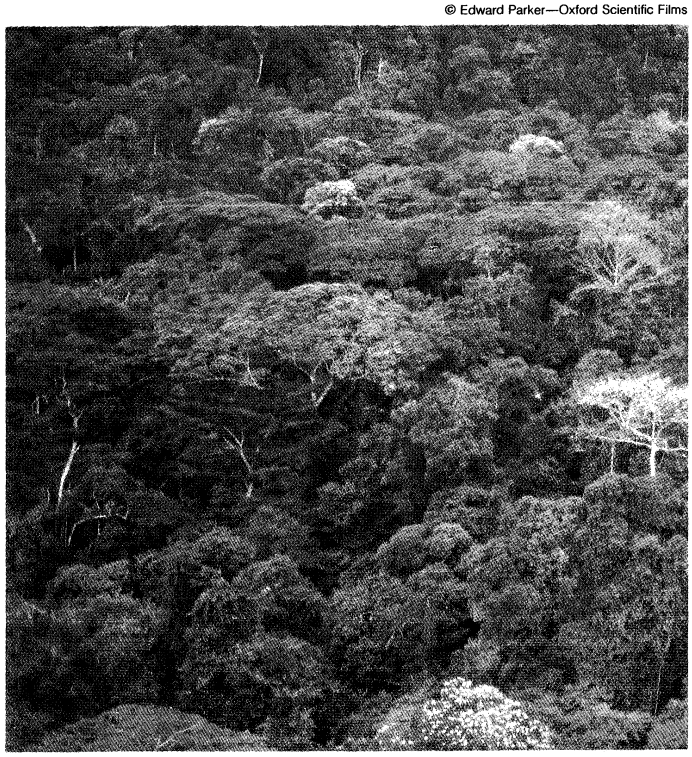
all the vegetation north of the rain forests. Indeed, the extensive savanna woodlands are regarded as fire climax communities, and vegetation in small sites protected from fire and cultivation is quite different in morphology and composition from the bulk of the surrounding vegetation. Consequently, in the more sparsely populated regions there are vegetation communities that may be not unlike the original ones, but elsewhere the landscape is a mosaic of farmland, fallow land, grasslands, grasslands with bushes and trees, open woodlands with grassland, coppices of dense woods, and forest blocks.

The progressive northerly reduction in rainfall, humidity, and length of wet season, and increase in evaporation is associated with clear vegetation changes characterized by species composition and community morphology. Rain forest in the south is replaced in the north by several types of savanna woodland.

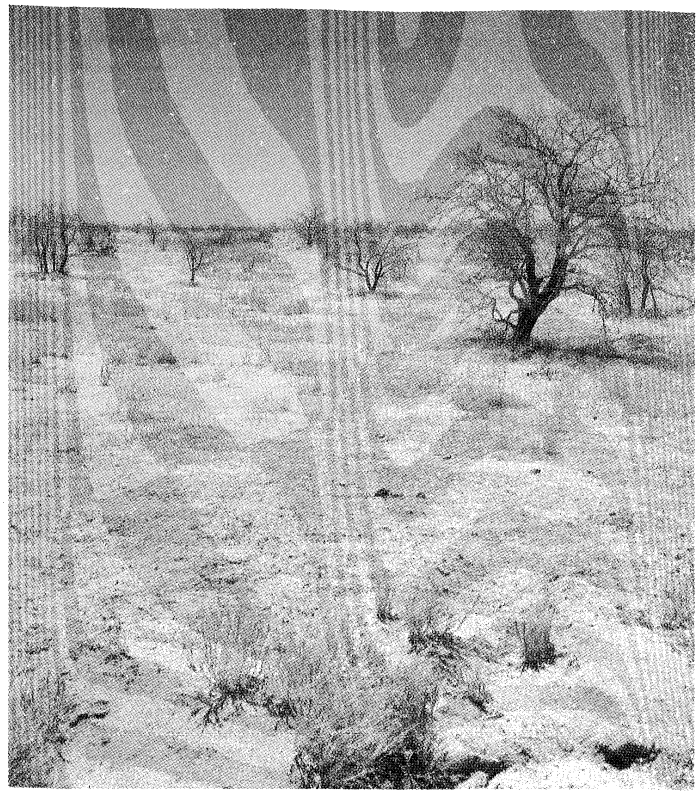
The tropical lowland rain forest is a species-rich community dominated by broad-leaved trees up to 200 feet high with slender unbranched, buttressed trunks. Saplings and ferns are numerous but the undergrowth is thin. The canopy carries climbing giant lianas and epiphytes. Many of the forest trees are commercially useful—*e.g.,* mahoganies, sapele, guarea, African walnut, iroko, utile, and obeche. This vegetation lies along the southern area between Cameroon and Sierra Leone, with the exception of a gap in southern Ghana. Large areas, however, have been cleared for agriculture. An impoverished but often denser secondary forest colonizes abandoned farmland. The brackish, muddy, tidal creeks of the coastlands support distinctive mangrove communities, and tall swamp forests or grasses, shrubs, ferns, and raphia palms grow in the freshwater swamps of the deltas and lagoons.

Originally the evergreen rain forest gradually merged, through an ecotone (transitional area) of mixed deciduous and evergreen forests, into the deciduous woodlands of the seasonal tropics. Today, however, the boundary between the rain forest and the savanna woodlands is very sharp and maintained by annual burning. The southernmost of the savanna woodland communities, called the Guinea savanna woodlands, is dominated by trees such as *Lophira, Terminalia,* and *Anogeissus;* fan palms; and tall tussocky grasses (*Hyparrhenia, Andropogon, Pennisetum* species) in the south. *Isoberlinia, Monotes,* and *Uapaca* characterize the northern parts. This formation extends into eastern

© Edward Parker—Oxford Scientific Films



Dense rain forest of the Guinea Coast region, Korup National Rain Forest Park, Cameroon.



Sahelian landscape near Zinder, Niger, typical of the western Sudan.
A.G.E. FotoStock

Africa, where it is colloquially called Miombo woodland. Today the zone is a mosaic of grasslands with and without isolated trees and palms, riparian rain forest, and dense coppices of deciduous forest.

Farther north, where the dry season is six to eight months long, are the Sudan savanna woodlands, characterized by shorter grasses and fire-resistant but agriculturally useful trees with wide spreading crowns growing 25 to 50 feet tall, such as *Acacia albida,* shea butter, locust bean, kapok, red iron wood, and baobab. This belt has been densely populated for centuries, and the natural woodlands have been replaced by extensive grasslands with occasional trees. In both the Guinea and Sudan savannas most of the trees are contorted from repeated fire damage.

North of the Sudan zone are the more sparsely populated Sahel savanna woodlands, where the dry season is more than eight months and cultivation is restricted to valley floors. Low, umbrella-shaped deciduous thorn trees (such as the *Acacia seyal*) and shrubs (such as *Commiphora africana*), succulents, and short, tussocky grasses give a distinctive vegetation strongly modified by burning for seminomadic cattle grazing.

The rain forests and the savanna woodlands are ecologically very different. The savannas have far less biomass, slower rates of nutrient cycling, fewer species and lifeforms, and harsher microclimates than the forests. Nutrient cycling is accomplished mainly by burning and macrofauna (termites and ants) in the savanna woodlands and by bacteria, fungi, worms, and termites in the rain forests.

Disappearance of wild animals

Long settled, and in some areas densely populated, western Africa is famed not for its wild animals but for its large flocks and herds of sheep, goats, and cattle, primarily in the savannas. Most of the indigenous larger wild animals—lions, leopards, cheetahs, serval, crocodiles, rhinoceroses, giraffes, elephants, antelopes, gazelles, bushbucks, buffaloes, and oryx—have been hunted to, or almost to, extinction during the past 200 years, and the wild dogs, hyenas, and jackals that were dependent upon them have also largely disappeared. Most countries now have conservation policies, and in national wildlife or game parks, such as at Yankari, Nigeria, some of the formerly more common species can be observed. Smaller species—

monkeys, chimpanzees, baboons, wildcats, duikers, hogs, cane and great rats, hyraxes, lizards, iguanas, and snakes—have not been so severely affected. Bird life, both resident and migrant species, is diverse and abundant, often to the extent of being an agricultural pest. Ants, termites, and insects are ubiquitous and play a fundamental role in plant nutrient cycling. Termite mound nests are a common and distinctive sight in the savanna areas. (M.B.T.)

TRADITIONAL CULTURES

People and languages

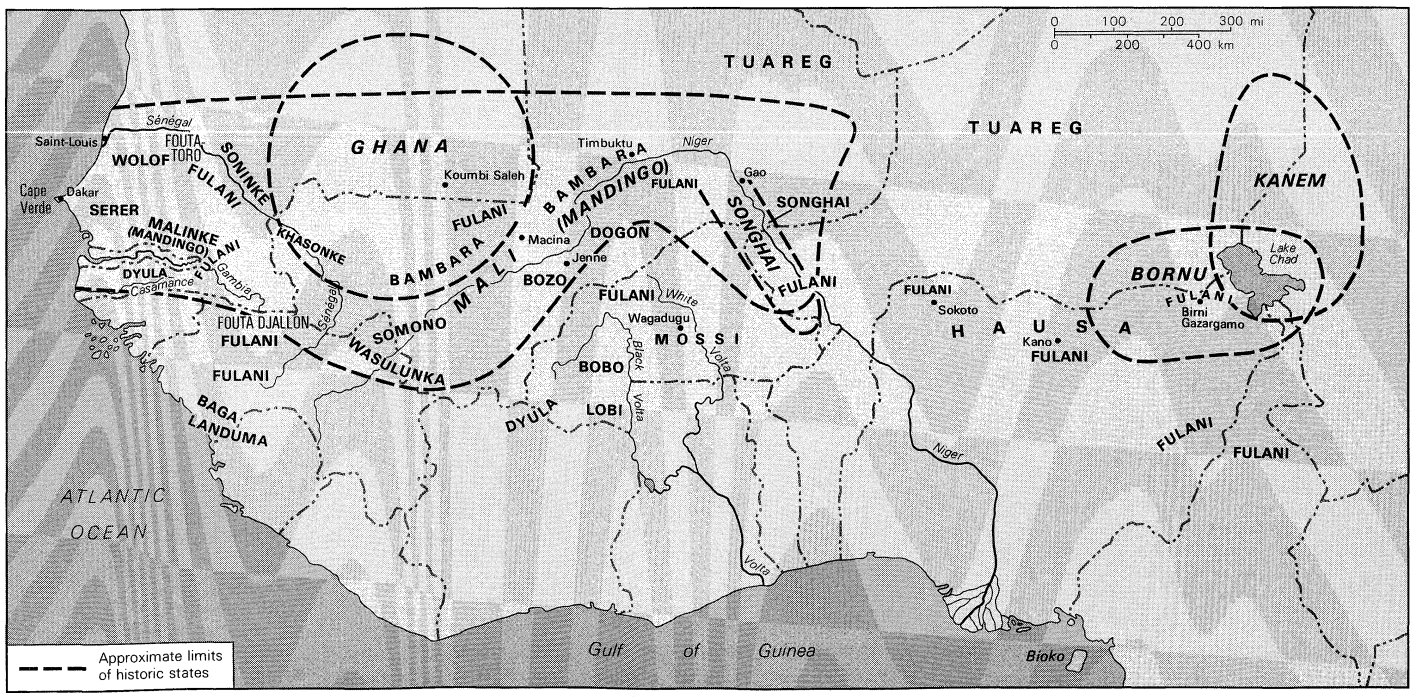
The western Sudan. The major ethnic groups today, as shown on the accompanying map, are the following: the Wolof of Senegal, the Serer to the south, and the Mande-speaking peoples to the east, comprising such subgroups as the Malinke, the Khasonke, the Bambara (Banmana), the Wasulunka, the Dyula, the Marka, and the Soninke (Serahuli). The Songhai are located largely in the region south of Timbuktu along the Niger, the Mossi are in the Volta basin, and a variety of smaller groups, such as the Dogon, Lobi, and Bobo, survive within the great bend of the Niger. Other small groups, such as the Diola (Jola), Landuma, and Baga, are to the southwest. The Hausa are concentrated largely in northern Nigeria, though they are scattered in all the major trade centres of western Africa. The Fulani (Fulbe or Peul) are distributed widely from the west Atlantic coast to Chad and Cameroon, though particularly concentrated in Senegal, Guinea, and northern Nigeria.

The continuous movements of people over the centuries have led to a complicated pattern of languages, but it is now held by some authorities that most of the languages should be considered as branches of one great Niger-Congo family. This would include the Mande, Voltaic, Kwa, Adamawa-Eastern, and West-Atlantic groups. The last includes such varied languages as Wolof, Serer, Fulani, and Diola. Linked to the Niger-Congo are the Kordofanian languages spoken in the area of the Nuba Hills. Other major families that have been distinguished are the Nilo-Saharan group, which includes Songhai, and the Afro-Asiatic, which comprises Ancient Egyptian, Berber, Cushite, and Hausa, among others. French is the language of communication among the elite of most nations of the western Sudan—namely, Senegal, Mali, Niger, Chad, and Guinea—but English is used in The Gambia, Ghana, and Nigeria.

Traditional culture patterns. Natural conditions in the Sudanic zone—drought, crop failures, epidemics of human and animal diseases—cause a great deal of uncertainty in peoples' lives, and they turn to the supernatural either in traditional rituals or in the Islāmic faith for reassurance and hope in time of trouble and the possibility of a greater reward in the next world. Life is also affected by the rhythm of the seasons, with a great contrast between the rainy season, the time of intense work on the farms, and the dry season, when the pace of life is slower. People must adjust their pace to natural conditions to gain the best advantages from them and must also be in harmony with the unseen powers behind them. Conditions may often be harsh, but farm work, though hard, is an honourable occupation, and the average inhabitant remains surprisingly optimistic and enjoys life to the full. Some of this feeling derives from the fact that a person does not face trouble alone but as a member of a group, linked to others by a complicated system of obligations—to kinsfolk, neighbours, and members of the same age group—maintained by constant visits, economic exchange, and mutual help at ceremonies. Everyone also feels links with the ancestors of the tribe. In general, the philosophy is one of bearing troubles patiently.

Local and territorial organization

Social organization. In the period from about AD 500 to 1470 the Sudanic zone was characterized by the rise and fall of a series of states and empires. The first to achieve eminence was Ghana (not to be confused with the modern state of that name), situated between the Sénégal and Niger rivers. It derived great wealth from trade in gold from the south and salt from the mines of the Sahara to the north. Ruins excavated at Koumbi Saleh are believed to be its capital, a town that could have contained 20,000 inhabitants. Ghana's power declined during the 11th century after nearly 20 years of attacks from the Almoravids, a Berber military and religious order from the Sahara, devoted to converting nonbelievers to Islām. The Mande-speaking people of Mali, on the Niger, developed the next great state, expanding rapidly in the mid-13th century, absorbing Ghana, and then gaining power over the trading cities of Timbuktu and Gao at the end of the major trans-Saharan trade routes. In the early 14th century the emperor of Mali, Mansa Mūsā, visited Egypt and Mecca. A large number of Arab scholars—teachers, lawyers, architects, doctors—established themselves in Mali at this time. After the death of Mansa Mūsā the empire began to break up. The city-state of Gao, under the Songhai, broke away toward the end of the 14th century and by the early 16th century had taken over control of the central region of the western Sudan. The power of Gao was extended over Timbuktu and Djenné, which were then at



Distribution of the peoples of the western Sudan and locations of major historic states.

their height as centres of trade, learning, and religion. The power of the Songhai, however, was broken in 1591 by an invading army from Morocco, whose firearms provided a great advantage over the swords and spears of the Songhai.

Farther east, the Chad region received various waves of immigrants—hunters, fishermen, and farmers who introduced weaving, bronze work, and pottery. They came under the influence of two states: that of Kanem, north and east of Lake Chad, which was powerful between the 11th century (when Islām began to make itself felt) and the 15th; and that of Bornu, to the west of Lake Chad, the dominant state in the 16th and 17th centuries. Bornu's army had a strong cavalry force, wore chain mail, quilted armour, and iron helmets, and retained its medieval splendour down to the 19th century, with something of its former pageantry still to be seen at Islāmic festivals.

Patterns of leadership

Society in all of these states was highly stratified, with a powerful ruling class controlling the wealth. A central ruler appointed regional governors or obtained the allegiance of outlying vassal chiefs, who were obliged to pay annual tribute and supply labour as needed. Well-organized armies both suppressed rebellion within the state and defended the boundaries against external enemies. War captives became slaves and performed much of the physical labour, carrying loads and working on farms. Islāmic religious teachers often formed part of the ruler's court, and gradually the people were converted to Islām. The ruler himself often filled a sacred role, as it was believed that the vital forces of the kingdom—rain, good harvests, and fertility—depended on him. The rulers were patrons of various arts and crafts, and the courts included musicians, praise singers, storytellers, goldsmiths, leatherworkers, and so on. Men of slave origin could rise to high rank as court officials and enjoy power over the freeborn.

Traditions reflecting the greatness of the former states have been handed down in songs and legends, and traces of old social patterns are to be seen in the behaviour of present-day chiefs. But the average person's primary allegiance is to the village rather than to a larger unit, and people are concerned with the ruler only when they have a court case or must pay taxes.

Villages are divided into wards or quarters, the nucleus of each ward being the descendants of an original settler, though as time goes by later settlers are absorbed. Often there is a meeting place for the men of the ward. Disputes within the village are commonly settled by the village head and the elders at a general meeting, or moot, in which the aim is to permit the parties to a dispute to state their grievances freely. The elders then arbitrate and seek to restore harmony and achieve a settlement, the meeting concluding with the group either praying together or sharing food or kola nuts imported from the south.

The village head is drawn from the clan that was the first to settle in an area and clear land. In general, rights of land ownership are determined in the first place by the act of clearing, and then the descendants of this person have the right of usage. Land is held in units large enough to support a family and is not fragmented. Where land is short, people break away and found new settlements; and migration also tends to occur under political pressure.

The pattern of descent since the coming of Islām has generally emphasized the male line (patrilineal descent), though links with relatives on the mother's side are also important—the mother's brother, for example, always giving help and support to his nephews and nieces. In families of slave origin, ties through the female line remain strong because the owner of the mother also owned the children born to her. In pre-Islāmic times, rights to rule could be transmitted through the female line, though the incumbent had to be male. People recognize kin groups in which they know the exact blood relationships (lineages) and that share rights over land and other privileges and obligations; but they also feel themselves related to larger groups (clans), descended from remote ancestors who bear the same name or observe the same ritual prohibitions.

Family and kinship patterns

In arrangements for marriage the family of the groom pays the family of the bride a sum of money known as the bride-price, which is returnable if the marriage fails through the fault of the wife. Ordinarily parents use the money received for a daughter to pay for a wife for a son. Households are normally larger than the nuclear family of husband and wife, for polygyny is permitted. The general pattern, however, is for each wife to have her own house, in which she cooks for herself and her children, taking her turn to cook for and sleep with her husband (though when a woman is pregnant or suckling a child she refrains from intercourse with her husband). A young widow will remarry, usually the brother of her deceased husband (the institution of the levirate), thus maintaining the link between the two families. If a marriage payment has been made and the fiancée dies, then a sister will replace her (the sororate). A widow who is old will often nominally remarry, but in practice be supported by her children, and often will stay with a son. The strongest emotional ties are not so much between husband and wife as between mother and children. An affectionate type of joking prevails between grandparents and grandchildren, and at weaning time small children are often sent to stay with maternal grandparents. A joking relationship also holds between cross-cousins (children of siblings of the opposite sex), and a cross-cousin has specific roles at naming, marriage, and burial rites.

Socialization and education

In African cultures ties between mother and child are extremely close. From birth to 15 months the child is carried on the mother's back; the child is fed at the slightest demand; and the mother plays with, sings to, and cuddles the child at every opportunity. Adults in general like children, hold them, protect them, and play with them. Grandmothers, in particular, are indulgent. The result is that the child grows up feeling valued and loved. There is, however, the very real threat of illness and death. Infant mortality, caused mainly by malaria, is high, and childhood illnesses—measles, whooping cough, influenza, pneumonia, and meningitis—take a heavy toll up to the age of five.

Children begin to take part in the work patterns of the community at an early age and are not as a rule kept separate from adults. Small children imitate adult activities in their games and then begin to undertake the lighter tasks. Girls help to pound grain with pestle and mortar, draw water, use winnowing baskets, fetch firewood, and so on. Boys drive off birds and monkeys from the farms and then help in weeding as they become older, until finally they are able to undertake a young man's share of cultivation. Among cattle-keeping groups, boys learn their adult roles by handling calves.

Initiation ceremonies

Initiation ceremonies mark the transformation from the status of child to that of adult. For males the ceremonies are generally associated with either circumcision (though many Muslims now have their sons circumcised in infancy) or, in some areas, scarification; in the past, death through infection was not uncommon, though now modern antiseptics may be used. The ceremonies involve both a physical separation, the initiates living outside the village in the bush, and a ritual death, followed by rebirth when they return to village life. During initiation they are harshly punished for their faults (sometimes for crimes committed before initiation, such as stealing); they are instructed in the traditions of their society and learn secret means of male communication, unknown to women or children. The initiates remain in seclusion for a period varying from several weeks to several months, and those who have gone through initiation together, irrespective of their age, retain close ties throughout life. In many societies there is also female initiation involving clitoridectomy. While in seclusion the initiates receive training for their future marital role.

There is a marked division of labour between the sexes: the women are concerned with preparing food, caring for the children, drawing water, washing clothes, making pottery, dyeing cloth, gathering leaves, fruits, and firewood, and so on, whereas men are concerned with looking after large animals such as cattle, horses, donkeys, and large sheep and with hunting, clearing land for agriculture, fishing, butchering, house and fence building, woodcarving, leatherworking, and smithing. Carding cotton and spinning are women's work; weaving is men's. Women have their own meeting place at the well or stream; men have a

place in a village square. The farming tools used are often different for men and women.

Stratification

Although egalitarian relationships are found in many non-Islāmic groups, such as the Diola and some pastoral Fulani, a system of social stratification is characteristic of most western Sudanic peoples. The essential pattern consists of such categories as (1) the royal families, often deriving from foreign conquering elements; (2) the nobles, members of high-ranking lineages who constitute the military leaders or provincial governors and who may have the power of electing the ruler from among suitable royal candidates; (3) the freeborn, who are landowners and farmers and sometimes traders; (4) the people of slave origin, who are objects of social discrimination or differentiation despite official abolition of the slave trade; and (5) the casted craftsmen, which include musicians, praise singers, professional storytellers and entertainers, smiths (whether blacksmiths, goldsmiths, or silversmiths), leatherworkers, and certain types of woodworkers. These craftsmen can marry only within their own category, cannot lose their status even if they abandon their trade, and are looked down upon by the freeborn, even though they may be richer and perform crucial services. The women of various caste groups are often expert hairdressers and tattooers. Islāmic scholars and their families are accorded high prestige.

Officials appointed by rulers could be either freeborn or of slave origin, with the result that "slaves" often came to enjoy greater power than the freeborn. Today there is also differentiation based on wealth acquired by trade and on status in positions of government service. Since independence from colonial rule, those who have attained prominence in political parties or the military have reached the chief positions of power.

Settlement patterns and housing

Economic systems. The characteristic settlement is a concentrated village consisting of fenced-off clusters of houses (compounds) occupied by members of a lineage and their spouses. Because for hundreds of years villagers led an uncertain life, liable to ravage by invaders and slave traders, many villages were built on sites afforded some protection by rivers or fortified by earthen walls. These old fortifications have almost all disappeared except for the great walled cities of northern Nigeria. The general trend during the 20th century has been for smaller and more widely dispersed villages, as people have cleared more land for agriculture, but as a result of periods of drought in the 1970s and '80s many people have resettled in the larger urban centres.

Among the Wolof the average population of a village is only about 100, and the compounds are built around an open village square. In the past the houses were generally circular, with walls of either millet stalk or reed, but now a rectangular form is more common, with mud walls and roofing often of imported zinc or aluminum sheeting. Fulani settlements tend to be even smaller and consist of temporary huts with walls of reed or of matting woven from grass. Malinke villages, on the other hand, tend to be larger, with some 1,000 to 2,000 people; the houses are generally mud-walled with a conical roof thatched with grass. In the north the Tuareg are tent dwellers. In general, the most common western Sudanic house is circular, but the square or rectangular form is not necessarily recent, for excavations of old towns in ancient Ghana have revealed rectangular houses.

In addition to the farming villages, the western Sudan is characterized by very large towns, either the capitals of the old states or trading towns at the southern end of the various trans-Saharan trade routes, such as Kano, Nigeria, and Timbuktu, Mali. There, architecture is often influenced by North African forms, mosques and fortresses in particular being designed after Arab patterns. At present, with the availability of imported sheeting for roofs and locally manufactured cement, richer people are building more permanent houses that are square or rectangular.

Few traces are now to be found either of purely hunting and gathering groups or of fishing communities, except for such peoples as the Somono and Bozo fishermen of the Niger. Most inhabitants are agriculturalists, dependent on the cultivation of such crops as millet and sorghum for food and peanuts (groundnuts) for cash. The system is one of shifting cultivation: the savanna land is cleared by ax and cutlass, the residue burned, and the crops planted. The hardest work is the continuous backbreaking struggle against weeds, the principal tool being a small hand hoe, and when the crops are ripening, they need to be protected against swarms of birds. Crop rotation is customary, but the soil nevertheless becomes exhausted after several years of cultivation and is then allowed to revert to bush. Only patches around the villages, fertilized by household rubbish, animal manure, and so on, can be kept under continuous cultivation. In such patches tomatoes, peppers, beans, eggplants, maize, and condiments can be grown. On good soils cotton cultivation is possible.

Patterns of agriculture

The single rainy season demands a considerable amount of extremely hard work in a very short period of time, and irregularities in the rainfall, whether drought or cloudburst, can cause crop damage and a food shortage. Farmers suffered losses through plagues of locusts, and despite international efforts at control these plagues still occur. People supplement the cultivated crops, particularly in times of shortage, with wild fruits and roots. The fruits of the baobab (*Adansonia digitata*), nete (*Parkia biglobosa*), and tamarind (*Tamarindus indica*) are particularly important, while over much of the more northern zone the shea-butter tree (*Butyrospermum parkii*) is a source of vegetable oil. In regions with a more abundant water supply, like the Gambia and Casamance valleys, rice is the dominant food crop.

The dry season is a period devoted to visiting friends, trading, building houses and making repairs, engaging in arts and crafts (pottery, mats, textiles, basketry), and participating in such ceremonial events as marriages and initiation ceremonies.

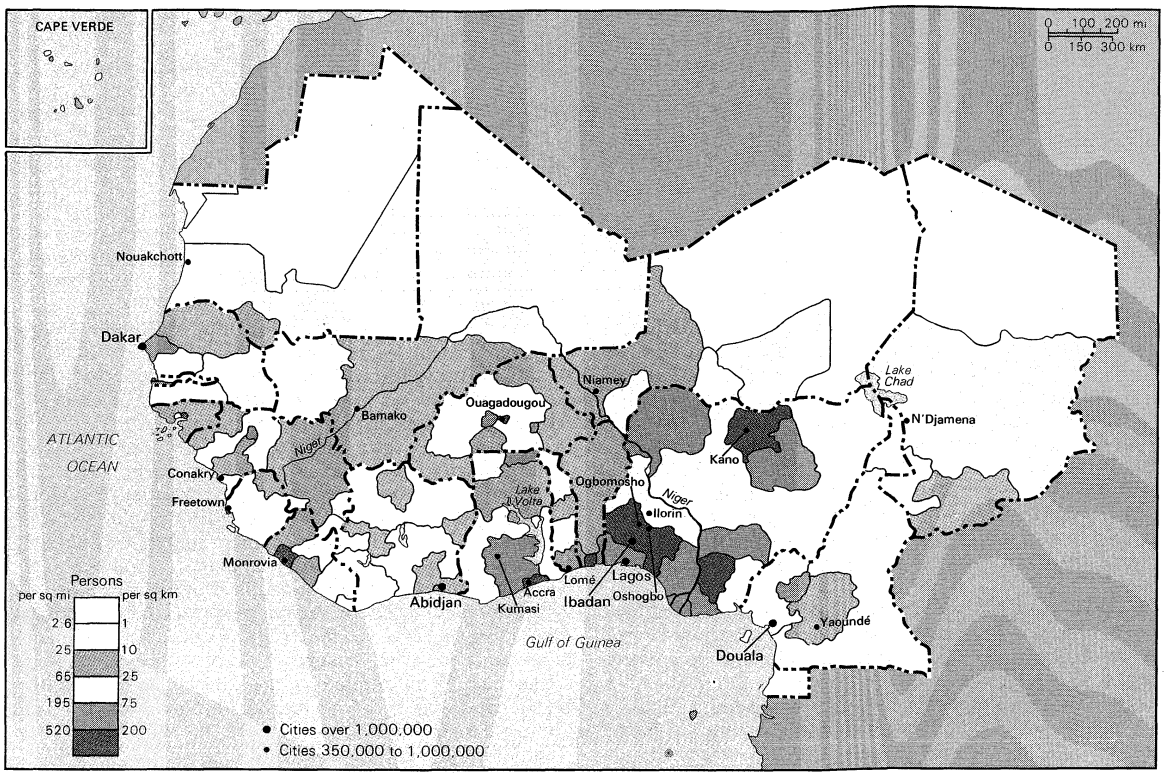
Pastoralism

Another important life-style is that of the pastoralists. The central region of the western Sudan, with its extensive grasslands, provides opportunities for raising large herds of cattle. The Fulani are the main cattle owners and are found from the west Atlantic coast to Chad and Cameroon. Some are entirely nomadic, living in temporary shelters and moving between wet-season and dry-season grazing areas. Others stay in fixed villages, build more permanent houses, engage in farming—producing excellent crops because of the cattle manure—and still maintain large herds of cattle. The men are concerned with the care of animals, determining the time for movement, selecting grazing areas, and seeing that the animals are watered. Women's occupations include domestic tasks, drawing water for household use, collecting firewood, cooking, and processing and selling milk. Both the watering places and markets in the towns are important meeting places where the Fulani exchange vital information about the state of pastures, water, animal diseases, and politics. The cattle population has increased rapidly as the result of inoculations against such diseases as rinderpest, but the Fulani feel personally attached to their cattle, in which their wealth and well-being are involved, and are reluctant to sell animals for slaughter.

Exchange system

The western Sudan has always been a highway and crossroads for long-distance trading. Today, manufactured goods entering the Atlantic ports, including cloth, medicines, and transistor radios, are taken inland. Sudanese traders travel to Sierra Leone, Ghana, and the Congo basin to deal in diamonds, which eventually find their way to the Middle East. Gold still crosses the Sahara to North Africa, and salt from the Saharan mines comes by camel caravan to the Niger River. Caffeine-containing kola nuts from the Guinea Coast rain forests are transported north in large quantities; they are in great demand among the farmers for assuaging hunger and thirst and for use in ritual and sociable contexts. Hausa leatherwork is widely traded.

The major trade routes across the Sahara are supplemented by lateral routes along the Niger and along the Sénégal and Gambia river valleys. For certain commodities a system of direct barter prevailed; Fulani herdsmen, for example, exchanged milk and butter for grain. Along the Niger is a substantial trade in dried fish, either bartered for grain or sold for cash.



Population density of western Africa.

Animistic beliefs

Belief systems. Indigenous systems of belief unaffected by Islām involved the concept of the essential unity of the visible and invisible worlds, humanity being accorded the dominant position in the system. Forces in plants, animals, and minerals are made known to humans through ancestors and can be used for either good or evil, humans having the moral responsibility for making the choice. Living persons are a continuation of the life stream of the first beings. Ancestors watch over the living and act as intermediaries between them and the creator of the universe, who is now remote from humans, even though his power is supreme. The ancestors indicate their wishes through dreams sent to the elders, while the living communicate with the ancestors through prayer and sacrifice, the blood of sacrificed animals setting in motion certain latent spiritual forces. The spirit world is also evoked by persons wearing carved masks during special ceremonies; associated dances and drumming cleanse, reinvigorate, and protect the community. The masks themselves are the abode of spirits, and carvers feel inspired by supernatural powers.

Religious beliefs reinforce the traditional values of society, for it is believed that lack of harmony in the community and breaches of traditional law and custom are followed by such disasters as drought, disease, and crop failure. Society is threatened by forces outside the community—evil spirits that are believed to cause mental disorders and physical abnormalities—and by people inside the community in whom evil grows—witches, who can cause harm to both human beings and crops through a witch substance inside them, and sorcerers, who perform deliberate acts of evil magic. Charms are worn and protective devices set up to guard against such dangers, while diviners seek to detect both witches and sorcerers. Diviners are consulted by individuals with problems; it is their role to trace the causes of troubles, using such techniques as casting cowrie shells or reading patterns in sand to see into the spiritual world, and then to indicate the proper measures to be taken. Diviners provide treatment at the physical level by prescribing herbs and medicines, at the psychological level by listening to confessions and providing reassurance, and at the social level by trying to disperse tensions between individuals.

Influences of Islām

Islām has spread widely throughout the western Sudan, and between 60 and 70 percent of the people are nominally Muslim. Most attend services at the mosques, observe Ramaḍān (the month of fasting), say the daily prayers, and give alms generously; and a few are able to make the pilgrimage to Mecca. Wherever Islām is the dominant faith, Muslim religious teachers have taken over the role of traditional diviners in determining the causes of troubles, and they provide remedies in conformity with Islāmic patterns. Children are given religious instruction in which they learn the prayers, recite long passages from the Qur'ān, and acquire the rudiments of Arabic writing. The traditional ritual dances and masked performances are gradually disappearing or have been greatly modified as a result of opposition from Muslim teachers. Nevertheless, many of the traditional beliefs about spiritual beings still remain relatively unchanged in people's minds. Christianity has had little effect in the western Sudan, except marginally in the coastal cities of Senegal and The Gambia.

Oral literature. The region is extremely rich in oral literature: proverbs; myths of origin; animal stories in which the lion, hyena, and hare play prominent parts; epics; and tales about people (neglected orphans, disobedient children, the rivalry of co-wives, jealous husbands, deceitful wives, unjust rulers) and about supernatural forces (encounters with good and evil spirits, the terrible deeds of witches and sorcerers). Narration involves both entertainment and education, for the young learn something of the values of the community and acquire knowledge of approved and disapproved behaviour. There is always a high degree of audience involvement, the listeners either replying to the narrator's questions or singing the rhythmical choruses that form part of the narrative style. Storytellers act the roles of the various characters with great effectiveness, and the narration becomes a highly rhythmical performance leading up to a dramatic climax.

Evolution of the cultures. In general, the western Sudan has been slow to change, largely because of long distances and problems of communication and because of the generally low level of income throughout the area. Migrants from the interior have gone south to work in the mines and cocoa plantations of Ghana or in the coffee, cocoa, and banana plantations of Côte d'Ivoire or to cultivate peanuts in the Sénégal and the Casamance river valleys; but for the most part they revert to traditional ways when they return home.

The new town dwellers

In moving to the cities and towns of the coastal zone, migrants who are from the same ethnic group or village have a tendency to live in the same area of town and tend to spend their leisure time together. Disputes that arise between people of the same ethnic group are ordinarily settled by the elders of that group according to traditional law and custom and are not taken to the state court. Voluntary associations are formed for mutual aid and entertainment.

In general, Islāmic influence has continued to spread slowly and steadily among animistic groups, for a Muslim has higher status than an animist outside his own community.

Change in agriculture has been slight. Mechanized agricultural projects have usually been unsuccessful because of the high costs of maintaining machinery. In Senegal, The Gambia, Mali, and Niger, light plows and weeding equipment drawn by donkeys, horses, and oxen have become popular and have led to an increase in the production of the cash crop, peanuts (groundnuts). The growth of cooperative societies has enabled farmers to receive greater cash benefits, and the improvement of roads and the increase in motor transport have made the marketing of produce easier. Gains have been offset by years of low rainfall, however.

In Niger and Chad, spectacular success has followed programs of water prospecting. Artesian water is now tapped by boreholes, and artificial water holes have been created, designed to store water during the dry season. This has enabled the seminomadic cattle owners to settle and make use of the rich pastureland available, the limiting factor previously being the lack of water for cattle. In the 1970s and '80s, however, recurring droughts caused heavy loss of livestock and led to desertification of overexploited marginal lands. (D.P.Ga.)

The Guinea Coast. Guinea is a term used originally for the coastlands and adjacent forests of western Africa between the Republic of Guinea on the west and Equatorial Guinea on the east, including the whole, or the southern parts, of Guinea-Bissau, Sierra Leone, Liberia, Côte d'Ivoire, Ghana, Togo, Benin, Nigeria, and Cameroon. There have been conflicting accounts of the derivation of the name Guinea, but it would seem to be a version of the Berber word *aguinaw,* or *gnawa,* meaning "black man," or "Negro."

The environment and the people. In western Africa, in the general absence of major mountainous areas, natural regions are determined primarily by climate and vegetation, and the Guinea Coast societies are those that have in the past been associated with the equatorial forest zone. This forest has long been cleared for agriculture in some areas and has been subject to increasing and widespread threat over recent decades, but it is the natural vegetation of most districts within 100 miles of the coast. In the east, heavy forest formerly extended from the borders of the Cameroon highlands to the area west of the Niger River. In the west, forest stretched from Sierra Leone to western Ghana. Between these two belts of forest there is a drier region, where for centuries tree cover has been thin; once cleared for farming the forest does not regenerate, and even areas left fallow for long periods remain as grassland. Societies in this area, while culturally similar to true forest societies, have historically been different in significant ways.

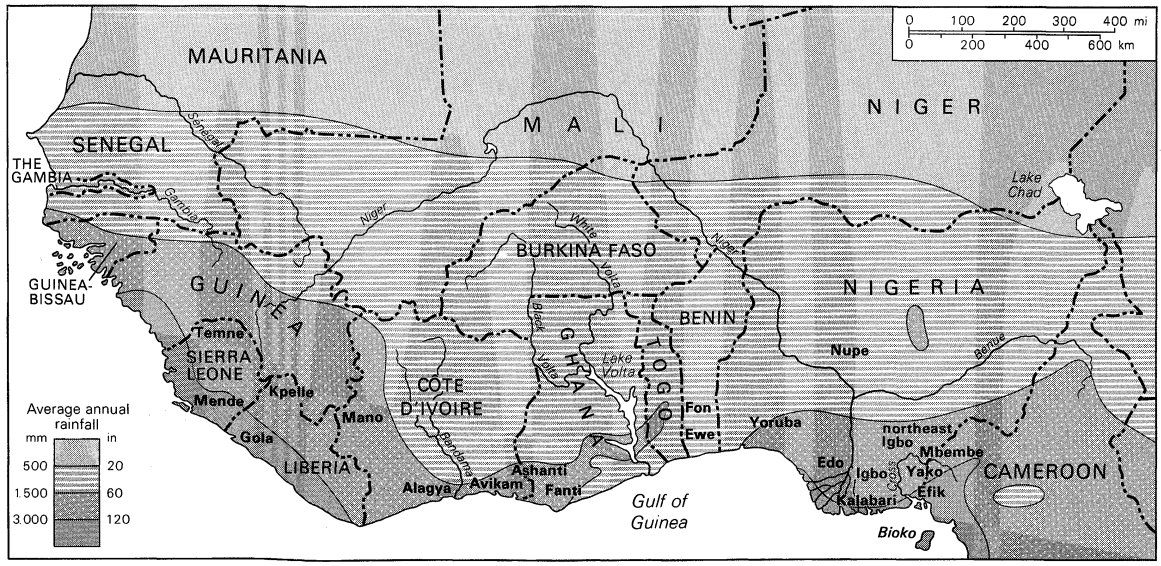
Cultural effects of the forest

The forest greatly influenced the cultural development of the Guinea Coast by affecting the movements of peoples and the development of agriculture and commerce. People occupied the forest areas relatively late because farming there had to await the development of suitable tools and crops. Iron axes are needed to clear equatorial forests and only with the introduction of the shade-tolerant crops—the plantain and the cocoyam (taro and eddo), brought from Asia in the 9th century AD—could forest farming become an economic alternative to hunting and gathering. Moreover, until recently forest farming did not include animal husbandry, for the forest harboured species of tsetse fly that are particularly dangerous to cattle and horses. This had advantages for the forest people, however, because the tsetse fly and the dense vegetation protected them from marauding cavalry. Gradually the forest gave its inhabitants commercial advantages: kola nuts and, later, palm oil were so highly desired by distant peoples that traders by sea, and overland from the north, were drawn to the Guinea Coast.

Language distribution

The cultural significance of the original forest environment is shown even today in the linguistic map of western Africa. In detail there are many different languages in the forest area, some spoken by millions, some by a few thousand people. What is striking, however, is that the boundaries between the major language families roughly coincide with the old boundaries of the forest. Only in the extreme east does this clear division disappear. The linguistic division between the forest societies and the hinterland is good evidence for the long historical distinctiveness of the Guinea Coast cultures.

Cultural patterns. This section cannot deal individually with all the groups in the area but only the more important or better-studied groups, for Guinea Coast societies vary enormously. Today many similarities are due to patterns of development in the colonial and postcolonial periods; nevertheless, precolonial variations still show themselves. This is true even in such obvious ways as population densities and types of settlement away from modern cities. Even within a small area, such as southeastern Nigeria, great variations exist. Prior to the Biafran war (1967–70) in certain Igbo (Ibo) areas there were 700 or more peo-



Distribution of peoples and of rainfall in the Guinea Coast region.

ple per square mile (270 or more per square kilometre), whereas in the equally fertile forest hinterland of the Cross River densities were well under 100 people per square mile. Moreover, the Igbo settlements were characteristically spread out through their cultivated lands, whereas the Cross River peoples to the east of the Igbo lived in large nucleated villages. To the west of the Igbo, the Yoruba built in precolonial times some of the largest indigenous towns in Africa.

Interplay of environmental and historical influences

Such differences are usually caused by the complex interplay of environmental and cultural factors and historical contingency. Even in the choice of crops, which might seem to be more or less dictated by physical conditions, culture is very significant. It is true that in the eastern, tuber-growing areas the choice between one tuber and another may be determined ultimately by environmental factors. For example, the modern selection of cassava rather than yams is usually the result of the greater ability of cassava to tolerate depleted soils. There is a major crop boundary, however, roughly corresponding to a linguistic boundary along the Bandama River in Côte d'Ivoire; to the west rice is the staple crop but to the east tubers are the staple. This seems clear evidence of the significance of cultural choice. The importance of culture has also been shown by recent work on the agricultural skills of local peasants that has shed light on the ingenuity with which indigenous farmers develop new strains of preferred crops and adapt their techniques in order to get reasonable yields even under adverse conditions. Their techniques are often clearly superior to those advocated by official development agencies.

Historical background of trade and politics. The more complex the cultural patterns the more complex is the interplay of environmental factors and historical contingency behind them. The evidence is to be found in the political variations of the indigenous units of the Guinea Coast societies. At the end of the 19th century there existed tiny, independent political groups whose inhabitants had, for many purposes, solved the Hobbesian puzzle of how to avoid the "warre" of all against all without having recourse to the development of the state. At the same time there also existed powerful kingdoms. Some were of medieval foundation, and some had been developed in recent centuries, but the rulers of the largest held sway over hundreds of thousands of people. Environmental factors were obviously of significance in influencing political growth. The authority of a central government depends on the "message-delay" time involved in conveying commands to the periphery of the state. In the forest, where riding horses could not be kept and passage on all but the largest rivers was repeatedly blocked by huge fallen trees, political power might depend, as in 19th-century Asante (Ashanti), on enormous efforts put into keeping roads open for runners. The success of some of these states is particularly notable for the triumph of human ingenuity over adverse natural conditions.

Shift from Arab to European trade

Before the end of the 15th century, most of the region's external contacts were made through the savanna kingdoms to the north, whose merchants wanted slaves, gold, and kola nuts (a stimulant lawful for Muslims). From the end of the 15th century, however, the interests of the Guinea Coast peoples were partly reoriented toward trade with European merchants, who sought successively gold, slaves, and palm oil. European trade was significant partly because, unlike the northern trade, it was controlled within the Guinea area entirely by local people. Europeans were prevented from penetrating inland by climate, disease, and the express action of African authorities. The merchants at the coast provided inducements to sell; how their wants were supplied was a matter for local traders. These overseas merchants were also significant because of the nature of the goods they brought to sell. They were mainly consumer goods, but they also included such capital goods as iron, guns, and gunpowder, and these gradually introduced a crucial new factor into local warfare. Political authorities were forced into trade because it became militarily vital to acquire the new weapons. Even imported consumer goods had a political significance, for it was normally the political authorities, able to tax European merchants and the local traders, who could acquire most goods. They thus had the best resources with which to exert general influence; consequently new power was put into their hands with important political consequences.

The development of Guinea Coast societies was radically influenced by the nature of their exports, especially the slaves. Slave trading did not everywhere lead to raiding, but where it did it led to changes in military and political organization. In Dahomey a strong government sent its army to raid slaves in every dry season; in Yoruba areas one factor in the violent relations between the city-states in the 18th and 19th centuries was probably their involvement in the trade, and new political forms were bred in response to this situation. Wherever war captives were traded as slaves, those central authorities who controlled captives developed an economic advantage over their rivals.

In the 19th century palm oil gradually became the most important Guinea Coast export because of its increased use as an industrial lubricant and because European humanitarians were successful in applying political pressure on their governments, especially the British, to end overseas transport of slaves. Nevertheless, slave dealing and raiding remained important internally throughout the century. Even when slave exports declined, the growing palm-oil traffic in itself stimulated these activities, for the transport of the bulky oil required slaves to paddle the canoe transports or to headload the oil to the ports. Moreover, a trader using many men to transport oil needed others, often slaves, to produce food for them; thus, slaves were important economically and politically to the Mende of Sierra Leone in the west and the Fanti (Fante), Dahomean, Yoruba, Niger delta, and Efik peoples in the south.

Effects of the slave and palm-oil trade

The growth of the palm-oil trade brought other economic and political changes. African traders exporting palm oil needed more capital than slave traders did, because slaves transported themselves and worked while awaiting shipment, whereas oil was expensive to transport and constituted idle capital while at the ports. An unanticipated result of the change to the oil trade was, therefore, that the exporters, needing capital, became increasingly reliant on European firms that advanced them goods on credit and therefore took increasing interest in local political affairs. This was one factor leading to colonialism. Paradoxically, in the rural areas many men who could never have been slave traders could easily gather and sell palm produce. Apart from the exceptional case of Dahomey, where most palm oil came from large government-fostered plantations, oil was drawn mainly from the fruit of trees growing naturally in the bush. In the eastern forest areas especially, participation in this production and trade was very general.

Markets and traders

By the end of the 19th century a network of local markets had been developed over much of the hinterland of the Guinea Coast. The great centralized kingdoms were naturally associated with great trade routes, but there is evidence that traders were protected by the common interest of many people—and almost all political authorites—that the routes should be kept open. It is remarkable that for the most part, even in the absence of strong governments and in areas where adjacent peoples might be regarded as fair game for attacks, the accredited trader passed unharmed; appropriate punishment was dealt out by local authorities to any person who robbed or injured them.

Kingdoms and chiefdoms of Guinea. Although trade could flow across political boundaries, the political development of western African societies was much influenced by the growth of trade and through the warfare and the struggle for trade routes that accompanied it. Where trade was limited, most political units seem to have remained weak and small in scale. This was true in the earlier part of the 19th century in much of the area of Sierra Leone and Liberia and is a major reason that it was possible to settle freed slaves there without their being dominated by indigenous powers. Later, however, hinterland Mende tribes, located in rich oil-palm areas, fought for the control of trade routes and developed into more centralized groups under warrior chiefs. At the other end of the western forest, and at the other extreme as regards trade, the

Asante Confederation of kin-based states developed much earlier and quite differently. The area had early commercial importance in the northern trade as a source of gold and of the best kola nuts. From the 17th century the Asante exploited their gold resources, which were easily made a government monopoly, in order to gain local control over the import of firearms. Extending their influence over their immediate neighbours by diplomacy and over those more distant by warfare, they eventually subjugated peoples as far southeast as Accra, on the coast, and as far north as the savanna. It has been shown that in the 19th century the authorities—with remarkable insight into their own social structure, which was based on matrilineal clans—created a semiprofessional civil service in which offices were passed from father to son; this ensured that new officials were trained but that the offices escaped the clutches of the major kin groups.

In the eastern forest, northwest of the Niger delta, was the kingdom of Benin, whose rulers claimed to have come originally from the Yoruba area. It was of medieval origin and was so well established in the late 15th century that the king of Benin sent an emissary to the king of Portugal, who in turn sent missionaries to Benin. Its internal political development from that time involved complex power struggles between the party of the ruler, the oba, and the nobles. In theory there existed a complex balance between different interest groups, but that balance shifted in different generations. Interestingly, it seems that the central authorities were so well aware of the dangers of allowing power to pass to locally based kin groups that much ingenuity was exerted in creating structures that ensured that commoners' kin groups did not develop and that politically ambitious individuals had to seek advancement by moving to the capital of Benin and could not create power bases in their home areas. In this the social structure of Benin was almost the reverse of the adjacent Yoruba areas.

To the east of the city of Benin lies the Niger delta, one of the greatest mangrove swamps in the world, an area without land for farming or any resources needed for the development of large states (until the discovery and exploitation of petroleum in the mid-20th century). In the 18th and 19th centuries, in that area and farther east at the mouth of the Cross River, there developed small, independent trading settlements. At first most were villages exchanging fish for agricultural produce; later, wherever deepwater anchorages existed—suitable for European ships and close to rivers giving good access to the interior—these settlements became large trading centres interested in exerting commercial control over the palm-oil-rich and slave-rich hinterlands.

Throughout much of the Guinea Coast the king or chief was the keystone of the political system because, although his actual powers varied enormously, his ritual relation to his predecessors and, usually, the gods provided the ideological framework for that system. In some cases he was the only appropriate intercessor with his deceased ancestors who were believed to exercise a controlling influence over group affairs. In other cases he was transformed, by his installation rites, into a person so sacred that all his actions had to be circumscribed lest, by breaking taboos, he brought disaster on his people.

Political structure of the kingdoms

In Oyo, one of the best-described Yoruba kingdoms, the king, at the culmination of his installation rituals, ate the heart of his predecessor and was transformed into a personification of his ancestors. Thereafter, on his only public appearances, at rituals held three times a year, he appeared veiled, his face hidden by a beaded fringe. Those who formally represented him in judicial, religious, military, and administrative capacities were slave eunuchs, chosen because, having neither kin nor affines, they were presumed to have no interests to serve but their master's. Although secluded, it appears that the king was involved in important political maneuverings, playing one group of hereditary chiefs off against a second and trying to avoid the great danger that would ensue if both groups were to unite against him.

Such political structures can be described as if they were frictionless systems of checks and balances persisting unaltered for generations, but modern research suggests that these structures were changed in detail whenever the balance of political power shifted. Points of particular struggle were the rules for choosing the successor to king or chief. In polygynous societies even a rule of inheritance by the eldest son does not necessarily indicate the true heir as there is room for dispute over the status of the mother. Any rule that widened the choice—*e.g.*, to any member of a lineage—gave increased powers to the selectors. Furthermore, military success could bring great problems, for if new territories were conquered there might be great competition between king and barons over claims to control the offices essential for the administration of these areas.

Variety of chiefdoms

These structures, based on similar beliefs in the ritual powers of chiefs, are to be found in many chiefdoms of the Guinea Coast, even very small ones. Nevertheless, there were some political units that might be called essentially secular. Mende chiefs, for example, were explicitly leaders whose rule was based on military prowess. In Niger delta and Efik towns senior priests had ritual headship but lacked any political importance, for power was held by rich traders. Leadership in many Igbo villages often lay de facto with wealthy men who were members of influential societies; there were few formal political offices, and decisions were reached through public discussions at village meetings. It is not coincidental that even in modern times attitudes to social hierarchy are markedly different between the Yoruba and Igbo and that the latter adopted with particular enthusiasm the patterns of "democratic" politics.

Kin groups and other associations. Kinship ties are almost invariably of great significance in all traditional African societies, and the Guinea Coast is no exception. For the individual, ties through both the father and the mother are significant, but for inheritance and for political and legal purposes kin groups were commonly organized by singling out a particular line of descent. Such kin groups still exist in this area, but they have lost much of their former importance; in the past, however, they were of great significance. Their bases might be patrilineal or matrilineal, or both these lines of descent might be recognized simultaneously in different contexts. Most Yoruba kin groups are patrilineal, the Asante groups are matrilineal, and along the Cross River there exist a number of variations of "double unilineal" descent, a system in which each individual belongs to both a patrilineage and a matrilineage that share areas of authority. Kin structures such as clans were usually of great significance in the administration of groups even in large and complex kingdoms although, as in the case of the Benin kingdom, there were exceptions. Even in matrilineal societies almost all important offices were held by men but, because women in such groups determined the group affiliation of their children and were of great formal significance in establishing a man's rights (as he claimed political office through his mother), women commonly attained a freedom of action and a degree of public significance that was difficult for them to acquire in patrilineal kin groups.

The rights of women in marriage varied considerably from group to group, but in many Guinea Coast societies, even patrilineal ones, it would generally be a mistake to regard women as having been particularly downtrodden in the past. Even when this superficially appears to have been the case, careful research reveals that, as very active economic partners of their farmer husbands, wives might exercise much influence over the allocation of crops and even of patrilineally inherited land. Among the double unilineal Mbembe of the Cross River, although the land a family used was normally given to the husband by his patrilineal kin, the wife's labour was clearly recognized as entitling her to part of the crops. The husband had an absolute obligation to retain enough of the crops to provide her and her children with food, and his wife was regarded as justified in divorcing him if he were irresponsible in his use of any money he had gained through their sale. The specific duties of husbands and wives in relation to different types of plots were detailed with great care, and a wife was entitled both to an area of her own for raising cash crops and to her husband's labour on it for

certain tasks. Among the Yoruba the husbands and their sons did almost all the farm work, and the women were responsible for marketing. Today sons move to the towns and husbands may employ labourers, but the old pattern remains and women continue to sell the produce. In an era when commercial marketing is much more important than formerly, many wives derive greater profit from the family's activities than do their husbands. Even in the modern urban environment the tradition of financial independence between husband and wife continues, but there the balance of advantage between husband and wife is less clear-cut.

Age-sets

In most tribal societies age is an important basis for group formation, and in some there exist "age-sets"—compulsory groupings of individuals of roughly similar age who advance through life together. No society in western Africa attached the same political significance to age as did some East African societies, where age stratified the male population into groups with markedly different levels of rights and authority. In the Guinea Coast, age groups tended to be more important in societies with weakly developed formal political structures; Igbo groups, for example, attached considerable importance to status given by age-set membership. In the Benin kingdom groups structured by age were, similarly to most other aspects of life, subjected to manipulation from the centre. If a man remained in his village his age-status was determined solely by his birth; if he went to Benin City and served in one of the so-called palace societies he might return home after a period and be entitled to promotion to an age-set beyond his years. This was a distinct inducement to go to work at the capital, and, since ultimately elders held sway in the village, the effect was probably that the most vigorous and, therefore, leading elders were relatively young men who had been influenced by the Benin City "establishment." Quite apart from the political signficance of age-sets, however, they might provide very significant personal ties for individuals who existed independently of any kin group organization to which they were attached. Among the Mbembe of the Cross River, for example, women as well as men derived great personal support from their age-mates. While kin-group membership established formal rights and obligations both between members and as collectivities in relation to the rest of society, age-sets provided friends who supported the individual through thick and thin. Age-mates were the witnesses called in when husbands and wives quarreled; age-mates would continue to find money to help the chronically sick even after kinsfolk gave up trying to aid; and in cases in which it was believed that a sick person was the victim of a sorcerer, always believed to be a kinsman, it was the age-mates who as a body would demand that he desist. The links between age-mates there, and elsewhere where they united individuals in different kin groups, did, however, have a political aspect, for they were of great value when kin-group elders met to deal with intergroup arguments; often the elders were age-mates and had the closest personal ties with one another.

Secret societies

One of the most characteristic of Guinea Coast institutions, especially in areas in which central government was weakly developed, was the so-called secret society. Such societies had a significance similar to that of age-sets because they cut across kin-group lines and united people in different settlements or of different political groups. Moreover, the fact that membership was often graded and the higher grades were open to those who could pay the fees meant that in societies where new wealth from trade became important it was often through these societies that wealthy men (few were open to women) achieved political influence to which they might not otherwise have had access. In two major areas—in Sierra Leone and Liberia, and in the area east of the Niger delta—these associations achieved such power that they were crucial to the pre-colonial systems of law and order. Among the tribes in the former area, there was a women's society, Sande, but Poro, for men, was the major organization responsible for punishing such serious offenses as incest and homicide. There were local Poro councils composed of members of the highest grade, and a chief's authority often rested on his Poro rank. Poro spread among the Gola, Kpelle, and Mano of Liberia and the Mende of Sierra Leone. Sometimes its members forged links between autonomous chiefdoms, and in 1898 the Mende Poro even organized a general uprising to try to oppose British expansion into Sierra Leone.

The most interesting example of the politically powerful secret society, however, was probably that of Ekpe in 19th-century Calabar, the Efik capital at the mouth of the Cross River. There the Ekpe society was the main instrument of the governing oligarchy of wealthy traders. There was no strong central government to ensure that traders honoured their commitments either to one another or to European traders, but the threat of Ekpe action usually ensured compliance. The power of Ekpe is credited with having made Calabar society one of the most stratified on the Guinea Coast. Its membership was reserved almost exclusively for freemen, and its power was used to subordinate the large slave population. In the neighbouring delta area many able slaves who escaped exportation were eventually incorporated into local groups and became almost indistinguishable from the local population. In Calabar, perhaps because there was more agricultural land available, unexported slaves were kept as serfs. A few became prosperous and had slaves of their own (no one who had any pretensions did his own manual work). Ekpe members, however, banded together to maintain the free–slave distinction. Not surprisingly perhaps, Calabar was a place of bitter friction and saw slave uprisings in the middle of the century. In general, secret societies were institutions for translating slight advantages of wealth into political influence, and wherever they occurred they indicated the existence of a measure of social stratification greater than that which commonly distinguished the successful elder from his junior kinsmen. The idea that small-scale tribal societies were essentially egalitarian has become much less tenable with recent research. It is recognized that, even in apparently unstratified societies, successful men usually depended for their position on their ability to control the labour of junior male kin and wives. Often they used their positions to increase their control over both at once, for by marrying polygynously while keeping young men wifeless they ensured an excellent labour supply for themselves. This was crucial because, until the mid-20th century, the densities of population were so low that land was not really scarce; what was scarce was the labour with which to work it. It is significant, for example, that even among the apparently unstratified societies of the Cross River, the Yako and the Mbembe, one of the most important of their secret societies was that used to punish young men who failed to work adequately for their fathers. Land was freely available, and for various reasons it was difficult to prevent young men from marrying and setting up their own households; the main sanction the elders could use was to call on members of this society to beat any recalcitrant junior who showed his face in the main settlement.

Stratification was most obvious among wealthy and centralized states. However, because wealthy men were always polygynous and usually had many offspring, the wealth of one generation was commonly dispersed in the next, so that class formation was limited—the hereditary basis of high status was lacking.

Freed slaves from North America

In the 19th century new forms of stratification emerged in Sierra Leone and Liberia when freed slaves educated in North America were settled in these areas to become the "Creoles," shopkeepers and white-collar workers—an elite vis-à-vis the natives. Some Sierra Leonians moved to other British West African possessions on the coast, where they joined with tiny indigenous elites drawn from wealthy, educated coastal families to form with them a new bureaucratic class. In the 20th century access to good education has done much to confer hereditary advantage on the children of the elite.

Belief systems. There are, at least in outline, similarities between the various belief systems on the Guinea Coast. Most systems contain these features: belief in a withdrawn high god; belief in lesser gods that are useful because easily manipulated; concern with the dead, usually but not necessarily ancestors, who are thought to exercise influence over the groups to which they belonged

in life; and belief in witches and sorcerers, whose existence explains undeserved misfortune. Finally, there is common acknowledgment of the power of diviners who can determine the cause of a particular misfortune. Beliefs as to what constitutes the basis of diviners' powers vary widely, but there is such a pragmatic attitude that what matters is the apparent success of the divination, not its conceptual foundation. Diviners in the past traveled widely between societies. They often advertized the power of distant cults, and sometimes priests were brought long distances to establish new local shrines. In this way famous cults spread widely, and this fact may help to account for the existence of broadly similar beliefs between societies that have had apparently rather little contact with one another. In general, however, these religious beliefs are broadly compatible with the type of society in which they are found, so that there tended to be complex pantheons of gods in hierarchical and stratified kingdoms such as Dahomey, but small-scale, stateless societies lacked that kind of complexity among their deities.

These beliefs accounted not only for misfortune but also for individual success. There was a widespread belief that individuals, in an existence before birth, had to choose whether their earthly lives should be fortunate or unlucky, and consistent lack of success might be ascribed to wrong choices—so-called prenatal fate. In some societies particularly fortunate individuals were presumed to have established a relationship with a luck-bestowing supernatural being—for example, the water spirit of the Niger delta. In other societies undue prosperity was suspect and might be ascribed to membership in a sorcerers' society, which was believed to give wealth to its initiates in return for a sinister fee: the life of a relative. In such societies the implication was that the individual could get ahead only at the expense of his kin.

Throughout the Guinea Coast the influence of Christian missionaries has been prolonged and considerable, and in many areas there are flourishing and orthodox locally run churches. This does not mean that all traditional religious explanations of the events in a human life have disappeared. Sometimes they continue to constitute a kind of fall-back position, so that local diviners may be consulted in cases of illness where modern medicine seems powerless. Sometimes they are blended into complex new sects with Christian and Muslim elements, as in the Yoruba-centred Church of the Cherubim and Seraphim. This has spread into distant areas because it combines many of the characteristics of the churches—comparable buildings and forms of worship—with a traditional concern for healing and the explanation of untoward events. In general, however, Guinea Coast cultures have come under significant Christian influence, and this distinguishes them from the northern hinterland that is predominantly Muslim. Insofar as Islām has been spread by the sword, it is not a matter of chance that the old forest zone, inimical to cavalry, formed a barrier beyond which this faith has been slow to penetrate. European colonialism subsequently protected the southern peoples from further forcible conversion by jihad. Thus the natural vegetation zones described at the beginning of this section have continued, for complex reasons, to be distinct cultural zones, particularly because education has depended very much on the activities of Christian missions, so that until recently the new educated elite of much of western Africa has been drawn predominantly from the southern Guinea Coast area. This elite has staffed the state bureaucracies and the modern commercial enterprises, giving southerners many advantages but also increasing the political problem of regional jealousy.

Modern relationships. When the present-day states of the Guinea Coast are considered in the context of the modern world, it is important to be aware of the ethnic diversity within nations whose boundaries have been dictated by colonial powers that paid scant regard to indigenous ethnic divisions. Despite the many changes that have taken place, old loyalties to these divisions sometimes remain strong. Ethnic patriotism may lead to a reassertion by the educated elite of the value of their particular local culture.

Ethnic divisions may, however, be less ultimately divisive than the new patterns of economic stratification. There are now vast differences in income and power between the rural population and the urban poor on the one hand and the politically and economically successful minority on the other. This is particularly noticeable in Nigeria, where a devastating war was followed by massive inflation, partly related to petroleum development, and where new laws have allowed powerful men to expropriate the land of peasant farmers.

Modernization has brought political and economic problems, but the Guinea Coast continues to be an area rich in cultural tradition. The region has produced not only scholars of high quality in many academic fields but also many imaginative writers, the best of whom are among the outstanding contemporary novelists and playwrights writing in French and English. (R.L.Ha.)

History

A reasonable body of sources for the writing of western African history begins to be available from about AD 1000. Three centuries earlier, the Arabs had completed their conquest of Africa north of the Sahara and so came into possession of the northern termini of trade routes reaching across the desert to western Africa. The lively school of geographers and historians that flourished in the Muslim world from about the 9th to the 14th century thus secured access to growing amounts of information about what they called the *bilād as-sūdān,* the territory of the black peoples south of the Sahara.

MUSLIMS IN WESTERN AFRICA

This information has its limitations. The Muslim writers, contemptuous of non-Islāmic societies, passed on little of what they must have known about the organization of pagan black societies and tended to concentrate on and condemn what struck them as their more monstrous aberrations. Conversely, they doubtless exaggerated the importance of the Islāmization that entered western Africa with the Muslim traders crossing the Sahara. The earliest firsthand account of western Africa is probably that of the world traveler Ibn Baṭṭūṭah, who visited the western Sudan in 1352–53. Finally, the North African merchants did not penetrate into western Africa beyond the urban centres of trade and government that existed or came to develop on the northern fringes of the cultivable savannas fronting the Sahara. Their *bilād as-sūdān* was in fact only the northern marches.

Nevertheless, the picture of western Africa given in the early Muslim writings is of major interest. It is apparent that, right from the beginnings of Arab contact, the organization of the more northerly western African peoples was not solely tribal. They had considerable towns and cities that were supported by a developed agriculture. They had organized networks of markets and trade and a developed system of monarchical government. Kings, whose claim to power was based on descent from the mythical divine founding ancestors of their ethnic groups, taxed trade and levied tribute on the agricultural villages through their possession of bodies of retainers who provided them both with military force and with a hierarchy of officials.

It seems likely that there was an increase in the volume of trans-Saharan trade following the organization of North Africa under Muslim dynasties and that this growth of international trade with western Africa stimulated the growth there of internal trade, urbanization, and monarchical government. Certainly the control of trade, towns, and government in western Africa became increasingly Islāmic in form. But it is quite clear that the foundations for the economic and political development of the western Sudan were in existence before the time of contact with Muslim traders or authors. Early Muslim interest was concentrated on two major western African kingdoms: Kanem, in the east, north of Lake Chad; and Ghana, in the extreme west, on the borders of modern Mauritania and Mali. The Muslim sources, which are broadly confirmed by local tradition, indicate that the kingdom of Kanem was being formed during the 9th and 10th centuries through an interaction between Saharan nomads

The kingdoms of Kanem and Ghana

and agricultural village communities. But ancient Ghana (not to be confused with its modern namesake, considerably farther to the south and east) had already reached levels of organization that presuppose several centuries of continuing development.

The earliest extant Arabic reference to a kingdom of Ghana dates from the early 9th century. In the middle of the 11th century, the Córdoban geographer Abū U'bayd al-Bakrī described its capital, court, and trade in some detail. The capital was made up of two towns, a stone-built town inhabited by the Muslim traders and a mud-built one of the local Mande in which the king had his walled palace. Their centres were six miles apart, and the whole of the intervening country was more or less built up. The considerable population was supported by the produce of surrounding farms, which were watered from wells. The court displayed many signs of wealth and power, and the king had under him a considerable number of satellite rulers. A principal part of his revenue was derived from regular taxes on trade. The mainstay of this trade was the exchange of gold, which Ghana's own merchants brought from lands to the south, for salt, which the northern traders brought in from salt deposits in the Sahara.

Al-Bakrī's description is broadly confirmed by archaeology. The region in which ancient Ghana was situated contains the ruins of a considerable number of stone-built towns that must have been supported by considerable agricultural and commercial activity; those at Koumbi Saleh are generally identified with the capital described by al-Bakrī.

The relatively extensive Muslim interest in Ghana was undoubtedly due to its importance as a source of gold. Kanem seems to have been less important commercially; the main interest of the Muslim authors seems to have been in the quasi-divine status of its kings, which offended their Muslim principles. Other western African kingdoms undoubtedly existed at this time, but the Muslim sources record little of them beyond their names and approximate locations. Thus between Ghana and Kanem was Kawkaw, perhaps the nucleus of the later Songhai kingdom of Gao. Malel, to the south of Ghana, may similarly have been a prototype of the later Mande kingdom of Mali, which ultimately was to eclipse and absorb Ghana itself.

Other kingdoms

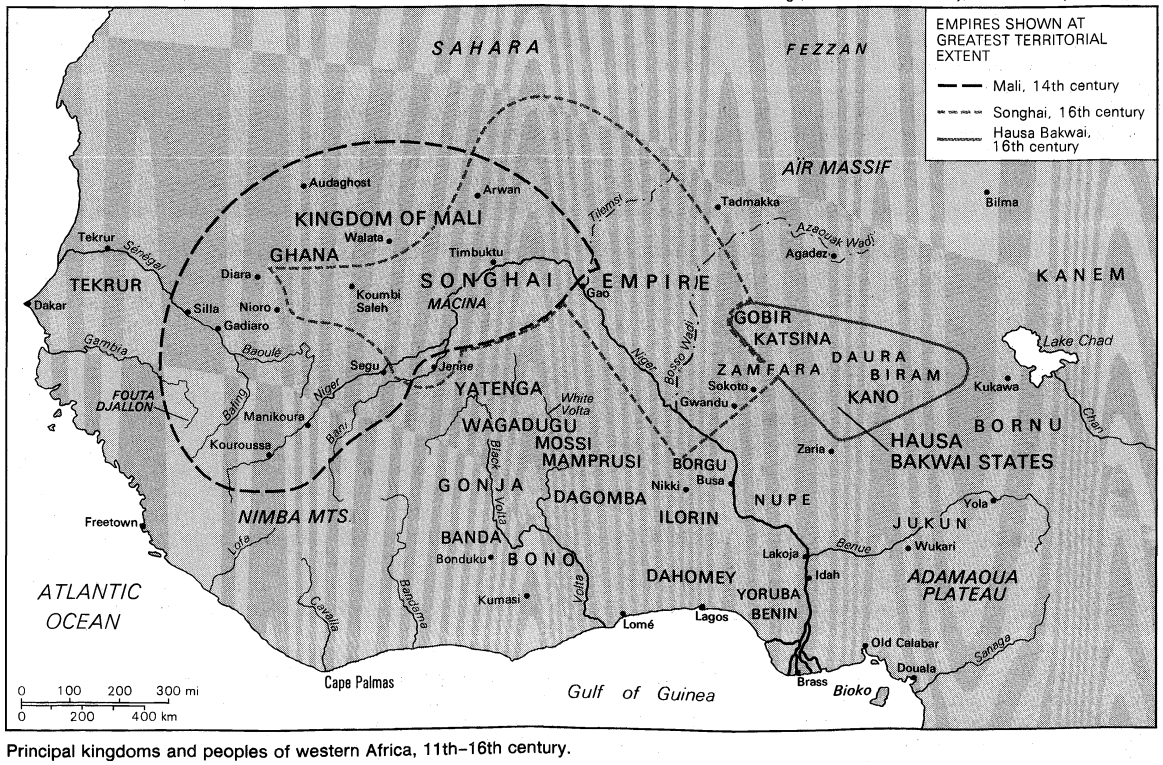
There are perhaps three possible—and not mutually exclusive—explanations for the origins and development of the kingdoms that Arab trade and scholarship had revealed by about AD 1000. The first is that they were the result of the invasion of agricultural territory by pastoralists from the Sahara who belonged to the Hamitic-speaking Libyan Berber tribes who were the dominant stock of North Africa before its conquest by the Arabs.

This is the explanation often given in western Sudanese traditions and chronicles. From about the 15th century onward, many of these were preserved by local authors who wrote in Arabic and were Muslims, and who thus had some incentive to link the history of their peoples with that of North Africa and with the adjacent Middle East. It was also the explanation favoured by European historians of the later 19th and earlier 20th centuries when Europeans were themselves conquering and colonizing black Africa. There thus evolved the so-called "Hamitic hypothesis," by which it was generally supposed that any progress and development among agricultural blacks was the result of conquest or infiltration by pastoralists from northern or northeastern Africa. Specifically, it was supposed that many of the ideas and institutions of tribal monarchy had spread through Africa by diffusion from the ancient civilization of Egypt and the Nile valley.

There can be no doubt that over the centuries pastoralists from the Sahara have indeed advanced and conquered southward. But not all of these were Libyan Berbers or Hamites; some, such as the dynasts of Kanem, were black African in language and culture. Nor is it easy to understand how mobile desert pastoral tribes could be effective transmitters of ideas and institutions from the settled civilization of the Nile valley to other agricultural lands in western Africa. It would seem more probable that conquering pastoralists who did succeed in establishing new kingdoms and dynasties in western Africa should do so by taking over existing monarchies, perhaps city-states or even "village-states," and amalgamating these into larger units. Some early western African traditions can certainly be interpreted in this sense.

This leads to the second explanation for the origins and development of monarchical statehood among the western African tribes. There is archaeological evidence for the evolution of a cattle-herding and agricultural economy among a mixed population of Libyan Berber and black agricultural peoples in the Sahara by at least 4000 BC—*i.e.*, more or less contemporary with similar developments in the Nile valley. The desiccation of the Sahara and the evolution of its present desert between about 8000 and 2000

From J. Fage, *An Atlas of African History*; Edward Arnold (Publishers) Ltd.



Principal kingdoms and peoples of western Africa, 11th–16th century.

BC must have occasioned an outflowing of population in which the blacks concentrated in the savannas to the south of it. There, in favourable riverine or lacustrine environments, it seems reasonable to suppose that the same desire to avoid conflicts over land and water rights and to control and exploit agricultural surpluses, which had led in the exceptionally fertile but extremely constricted environment of the Nile valley to the dramatic kingship and civilization of the pharaohs, should have occasioned the evolution of similar if less spectacular monarchies.

Formation of black African monarchies

It should be noted, however, that the major western African monarchies known to the Arabs by about AD 1000 were situated not in the well-watered lands along the Sénégal and Niger valleys nor around Lake Chad but north of these, in the less favoured agricultural territory between them and the southern edges of the Sahara. This suggests that a third factor in the evolution of these northern monarchies was the influence of long-distance trade. The western African kingdoms had their own resources of iron, which in some cases were being worked by about 500 BC, but they imported other metals, notably copper, together with horses, luxury manufactures, and—above all—salt, a vital commodity that was scarce in all of western Africa except the coastlands. In exchange they could offer gold, ivory, certain agricultural commodities, and slaves.

The exchange across the Sahara of such commodities probably goes back to times before the establishment of the modern desert. The emergence of the desert did not lead to the cessation of the trade but meant that its surviving pastoralists were encouraged to organize regular trans-Saharan expeditions for trade and plunder. It is known from Herodotus and other classical authors and from surviving rock engravings in the desert that horse-drawn chariots were in use in the Sahara by about 500 BC. Chariots would have been used for short raids rather than for trans-Saharan trade. But the fact that the engravings are deployed along two principal lines, from the Fezzan and southern Morocco toward the upper Niger and Sénégal rivers, suggests a North African interest in the alluvial gold of these rivers. This could well have occasioned the 6th-century Carthaginian expedition led by Hanno to explore the possibility of direct sea trade with western Africa along the Atlantic coast. Despite this expedition, the Carthaginians do not seem to have been capable of opening up a regular sea trade with western Africa. The links with western Africa remained firmly in the hands of the Saharan tribes, although, at about the beginning of the Christian era, camels and other pack animals came into use to supplant the horse-drawn vehicles.

Trans-Saharan trade

The profits to be obtained by distributing Saharan and Mediterranean produce in western Africa, and by controlling the collection and export of the western African commodities that were exchanged for them, must have been a powerful factor in encouraging the kings of communities on the southern fringes of the Sahara to extend their rule by conquest over adjacent similar communities. Control over more extensive territories meant that by tribute and taxation they could acquire greater stocks of goods for exchange with North Africa and the Sahara and more clients and slaves to extend their power at the expense of their neighbours. Some of their increased human power could be mounted on horses, obtained from the Saharan trade, to increase the mobility and power of their armed forces over the open savannas. It is tolerably certain that the power of the kings of ancient Ghana, controlling the export of gold from the Sénégal and Niger valleys, was built up in this way.

THE STATES OF THE SUDAN

The early kingdoms and empires of the western Sudan. In the 10th century the kings of Ghana extended their sway over the Ṣanhājah, the congeries of Berber nomadic tribes living around Audaghost, just north of their kingdom, who supplied them with salt and North African goods.

This move must have upset the economic balance between agricultural Ghana and the pastoral Ṣanhājah, and ultimately it provoked a reaction. Like the North African Berbers, the Ṣanhājah tribes were already to some extent Islāmized, and they shortly found in a militant, puritanical version of Islām the means to eliminate their tribal differences and to unite in the movement known to history as the Almoravids. In the middle of the 11th century they began to expand into the productive lands on either side of the western Sahara, and it would seem that later in the century Ghana became dominated by them.

One important result of this domination, following as it did upon some centuries of trading contact by Muslims, was that the ruling and merchant classes of the western Sudan became converted to Islām—though in the case of the rulers the conversion was for many centuries not wholehearted. The justification for a king's claim to enforce his rule over his subjects, who remained pagan, was his descent from the original ancestor who had first settled the land and, by accommodation with its deities and spirits, had developed and controlled it for agriculture. If he were not to be rejected and replaced as king by a rival member of its royal family, he had to continue to observe the ancestral and land cult rites in which he was the principal figure.

However, the depredations of the Almoravids' herds and their internecine quarrels must have undermined the prosperity of agriculture in a marginal environment and would have accelerated the decay of Ghana. More southerly Mande groups, many of which had formed satellite kingdoms of the Ghana empire, began to act independently and to compete among themselves for primacy. Eventually in about 1235, in the time of a king called Sundiata, the Keita kings of Mali, in the well-watered and gold-bearing lands of the uppermost Niger valley, gained ascendancy and incorporated what was left of ancient Ghana into their own considerably more extensive empire.

The rise of Mali

The Keita clan seem originally to have been traders from lower down the Niger, and the strategy of their empire was to extend their power down river to the Niger Bend and to its trading cities of Timbuktu and Gao, which lay at the foot of the shortest trans-Saharan routes. The initial success of the Almoravids and their subsequent rapid decline had upset the stability of the more westerly caravan roads leading to Ghana, while by the 13th century Ifrīqīyah (Tunisia and eastern Algeria) and Egypt provided more stable bases for trans-Saharan trade than did Morocco. The Niger River provided a natural means of communication from Mali and its goldfields to Timbuktu and to Gao and also provided Mali's merchants with the possibility of opening up trade elsewhere in black western Africa. By the 14th century, Mande merchants, the Dyula (Dioula), were trading as far east as the city-states of the Hausa, between Lake Chad and the Niger. By about the same time they had also begun to develop a new trade route southeastward from Jenne (modern Djenné, Mali), on a southerly tributary of the Niger, toward goldfields that were being opened up along the Black Volta and further south still, in what is now modern Ghana.

These Mande merchants were Muslims, and their activities led to a considerable expansion of Islām among the trading classes of western Africa and, with the qualification mentioned earlier, also among its kings. Thus the first conversions of Hausa monarchs seem to date from the 14th or 15th centuries. The Mali kings themselves valued Islām for the commercial and diplomatic advantages it gave them, and some of them, of whom the best known is Mansa Mūsā (1307–32), made notable pilgrimages to Mecca via Egypt. As may be seen from Ibn Baṭṭuṭah's account of his travels in 1352–53, the essentially pagan society of the western Sudan became open to a considerable degree of Islāmic influence, and literacy and even scholarship became firmly established in its major cities.

The success of the Mali empire depended, however, on its rulers maintaining firm control of the Niger waterway. This in its turn depended on their maintaining control over a non-Mande people, the Songhai, who monopolized the fishing and canoe transport of the middle Niger. The Songhai had an independent monarchical tradition of their own, and Mande control of their capital, Gao, proved somewhat fitful. During the 15th century it was lost altogether, and eventually a Songhai king arose, Sonni 'Alī (1464–92), who, appealing to traditional Songhai pagan-

ism against the Islāmic universalism of the Mande system, destroyed the Mali empire by ceaseless military campaigning and erected in its place a new empire ruled from Gao. But if this empire were to be profitable and strong, the Songhai needed the Mande as much as the Mande had needed the Songhai. After Sonni 'Alī's death, power passed to one of his former generals, al-Hājj Muḥammad Askia (1493–1528), who was both a Mande and a Muslim, and thereafter there was a continual struggle for power between the two groups.

The extent of the Songhai empire

The Songhai empire was never strong in the west, where a number of Mande kingdoms remained in the tradition of Ghana and Mali, but was more effective to the east. There the kingdom of Kanem, whose kings had become Islāmized in the 11th century, had declined during the 14th and 15th centuries following quarrels among its aristocracy when it was subject to pressure from new nomad invasions. Eventually, however, the Kanem kings reestablished their state in the former province of Bornu in the southwest, close by the Hausa kingdoms. But in the 16th century, Songhai was the most important external influence over the latter, which began to grow in power and importance. South of the Niger Bend the kingdoms of the Mossi-Dagomba peoples were emerging, founded by bands of cavaliers who may have been in some way connected with the ruling families to the northeast.

Songhai was strong enough to extend its sway northward across the Sahara to as far as the salt mines of Taghaza, close to the Moroccan borders. This upset the balance of trans-Saharan trade, as Ghana's attempt to control the Ṣanhājah had done, and in 1591 finally provoked effective retaliation from the Sa'dī dynasty of Morocco. An expeditionary force of some 4,000 soldiers was sent across the Sahara and took the important cities of Gao, Timbuktu, and Jenne. The Moroccans had firearms, but their success against the much larger numbers of the Songhai army was also facilitated by the internal divisions of the Songhai state. For a time the profits of this enterprise were considerable, but the Moroccans were not strong enough to control the network of trade routes within western Africa that brought gold and other produce to the Niger cities. Ultimately the main gainers from their conquest were the Saharan tribes, essentially Berber in origin (such as the Tuareg) but now increasingly Muslim and even Arabized, who finally levied tribute on the descendants of the Moroccan soldiers who formed the military caste (Arma) of Gao, Timbuktu, and Jenne.

Firearms also came to the central Sudan about the same time through the trading relations that existed between Bornu and the Ottoman Turks in North Africa. Together with Muslim cavalry, they enabled Idris Alawma of Bornu (end of 16th century) to impose a Muslim bureaucracy on his pagan subjects and to reconquer Kanem. This revival of the Kanem-Bornu dynasty, however, was relatively short-lived. By the 18th century it was the much smaller Hausa kingdoms, especially Kano and Katsina—which had learned much from Mande commercial and industrial experience and had developed a trading network to the south to rival that of the Mande themselves—that took the leading role in western Africa's external trade with North Africa across the Sahara.

The wider influence of the Sudanic kingdoms. The development of such major Sudanic kingdoms and empires as Ghana, Mali, Songhai, the Hausa states, and Kanem-Bornu along the southern fringes of the Sahara had a number of important consequences for the history of western Africa as a whole. For example, it provided the background for the expansion of the Fulani, the only pastoral western African people (also variously known as Fulbe, Fula, Fellata, and Peul).

The fact that, uniquely in western Africa, the Fulani are pastoralists and that physically they have some non-Negroid traits has led to suggestions that they were originally a Saharan people. The Fulani language, however, is classified as part of the Niger-Congo family of languages spoken by black Africans, and the earliest historical documentation reports that the Fulani were living in the westernmost Sudan close to ancient Ghana. The development of this organized kingdom thrust pastoral peoples outward, and the ancestors of the modern Fulani seem to have chosen to settle to the southwest, toward the middle Sénégal valley. But there another settled, and (from the 11th century) an Islāmized, black kingdom evolved, that of Tekrur. Some Fulani participated in this kingdom and became Tukulor (Toucouleur)—the Tukulor and Fulani languages being practically identical. Some, however, chose not to accept the settled way of life and, to preserve their traditional pastoral and religious customs, migrated eastward over the savanna grasslands. Grazing land was available between the agricultural villages, and the growing towns provided the Fulani with markets where they could exchange their pastoral produce for agricultural and manufactured goods. Eventually some Fulani settled in these towns, no doubt initially as trading agents for their fellows in the countryside, where the bulk of the Fulani continued to live under their own leaders, aloof from the social and political life of the cultivators, though increasingly paying rent for their grazing and rendering military services to the settled authorities.

The Fulani diaspora

In this way, by the 15th century, large numbers of Fulani had settled in the Fouta Djallon and in and around Macina, the inland delta country of the Niger upstream of Timbuktu, and they were beginning to appear as far east as Hausaland, where today many millions of their descendants live. By the 16th century the Fulani were appearing in Bornu, and by the 18th century large numbers of them were settling in the grassy uplands of the Cameroons. Although the bulk of the Fulani remained animists, gradually significant numbers of them became Muslims and indeed provided some of the leaders of Islām in western Africa.

Southerly movements

The growth of organized statehood in the western African savannas also had important consequences for more southerly lands and peoples. Unsuccessful contestants for power in the major savanna states sometimes moved off toward the south. Traders from these states, especially from Mali and, later, from the Hausa kingdoms, also settled in the south as their trading networks developed, and they often had important political, as well as economic, influences on the tribal groups with whom they came to live. The consequences of these two kinds of movement, which were sometimes interlinked, can best be considered in two sections—firstly, those deriving from developments originating in the Mande sphere in the western savannas, and secondly, those deriving from the Hausa-Bornu region in the east.

In the west one notable emigration was that of the Susu, a Mande group that had lost out to the Keita in the 13th-century struggle for the inheritance of ancient Ghana. This emigration created a wedge of Mande-speaking people close to the Atlantic in the modern Republic of Guinea and in northern Sierra Leone among peoples who had not advanced politically beyond the village level. With the subsequent growth of Mali's power, other smaller groups—sometimes traders, sometimes conquerors, often both—also infiltrated these West Atlantic coastlands. They began to organize their people into petty kingdoms that tended to owe a nominal allegiance to Mali. Major incentives for these migratory movements seem to have been the desire to gain access to coastal supplies of salt and, from the 15th century onward, to the foreign merchandise brought by European sea traders.

In the 16th century the West Atlantic coastlands were invaded by yet another Mande group, the Mane, who advanced westward parallel to the coast from Liberia onward. These were military bands that systematically attacked and overcame the villages of each tribal group they came across. Some of them would stay behind to organize these conquests into small kingdoms, while others, reinforced by auxiliaries recruited from among their victims, would proceed farther west to repeat the pattern. Their advance was halted only when they came up against the Susu. South and east of the Susu, however, the West Atlantic social and ethnic patterns were considerably altered by the actions of the Mane. New Mande-speaking groups emerged, such as the Mende and Loko, while some West Atlantic peoples who retained their original language, such as the Temne, accepted a new aristocracy of Mane provenance.

The Mane advance into the West Atlantic coastlands from the east may have been connected with the growth of Mande influence in the Ivory Coast and in modern Ghana. This was commercial in origin; Dyula merchants developed trade routes in search of gold, slaves, and kola nuts, in exchange for which they offered salt, cloth, and other Sudanic or North African goods. It is known that by 1500 the Dyula were trading as far south as the coast of modern Ghana, and their first contact with the Akan peoples who populate almost all the southern half of this territory was probably one or two centuries earlier than this.

This development of trade by the Dyula in modern Ghana and in the adjacent Ivory Coast had important political consequences, and sometimes military implications as well. Ambitious Akan chiefs began to develop and expand their political power to secure the maximum profit from the exploitation of the resources of as much territory and as many people as possible. On the northern fringes of the forest, astride the routes along which gold and kola nuts were brought for exchange with the Dyula, important new kingdoms emerged such as Bono and Banda, both of which were probably in existence by about 1400. As the economic value of gold and kola became appreciated, the forest to the south of these states—which had hitherto been little inhabited because it was less favourable for agriculture than were the savannas—became more thickly populated, and the same principles of political and military mobilization began to be applied there. Village communities became tributaries of ruling groups, with some of their members becoming the clients and slaves needed for the support of the royal households, armies, and trading enterprises. By 1500 most of the Akan territory seems to have been organized in this way and, as trade increased, so the political units—initially very small—tended to increase in size.

Increasing power of Akan

Sometimes these political changes were not to the advantage of the Dyula, who employed Mande warriors to guard their caravans and, if necessary, could call in larger contingents from the Sudanic kingdoms. Tensions between the Dyula and the increasingly powerful animist monarchy of Banda erupted in the 17th century into a civil war that destroyed the kingdom and led its Dyula merchants to establish a new trading base of their own farther to the west at Bonduku (Bondoukou). In the following centuries there were at least two major examples of the Dyula taking political authority for themselves at strategic points on the trade routes running through the eastern Ivory Coast.

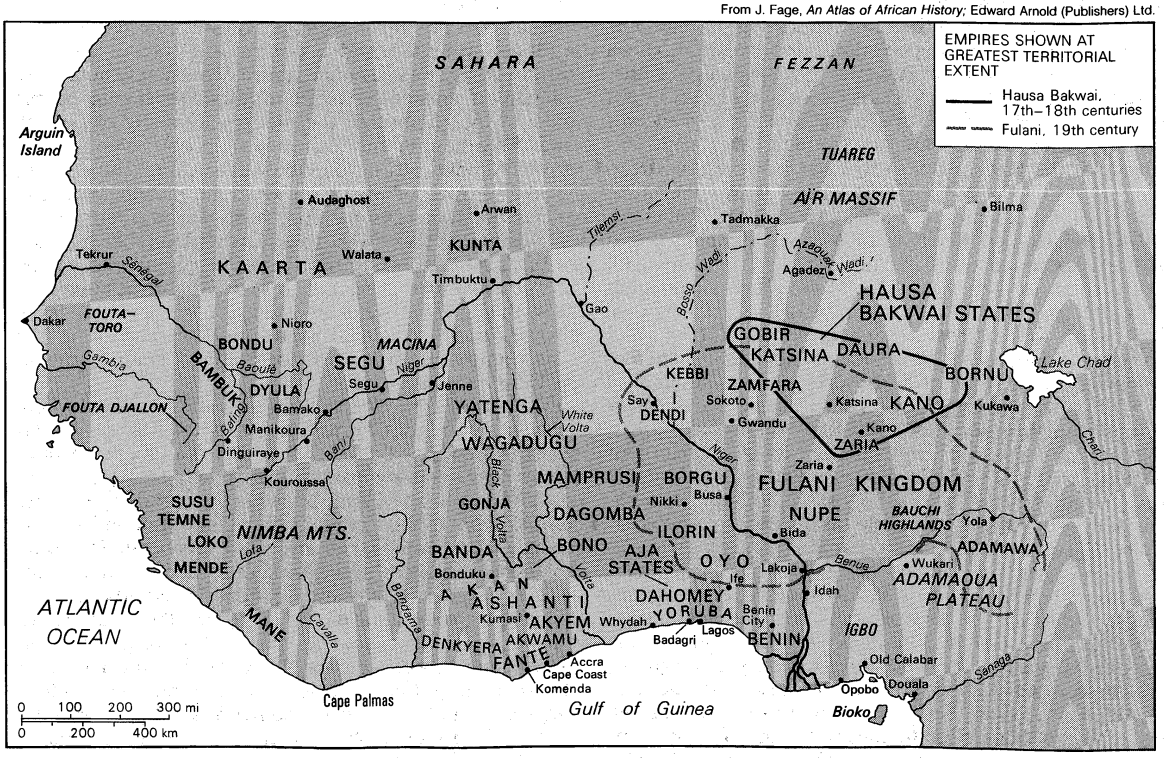
But the most interesting Mande political initiative along the trade routes south of Jenne was the creation in the early 17th century, just north of the Akan lands, of the new Dyula state of Gonja. This seems to have been inspired by a general worsening of the competitive position of the Mande traders, and it was occasioned by three factors: (1) the near monopoly in the control of the export of forest produce achieved by the Akan kingdom of Bono; (2) the rise to power farther north of the Dagomba kingdom, which controlled local salt pans; and (3) the arrival in the region in about 1500 of rival long-distance traders from Hausaland. The Dyula seem to have tried to combat these developments by erecting a major kingdom of their own in Gonja—the territory that the northern traders had to cross to reach the Akan forestlands. But much of Gonja was barren, and its kings lacked the resource base to withstand the growth of Akan power.

The kingdom of Gonja

Rather less is known about the nature of Sudanic influences in the more easterly zone south of Hausaland and of Bornu. However, it has already been suggested that Dagomba (and a number of similar kingdoms in the Volta basin, including Mamprusi) and the Mossi kingdoms—such as Wagadugu (Ouagadougou) and Yatenga (or Wahiguya), north of Dagomba and closer to the Niger Bend—were founded by conquerors coming from the east. The structures of these kingdoms, which were extant into the beginning of the colonial era, seem to have been erected about the 15th century by relatively small bands of immigrants who eventually merged with the autochthonous Gur-speaking inhabitants of the Volta basin. Their success in conquering and organizing the Gur villages into kingdoms seems to have been due to their possession of cavalry, which subsequently remained a badge of royalty and of aristocracy.

Somewhat earlier, and farther to the east, astride the Niger and closer to Hausaland, similar kingdoms seem to have been developed through the same kind of process by invaders who may well have been ancestral to the Mossi–Dagomba state builders. Examples of these survived in Borgu into colonial times. An interesting corpus of legends—such as that of Kisra, a character derived from the Sāsānian conqueror of Egypt, Khosrow II, who is supposed to have migrated southwestward from the Nile valley founding various kingdoms—suggests that state-building

From J. Fage, *An Atlas of African History*; Edward Arnold (Publishers) Ltd.



Principal kingdoms and peoples of western Africa, 17th–19th century.

invaders also proceeded south of Borgu and Hausaland through Nupe, Jukun, Igala, Yoruba, and Benin territory (all in modern Nigeria) to as far as southern Dahomey and to the southeasternmost tip of modern Ghana. Much of this territory, however, was subjugated in the 19th century by Muslim Fulani and Hausa, and the legends as they survive often have an Islāmic colouring, which makes it difficult to assess their historicity. It could well be that the traditions of the autochthonous pagans had become coloured by the folklore of incoming Muslim traders. But if the legends are really evidence of conquest, there can be no doubt that the conquered peoples had already achieved a high degree of cultural, economic, and—probably—political sophistication.

Skill in sculpting

The archaeological evidence of the Nok culture shows that the inhabitants of central Nigeria engaged in agriculture and were using iron and other metals by about 500 BC. It is moreover generally accepted that the terra-cotta sculptures associated with Nok are precursors of the later court sculpture of southwestern Nigeria, in bronze or brass and stone as well as terra-cotta. The bronze or brass sculpture of Yoruba and of the kingdom of Benin is especially famous for its naturalism and for the high degree of metallurgical skill used to cast the figures by the lost-wax process.

Tradition asserts that the Yoruba town of Ife (Ile-Ife) was the centre from which this art, and the type of monarchy that supported it, spread to other parts of the region. Modern archaeologists have found nothing to contradict this and have moreover provided evidence for the existence in Yorubaland of a high degree of urbanization by at least the 11th century. Thus in southwestern Nigeria urban civilization seems at least contemporary with the same development in Hausaland and in Kanem–Bornu, the point of departure for Kisra-type conquerors. The probability, therefore, is that if the Kisra and similar legends are evidence of conquering cavaliers, they were—like the Saharan invaders in the Sudan—exploiting rather than creating urban civilizations, though no doubt they may have welded smaller units into larger kingdoms.

It should be pointed out that the territory southwest of Hausaland toward Dahomey is often open country suited to the deployment of cavalry. The effectiveness of cavalry decreased southeastward toward the forest; in the Benin kingdom, horses were little more than residual symbols of monarchical and aristocratic power. Still farther east, in the forest on the other side of the Niger, where there are no traditions of invasion or of developed monarchical government, the archaeological finds at Igbo Ukwu revealed that ancestors of the modern Igbo (Ibo) had, as early as the 9th century, a sophisticated society with surpluses of wealth supporting considerable craft specialization, including a highly developed bronze art with a distinctive style of its own. Recent thinking suggests that the origins of the small, competitive city-states of the eastern Niger delta south of Igboland, with which Europeans were to develop flourishing commerce from the 17th century, may well be associated with earlier, purely indigenous trading activities of which very little is at present known.

On the available evidence, the establishment in eastern Guinea of the network of Hausa trade and trade routes in the form in which it has been known in modern times, essentially similar to the Mande system farther west, can hardly be dated earlier than about the 16th century. Yet the European traders arriving at Benin, in the Yoruba coastlands, and in the Niger delta, from the end of the 15th century, were able to secure regular supplies of goods, some of which (*e.g.*, cloth and beads) clearly originated in the hinterland and seem to have had an already extant market among the Akan peoples of southern Ghana. This presupposes appreciable indigenous commercial development in eastern Guinea before the emergence of the Hausa system.

THE BEGINNINGS OF EUROPEAN ACTIVITY

The advent of the Portuguese

The arrival of European sea traders at the Guinea coastlands in the 15th century clearly marks a new epoch in their history and in the history of all of western Africa. The pioneers were the Portuguese, southwestern Europeans with the necessary knowledge, experience, and national purpose to embark on the enterprise of developing oceanic trade routes with Africa and Asia. Their main goals were in Asia, but to reach Asia it was necessary to circumnavigate Africa, in the process of which they hoped, among other things, to make contact with Mali and to divert some of the trans-Saharan gold trade from Muslim North Africa to Christian Europe.

The colonization of the Cape Verde Islands, from the 1460s onward, provided bases for trade with the fringes of the Mali empire. The most momentous discovery in western Africa, however, came in 1471, when Portuguese captains first reached the coast of modern Ghana between the mouths of the Ankobra and Volta rivers. It was quickly appreciated that the Akan peoples of this coast had access to supplies of gold, which were plentiful by contemporary European standards, and that they were willing and organized to trade some of this gold for base metals, cloth, and other manufactures. The Portuguese called this coast Mina, "the mine," while in European languages generally it became known as the Gold Coast.

The wealth obtainable from trade with the Gold Coast was so important for the completion of the Portuguese design to establish regular commerce with Asia by circumnavigating Africa that the Portuguese crown quickly took steps to exclude foreign rivals from the western African trade and to bring it under its direct control. Portugal was not a naturally wealthy nation, however, and its overseas interests had become very widely extended by the beginning of the 16th century. The western African coastlands and their trade were only one element in a system that also embraced the Congo and Angola, Brazil, the East African littoral, and India and the East Indies. By and large it was the trade of the latter that was regarded as the major prize, and elsewhere activities tended to be restricted to those which might strengthen the prosperity of the overseas enterprise as a whole without unduly straining the limited resources, especially perhaps of labour, available for its control and exploitation.

The general strategy in western Africa—as elsewhere in the Portuguese trading empire—was to keep territorial and administrative commitments to the minimum necessary to develop and benefit Portuguese commercial activities that were already in existence. The main interest in western Africa was the gold trade of Mina, and it was there—and virtually there alone—that the Portuguese endeavoured to maintain a positive presence on the mainland. In 1482 they built the strong fort that they called São Jorge da Mina (the modern Elmina Castle) on the shores of the Gold Coast, on land leased from the local Akan, and in subsequent years this was supplemented by the construction of three additional forts, at Axim, Shama, and Accra. The purpose of these forts and their garrisons was to try to ensure that the local people sold their gold only to agents of the Portuguese crown. No other Europeans succeeded in establishing lasting footholds on the Gold Coast before the close of the 16th century, and the Portuguese purpose was largely achieved. The surviving records suggest that up to about 1550 the Portuguese were securing from the Gold Coast on average at least 12,400 ounces of gold each year, a sizable proportion of the production then available to Europe.

In exchange, the Gold Coast peoples needed to be supplied with commodities they desired, and this presented Portugal with a problem, as it was not a major manufacturing nation. The raw iron and copper, metal goods, cloth, and other items that were in demand on the Gold Coast had to be purchased elsewhere. Some of the cloth exported to the Gold Coast was in fact brought from Morocco (and may therefore have been in competition with a trade in cloth that had earlier reached the Akan from the north), and the requirements of their Gold Coast customers were a prime factor in leading the Portuguese to develop relations with the kingdom of Benin and the Niger delta, where further supplies of cloth, and also of beads and slaves that were in demand on the Gold Coast, could be obtained.

At first the Portuguese hoped to control the trade of Benin and surrounding areas by converting the kingdom,

or at least its court, to Christianity and turning it into a satellite protectorate of their empire. Although this kind of policy was initially successful elsewhere in Africa, notably in the Kongo (Bakongo) kingdom of northern Angola, the Benin monarchy was powerful enough to reject European pressures and infiltration. From about 1520 onward, the Portuguese were virtually excluded from Benin, and their trade with the Niger delta was conducted from São Tomé and from the other islands of the Gulf of Guinea that they had colonized. This trade was principally in slaves, from the Congo and Angola as well as from the delta, who were employed on plantations to grow tropical produce, sugar in particular, for the European market.

Rejection by Benin

Apart from an abortive attempt to intercept the western trans-Saharan trade from a fort that was erected on the island of Arguin off the coast of Mauritania, the other principal Portuguese activity in western Africa was the trade with the coastlands of Upper Guinea that was conducted by the settlers on the Cape Verde Islands (which, together with Madeira, were also developed as plantation colonies employing African slave labour). The empire of Mali was in decline, but the Portuguese were not strong enough to control trade so far into the interior. What ultimately developed, on the creeks and islands of the coast from the Gambia to Sierra Leone, was a number of informal settlements where traders from the Cape Verde Islands did some trade with Mande merchants and with the local peoples. Gradually they married into the local trading and ruling families and, escaping formal Portuguese control, became agents of the African commercial system who sought to secure the best terms they could from any visiting European trader irrespective of nationality.

It may be doubted whether this first period of European involvement with western Africa, from about 1450 to 1600, had much effect on the course of its history. The only Europeans consistently involved were the Portuguese, who were not strong outside the Gold Coast and who were really only interested in controlling some aspects of trade, and these only in a few selected areas of the coastlands where new opportunities had been opened up for a few members of the ruling and trading classes. Perhaps the main changes were that a few Africans acquired some acquaintance with Christianity and with elements of the Portuguese language—a pidgin variety of which became the lingua franca of coastal commerce for some centuries—and that western African farmers were introduced to some new crops and fruits, usually of tropical American provenance, which they quickly adopted if they were more productive than their established cultigens. For example, corn (maize) was more productive than millet and cassava more productive than yams under certain conditions.

Trade versus sovereignty

The new era of maritime intercourse with the outside world was probably of marked significance only on the Gold Coast. There new avenues of wealth had been opened up for some of the Akan in trade at the coast. There too a new political problem had emerged of how to ensure regular and profitable commercial dealings with the Europeans, while at the same time preventing the coastal footholds, which the Europeans required as entrepôts, from subverting the sovereignty of the indigenous states. This was a very real problem, because the coastal kingdoms were small and divided among themselves in competition for the trade with the Europeans. Elmina certainly, and to some extent Axim also, did in fact develop independent jurisdictions over the mixed European, African, and mulatto trading communities that developed beneath the walls of the forts. Beyond these, the difficulty of maintaining large and effective forces of European soldiers in the tropics meant that the Portuguese could only exert power through African allies. The coastal people were thus able to maintain the principle that the land on which the forts were built was not ceded, but only leased. If the Portuguese lost their allies' confidence, the latter could refuse to supply or help defend the forts, or even destroy them altogether (as happened for the first time at Accra in the 1570s).

The rise of the Atlantic slave trade. The coastal peoples generally, and those of the Gold Coast in particular, were certainly prepared to welcome the merchants of other European nations so as to decrease their dependence on the Portuguese. The first Europeans effectively to break into the Portuguese monopoly of sea trade with western Africa were the Dutch, who had been some of the principal distributors in northwestern Europe of the Asian, African, and American produce imported into Portugal and Spain. After the northern Netherlands had revolted against Spanish rule, however, and Philip II of Spain (who since 1580 had been king of Portugal also) had sought to punish the Dutch merchants by excluding them from the Iberian ports, they were stimulated to organize oceanic trading ventures of their own and from 1598 established on the Gold Coast the first Dutch trading posts in western Africa.

Portuguese and Dutch rivalry

The principal targets for Dutch aggression, however, were in the East Indies and in the Americas, and the effective assertion of Dutch power against Portugal in western Africa was a by-product of the success of the Dutch West India Company in destroying Spanish naval power in the Caribbean and in embarking on the conquest of the plantation colony that the Portuguese had established in Brazil. As a result, by the end of the 1630s the Dutch had established themselves as principal suppliers and customers of the Spanish plantations in the Caribbean, while in Brazil they were themselves in possession of a plantation colony.

The production on American plantations of tropical produce, of which sugar was the most important, and especially the marketing of this produce in western Europe, were extremely profitable activities. But plantation agriculture in the tropics required large and regular supplies of cheap labour. America did not have these, but, just across the Atlantic, western Africa seemed to have relatively great quantities of productive labour. As early as the 1450s, the Portuguese had begun to transport some African slaves to supplement the meagre labour resources of their own country (especially of the southern provinces they had reconquered from the Moors), and their own plantations in Madeira, the Cape Verde Islands, and, ultimately, on the islands of the Gulf of Guinea had come to be dependent on African slave labour. The Spaniards, and subsequently other Europeans, in America thus naturally came to look to Africa to make good their labour shortage, and a slave trade to the Caribbean had commenced on a small scale in the 1520s.

Spain had little in the way of trade in Africa itself, so the authorities gave out contracts for the supply of slaves to other merchants, and the ultimate suppliers were usually Portuguese. From the 1570s onward, Portugal's slave traders had a further American market in their own colony of Brazil. The large slave population on their Gulf of Guinea islands was getting out of control, with the result that many Portuguese planters were abandoning these islands and reestablishing their activities in Brazil.

Comparative statistics in the slave trade

By the time the Dutch West India Company entered the scene in the mid-1620s, in all probability about 400,000 slaves had been imported into the Americas, and the annual volume of imports had risen to about 10,000 a year. This was a sizable business compared with the earlier trades in African slaves to Europe (which had virtually ceased by the 1550s, when perhaps 50,000 slaves had been imported) or to the Atlantic and Gulf of Guinea islands (which by the 1620s had probably received rather more than 100,000 slaves, but whose imports were then less than 500 a year). The demand for slaves in the Americas was also beginning to increase rapidly. In the wake of the Dutch defeat of Spanish naval power, English and French colonists were beginning to settle the smaller Caribbean islands and were seeking to exploit their soils for tropical agriculture with as many slaves as possible, while Spanish official restrictions on the volume of slave imports in their territories were hardly effective.

The labour needs of the plantations it had conquered in Brazil made it imperative that the Dutch West India Company should begin to secure slaves from Africa, and it very quickly recognized that the supply of slaves to European colonists elsewhere in the Americas was an extremely profitable business. To ensure its sources of supply, the Dutch company embarked on the conquest of all the Portuguese bases on the western African shores. Some of these, for example the island of São Tomé in the Gulf

of Guinea, were subsequently lost to it, but by 1642 the Dutch were firmly in possession of the Portuguese forts on the Gold Coast, and they dominated Europe's trade with western Africa from Cape Verde to the Niger delta. By the 1660s probably as many as 15,000 African slaves were being landed in the Americas each year. Nearly half of this trade was Portuguese, for this nation had by then recovered its possessions in both Brazil and Angola and had a near monopoly of the trade south of the equator, but the remainder, the growing part, was in northern European hands.

The Dutch West India Company's activities in the Atlantic slave trade aroused interest throughout the ports of northwestern Europe, and soon merchants from France, Britain, Germany, and Scandinavia, as well as private Dutch traders, were competing with the Dutch company. French and British competition soon became of major importance. Both countries were resentful of the growing economic power of the Netherlands that was based on foreign trade, and both possessed colonies in the Americas. Their governments decided that their colonists should not be dependent on Dutch merchants for their supplies of labour (nor, for that matter, for capital or for the marketing of their produce in Europe). Through grants of monopolies of their nations' trade with western Africa or the West Indies, French and English merchants were encouraged to form companies strong enough to challenge the power of the Dutch West India Company, and these challenges were supported by the warships of their respective royal navies.

Between 1652 and 1713 there was a succession of wars involving France, Britain, and the Netherlands. The main battles were usually fought far away from Africa, but throughout the period the traders of each nation sought to increase the number of their trading posts on the western African coast and to deny trade to their rivals. A large number of new forts was built, and these forts were constantly changing hands.

By 1713, however, a pattern of European activity had emerged that was to remain more or less constant until the Atlantic slave trade was brought to an end during the first half of the 19th century (and that was considerably to influence the subsequent partition of western Africa between European empires). The hold that the Dutch had established over Europe's oceanic commerce was destroyed, and Britain and France competed with each other for its inheritance. The Anglo-French wars of the 18th century had less direct effect on western Africa than did the earlier wars involving the Dutch, but the development of trade with western Africa to supply slaves for their American colonies continued to be an important aim of both countries.

Anglo-French competition

France emerged as master of the coastal trade north of the Gambia River, where it had taken the strategic naval base on Gorée Island, close by Cape Verde, from the Dutch in 1677 and was developing a fort and town at Saint-Louis at the mouth of the Sénégal River as a major commercial centre. By and large, however, this part of the coast produced relatively few slaves, and the French companies operating from Saint-Louis sought compensation by developing the trade of the Sénégal basin in gum and hides and by penetrating upriver toward the alluvial goldfields of Bambuk. Little of permanence was achieved in the 18th century, however, in part because of the resistance of the local peoples, but mainly because of the growing naval power of Britain. Britain's strategic interests almost invariably led it to occupy Gorée and Saint-Louis in each of the sea wars of the century, and from 1758 to 1779 its government attempted to consolidate its conquests there into a formal colony.

But the Colony of the Senegambia was not a success. Britain's merchants were not willing to follow up its naval and military successes in this region, and French traders were allowed to creep back. The main results of Britain's initiative were to interrupt French imperial ambitions in the Sénégal valley for nearly a century, and, on the British side, to contribute to a growing opinion—associated particularly with the loss of the North American colonies and the views of Adam Smith—that formal empire was less important and valuable than the independent operations of a growing host of individual British traders, operating wherever there was profit to be found under the general cover of British naval supremacy.

British traders competed with the French on the Gambia River, where both nations' companies maintained forts, and also established themselves to some extent on the coast of Sierra Leone, but initially the main centre of British activity in western Africa was the Gold Coast. Because the Gold Coast had been the scene first of the major Portuguese and then of the major Dutch successes in western African trade, its peoples were better organized than most to provide European traders with what they sought, and the fact that they could supply gold as well as slaves remained of major importance.

During the period of intensive competition in the later 18th century, the number of major European forts on this 300-mile-long coastline had risen to approximately 30. The Dutch West India Company still had more Gold Coast forts than anyone else, and these by and large were the strongest and best maintained. But their headquarters at Elmina was now rivaled by the British castle at Cape Coast only a few miles away, and indeed almost every Dutch fort had an adjacent British establishment competing with it. The French never succeeded in getting a permanent foothold on the Gold Coast; Swedish and German ventures eventually came under Dutch or British control, and traders from Denmark retained forts only on the eastern Gold Coast, where there was little gold and slaves were the main articles of trade.

Although the formal position of the Dutch on the Gold Coast remained strong throughout the 18th century, and indeed into the 19th century, they steadily lost trade to the British merchants. One reason for this was simply that Britain had displaced the Netherlands as the major naval and sea-trading power in western Europe and, through the developing Industrial Revolution, was better able to supply overseas traders with cheap goods for the world market. The British were also more successful than their rivals in adapting the nature of their operations to changing conditions of trade.

The Dutch give way to the British

The Dutch on the Gold Coast and the French in Senegal tended to hold to the view that the African trade should be conducted through large corporations, which had military and administrative as well as commercial responsibilities and which were rewarded for these by their monopoly rights. Such companies had been essential if the original Portuguese monopoly of African maritime trade were to be broken, or if the national interest in it were to be maintained in the subsequent period of militant European rivalry for the trade. The intensity of the competition during about 1650 until about 1713 had made the business of building or capturing and of maintaining and defending coastal forts extremely expensive. On the other hand, the breakthrough into trade with Africans had been made, and the ever-increasing demand for slaves led to a great widening of European commercial activities along the coast. Thus by the beginning of the 18th century there were few parts of the coastlands where African rulers and merchants were not prepared and organized to sell slaves in some numbers.

Hence the future lay not with the cumbersome companies, with large amounts of capital locked up in costly forts, but with a host of small European traders, who were not tied to particular shore installations but were seeking for the best terms of trade they could find along the coast. This type of European trader was naturally welcomed by African rulers, newly embarking on trade with the Europeans, who had no wish to see the latter establish permanent bases on their coasts, which, as the Gold Coast forts had done, might develop political claims to challenge their own traditional jurisdictions. It was the British traders, protected by their country's command of the sea and backed by the abundant supplies of goods and capital produced by its revolutionary economic growth, who were most successful in exploiting the new pattern of trade.

The old monopoly companies had never been strong on the coast between the Gambia River and the Gold Coast. In the north this was largely due to the effects of the

breakdown of the early Portuguese attempts at controlling the trade. Farther south, the territories of modern Liberia and Côte d'Ivoire had not excited much interest among early European traders. Their coastline was treacherous for navigation, and their thick forests supported a scanty population that was little organized for commerce. Such trade as had developed, principally in ivory and in agricultural produce, did not warrant much investment in shore establishments. Thus, as the American demand for slaves increased, the Upper Guinea coast became frequented by increasing numbers of small traders who competed bitterly with each other for slaves.

The greatest opportunities for the new class of individual European traders lay not to the west but to the east of the Gold Coast, where populations were denser and much better organized, not only politically but also commercially and specifically in their access to developed trade routes linking them with inland centres of population, production, wealth, and organized government. European sea captains knew this coast well enough, for the best sailing routes to return to Europe from the Gold Coast ran via the Gulf of Guinea and its islands. European trade had been held back there only because there had been no staple export to compare with the gold of the Gold Coast. Nevertheless, their ships' need of provisioning for the return voyage, the local trade between the Niger delta and the Gold Coast, and the early slave trade to São Tomé and Brazil had led to a demand developing among the African peoples for European manufactures such as metals and metalware, cloth, guns and ammunition, and spirits, and there was a network of traders and trade routes to expand trade with the interior if a staple African export could be found.

The rapidly increasing demand for slaves as West Indian and American plantation production began to boom provided this staple. In the second half of the 17th century, Dutch, French, English, and Portuguese traders became increasingly involved in trade on the coast between the Gold Coast and Benin, which soon in fact received the name of the Slave Coast. Initially the company-fort pattern of trading was applied here, but it never took root to the extent that it had done on the Gold Coast, in part because the local rulers insisted that the forts should be built in their own inland towns. As the slave trade further developed, both on the Slave Coast itself and also further east, in the Niger delta region, it became typically an enterprise of individual European or American traders or small partnerships, who acknowledged the authority of the local rulers and paid the fees and duties these demanded.

The Slave Coast

As has been seen, in the 1620s, on the eve of the great growth of the Atlantic slave trade that followed the Dutch entry into it, the number of African slaves reaching the Americas was about 10,000 a year. In the last quarter of the 17th century, the average annual American import was some 25,000, and the total number of African slaves imported during the century has been estimated at 1,494,000. In the following century it operated on a vastly greater scale; the best available estimate for the number of Africans imported into the New World in the 18th century is some 5.2 million. After about 1815, and especially after the 1840s, the measures taken to outlaw European and American slave dealing, and also—more significantly—the possession of slaves in the Americas, began to take effect, and in the 1860s the Atlantic trade was finally brought to an end after a further 2,780,000 slaves had been landed in the Americas.

The peak of the Atlantic slave trade seems to have been reached in the 1780s, when on average some 78,000 slaves were brought to the Americas each year. About half these slaves were transported in the ships of British merchants. Their nearest competitors, the French and Portuguese traders, carried each about a fifth of the total. Subsequently the French trade (and also the Dutch and Danish trades) virtually ceased as a result of the British blockade of Europe during the Revolutionary and Napoleonic wars of the turn of the century, and the British predominance was even more marked, possibly 60 percent of the trade being in British hands, compared with perhaps 25 percent for Portugal and 15 percent for North American merchants.

The slave-trade era. All the estimates for the volume of the Atlantic slave trade that have been given so far are for numbers of slaves landed in the Americas, as such numbers are generally more readily ascertainable than figures for slaves leaving Africa. A fair proportion of these slaves never reached the other side of the Atlantic because of deaths from disease, maltreatment, or maritime disaster. Evidence from the 18th and 19th centuries, when the vast majority of the slaves were transported, suggests that on average the loss may have been about 15 percent; in earlier times losses are likely to have been higher, perhaps averaging 20 percent.

Not all the slaves were taken from western Africa as defined in this article. Considerable numbers were always taken from Africa south of the equator, and in the 19th century the measures taken to stop the North Atlantic slave trade were quicker and more effective than those against the trade across the South Atlantic. It seems safe to suggest that, up to and including the 18th century, 60 percent of the slaves were taken from the western African coasts from the Sénégal River to the Cameroons and that in the 19th century the proportion dropped to about one-third. It is thus possible to arrive at the following estimates for the loss of population to western Africa.

Comparative statistics of population loss

Population Loss Due to Atlantic Slave Trade

	from Africa as a whole		from western Africa north of the equator	
	arriving overseas	leaving Africa	percentage	estimated total
Before 1600	290,000	370,000	60	220,000
1601–1700	1,490,000	1,870,000	60	1,120,000
1701–1810	5,150,000	6,130,000	60	3,680,000
After 1810	2,780,000	3,270,000	33	1,090,000
Total	9,710,000	11,640,000	—	6,110,000

It is not easy to assess what effects such a loss of population may have had on western Africa and on the course of its history. In the first place, it must not be forgotten that almost all statistics concerning the slave trade involve some degree of estimation. Those used here are based on the analyses of the available data by the American historian Philip D. Curtin and by the Canadian historian Paul E. Lovejoy; they are unlikely to be more than ±20 percent from the reality.

Second, there is really no means of knowing the size of the population of western Africa at any time during the period of the Atlantic slave trade. Working backward from the population data available in the 20th century (which are not themselves always very reliable), and from the evidence these provide for rates of growth, it is possible to suppose that at the beginning of the 18th century, when the Atlantic slave trade was entering its dominant phase, the total population of western Africa may have been about 25,000,000, and that its natural rate of increase may have been some 0.15 percent per annum. Although these estimates can be little more than guesses, they do tend to suggest that the commonly held idea that the export slave trade actually depopulated western Africa is not likely to be right.

When the slave trade was at its height during the 18th century, the export of slaves was averaging 45,000 a year. This loss would have been about equal to the assumed natural increase in population, so that the effect might have been to have checked population growth rather than to have actually diminished the population. In earlier centuries, or in the 19th century, it would not even have had this effect: population would have been growing, albeit more slowly than with no export of slaves.

But these are gross calculations that take no account of the uneven selection of slaves for export. Since the American planters, and hence the slave traders, looked in particular for fit slaves in the prime of life, between about 15 and 35 years old, it may be argued that robbing western Africa of people particularly from this group of its population would especially tend to reduce births and thereby reduce the capacity of the population to maintain its numbers. On the other hand, however, the planters preferred their slaves to be male, and only about a third of

those exported were women. Thus, since western African men who could afford it were polygynous, the birth rate may have been less affected than might have been expected. There is also evidence to suggest that the fitter or more intelligent slaves were often kept at home, and that less fit individuals were in many ways prepared to deceive the European buyers as to their age or condition.

It can also be argued that, since some parts of the coast saw the export of many more slaves than did others, the regions adjacent to these coasts suffered much more severely than the overall figures for western Africa as a whole might suggest. In the peak period of the 1780s, the distribution of exports along the coast was approximately as follows: from the Senegambia and Sierra Leone, about 7,000 slaves a year (about 15 percent of the total from western Africa as a whole); from the Gold Coast, about 9,400 (20 percent); from the Slave Coast and the Benin region, about 16,000 (35 percent); and from the Niger delta and the Cameroons, about 13,400 (29 percent). The three last zones—Lower Guinea—today have populations as dense as any to be found in tropical Africa, and the available evidence suggests that their population was also relatively great in the 18th century—certainly by and large denser than that of most parts of Upper Guinea.

It is therefore possible to conclude that the largest numbers of slaves came from just those regions that could most afford to export population. It is also unlikely to be a coincidence that it was this same area—from the Gold Coast to the Cameroons—which was the most highly developed coastal region in terms of government, economic production, and trade. It was only in areas of low population and poor indigenous organization that foreign slave traders ever needed to set out to capture slaves for themselves. This naturally made the Africans involved hostile to further dealings with the traders, while it also tended to reduce the power of the population to maintain and feed itself, so that in both cases supplies of slaves were ultimately fewer. For the most part, the European traders bought the slaves they needed from African merchants and rulers who had organized to offer slaves for sale.

About half these slaves were unfortunates in their own societies: criminals, the mentally or physically handicapped, debtors or those who had been sold for debt or pledged as security for a debt, those who had offended men of power or influence, or simply those who in some way had become outcasts from the family and tribal systems. Selling such people was usually simply an alternative to keeping them in some kind of servitude in domestic society or, in more extreme situations, condemning them to execution or to serve as human sacrifices in the festivals of ancestral or land cults.

The remainder of the slaves exported were strangers to the societies that sold them, sometimes unwary travelers or border villagers who were kidnapped, but for the most part prisoners of war. Europeans sometimes argued that African kings went to war often with the prime purpose of securing slaves for the slave trade. In the 19th century, when the Europeans themselves had outlawed the slave trade, this argument was used to justify the advance of European colonial rule. On the other hand, in the 18th century, some European slave traders claimed that the acquisition of slaves was simply a consequence of wars which were natural occurrences. From this they argued that they were actually doing a service to such captives and to humanity by buying them and selling them into hard labour on the American plantations. They claimed that they were rescuing the slaves from the danger of being executed or of becoming human sacrifices and that slavery under civilized Christian masters was preferable to slavery in primitive, pagan African society.

Despite the speciousness of the latter claim, the 18th-century slavers' argument seems nearer the truth than that of the 19th-century abolitionists. African wars, like wars anywhere else, were the consequence of rivalries for wealth and power between states. Whereas elsewhere the wealth and power of a monarchy might be measured in terms of the amount of territory it controlled, or in terms of the monetary value of its resources, the prime measure of both power and wealth in Africa was people. By and large land in Africa had very little economic value. There was almost invariably far more land available than there were people to cultivate it or to develop its mineral and other resources. The key to the strength of a kingdom thus lay in its ability to gain control of human energy, and an obvious way to do this was to take people away from its neighbours and rivals. This, indeed, was how western African kingdoms had come to be built up, by the natural rulers of particular small kinship groups securing for themselves and their units more clients and slaves than their neighbours, and by using them to extend their power over these neighbours and even farther afield.

Slaves as the measure of wealth and power

People then were the important resource. Often, indeed, a person was the unit of value in which other resources were measured: thus the value of horses, or guns, or parcels of trade goods was often expressed in terms of the numbers of slaves (*i.e.*, disposable people) for which they might be exchanged. If more people were available, then more land, or gold or iron or salt, might be exploited or more trade might be done (the environment was hostile to transport animals, so that trade depended heavily on the availability of porters and canoemen). Thus there would be greater surpluses available to support the monarch, his household, his administration, and his army, and to maintain specialized manufactures, crafts, and services.

The purpose of wars was thus to increase the power and wealth of a kingdom by increasing its human power and diminishing that of its rivals. The ruling philosophy cannot therefore have been one favouring the export of slaves. But it was one in which the economic value of a person was very well established. With the growth of trade, and especially of international trade which made available desirable commodities that seemed as valuable or sometimes more valuable than people, it was natural for African kings and their traders to think of selling some men and women in exchange for these commodities, and especially so if the foreign merchants who offered these commodities were themselves interested in acquiring slaves.

Pre-European slave trading

This situation had first arisen, and at a very early stage, in the trans-Saharan trade. Labour was needed to work the Saharan salt deposits, and the civilizations of the Mediterranean and Middle East had long had a demand for slaves. Some North African and Middle Eastern exports, particularly perhaps horses, were so valuable in the Sudan that its kings were quite ready to exchange some of their scarce human power to secure these. However, the problems involved in marching slaves across the Sahara with its scarce and widely separated resources of water were formidable. Although reliable estimates are lacking, it is generally supposed that the trans-Saharan slave trade could rarely if ever have transported more than 6,000 or 7,000 slaves a year. After the middle of the 17th century, however, the demand of the Atlantic trade for slaves was practically insatiable, and, as has been seen, at its peak during the 18th century, each year about seven times as many slaves were leaving the western African coasts.

A high proportion of these slaves, nearly a third, were being exported—as has been seen—from the Niger delta region. The communities of Ijo (Ijaw), Ibibio, and Efik fishermen and salt makers, who controlled the waterways to the interior, developed city-states whose whole fortunes came to be bound up with the slave trade. Most of their slaves were brought from their immediate hinterland. It is probably significant that some of this hinterland, particularly that inhabited by the Igbo (Ibo) and the Tiv, today has the highest population densities to be found anywhere in tropical Africa—some Igbo densities being as much as 1,500 persons to the square mile. Igbo country is not rich in natural resources, however, and its water supplies are poor. In the 20th century one result has been that many Igbo have emigrated to sell their labour in other parts of Nigeria. It is not unreasonable to assume that something of this population pressure may have already been evident during the slave-trade era and that many communities in the hinterland of the Niger delta could only survive and prosper by selling some of their people.

This was almost certainly a unique situation. It is doubtful whether there could have been anywhere else in western Africa a buoyant surplus of labour to encourage this kind

of exploitation. It is noteworthy, indeed, that, although the American demand for slaves was rising steadily from about 1630 onward, every other part of the coast seems quite soon to have reached a figure for slave exports which thereafter remained more or less constant until the 19th century. Then two things happened: first, there was a breakdown in the system of law and order that had hitherto operated in Yoruba country to the west of the Niger delta, with the result that exports from the Slave Coast began to increase; and second, after the early 1860s, the American demand fell off sharply. There is a strong implication that in the 17th and 18th centuries the other major slave-exporting regions with relatively large populations had developed politico-economic systems which were able more or less consciously to calculate the balance of advantage for themselves of engaging in the export slave trade. By and large their conclusion seems to have been that it was more profitable to exchange some of their labour for European goods than it was to keep it all at home, but that it was dangerous to export more than a certain controlled quantity.

Other trading commodities

The Atlantic slave trade was not simply a rape of African labour to serve European purposes in the Americas. In any case, slaves were not the sole African exports by sea during the slave-trade era. Gold, gum, hides, timbers, palm oil, and other commodities were also traded, and the European merchants needed to buy large quantities of provisions to feed the slaves during the Atlantic crossing. The slaves and other exports had to be purchased, and in exchange Africans received supplies of other goods—cloth, metals, tools, knives and other hardware, guns and ammunition, beads and small manufactures, tobacco and spirits—that, however much their prices may have been inflated in relation to their cost in Europe or America, were thought by their purchasers to be at least as valuable as the slaves or other goods they had sold.

From the African point of view, the main importance of the slave trade may have been that it led to a great growth of all kinds of trade at the coasts and to a considerable stimulation of economic and political activity and organization for some distance into the interior of Guinea, a region which had hitherto been remote from the main centres of trade in western African history, which had been in the Sudan. It is difficult to quantify this growth of commercial activity, but coastal exports and imports combined, negligible prior to about 1500, may—in contemporary values—have been worth something like $3.5 million a year by 1700 and $8 million or more by 1800. This was a considerable trade by the standards of the time, as can be seen from comparison with an estimate in the 1850s of the value of the trade of Kano, then the most prosperous of the Sudanic kingdoms, at some $500,000 a year.

Some consequences of this rapid increase of trade on the Guinea Coast were fairly general. One was the appearance on the coast itself of a new class of African merchants, who freed themselves from some of the restrictions of traditional society and were able to accumulate personal wealth and power to rival that of the local kings. Sometimes, indeed, as at Komenda on the Gold Coast or at Opobo in the delta, the new men actually set themselves up as kings. This development owed much to the direct influence of the European trade on the coast. Some Europeans settled more or less permanently in Africa, married local women, and created new merchant dynasties, such as the Brews of the Gold Coast. Others of the new men were former slaves who had returned from America, particularly perhaps from Brazil to the Slave Coast. Many were local Africans, but usually men who had started by gaining useful contacts and training in the service of European merchants. All the new men were experienced in European ways and often secured for their sons elements of a European education.

Systems of currency

The growth of Guinea's international trade encouraged the spread and acceptance of regular systems of currency. Many, perhaps most, of the Guinea currencies were already extant on the coast when the Europeans arrived, but the growth of trade meant that certain of these became the sole acceptable currency within relatively large and well-defined areas, such as iron bars in Upper Guinea and much of the Niger delta, ounces of gold dust on the Gold Coast, and cowrie shells on the Slave Coast. Another development was the emergence on a considerable scale of production for sale. In Asante (Ashanti), for example, some villages were given over to the production of a particular commodity, such as cloth, and some of the chief men ran plantations with slave labour (as also, in the 19th century, did the kings of Dahomey).

The emergence of the trading city-states of the Niger delta represented a social revolution as well as a political innovation. The kinship system gave way to the "House" system, by which both freemen and the large numbers of slaves needed to operate trading canoes and strategic and trading settlements were bound together by common economic interests into large corporations headed by the leading merchants. On the Gold Coast and Slave Coast, however, political development was more akin to that which had earlier taken place in the Sudan under the influence of trans-Saharan trade.

Behind the Gold Coast the original centres of Akan trade and power had been north of the forest and northward-looking. The growth of trade at the coast led to new developments among the settlements in the forest, which had hitherto served only to produce gold and kola nuts for the northern trade. In the 17th century three major forest kingdoms emerged: Denkyera in the west, Akwamu to the east, and between them, Akyem. These competed with each other in expansion parallel to the coast to control as many as possible of the paths of trade to the European forts. Akyem lost, while Denkyera achieved such an overweening power that some of its northern tributaries secured guns and new techniques of political and military organization from Akwamu and rose in revolt. This was the effective origin of the new monarchy of Asante based on Kumasi, situated in the central forest where the major trade routes from the Gold Coast converged and met with the major routes of the Hausa and the Mande-Dyula traders from the north.

By the beginning of the 18th century the power of Denkyera had been crushed, and in the next 70 years Asante armies went on to build up an empire that in the north engulfed Bono and Gonja (Guang) and levied tribute on Dagomba and in the south incorporated or made tributary virtually all the small states that had been involved in the rise of Denkyera, Akyem, and Akwamu. The only part of modern Ghana that was not under the sway of Kumasi was the central coastlands, where the small Fanti (Fante) states, gaining some measure of protection from their close association with the various European forts, began to come together in a federation to resist Asante influence.

The southward expansion of Oyo

As the Atlantic trade began to expand east of the Gold Coast to the Slave Coast, similar political developments began to manifest themselves in its hinterland also. Toward the end of the 17th century the northernmost Yoruba kingdom, Oyo, began to turn away from its traditional rivalry with the adjacent savanna kingdoms of Nupe and Borgu and to use its cavalry to assert control of the trade routes through the open country southwestward to the small Aja states on the coast in which the Europeans had established trading posts. A measure of control was also asserted more directly to the south over other Yoruba peoples and kings in the forest. Here a boundary was established with the kingdom of Benin, which in the later 17th century decided that it was in its interests to open up its ports to European merchants and to sell slaves to them to secure a share of the goods they were offering.

By the early 18th century strains caused by the virulent competition between the European traders and their African associates were leading to the dissolution of traditional social and political controls among the Aja, who had a number of small kingdoms under the nominal leadership of the king of Allada. The resultant disorder was not to the liking of the kings of Dahomey, the youngest of these monarchies, who, in colonizing the northern marches of Aja territory, had evolved much more authoritarian and militant forms of government and society. Between 1724 and 1734, Dahomey enforced its concepts on the other Aja peoples by conquest and began to build up a centralized state to control the entire Slave Coast.

Initially Dahomey, wishing to conserve labour, was reluctant to sell slaves to the Europeans. This was not to the liking of Oyo, whose foreign trade was dependent on its being able to sell slaves to the Europeans on the Slave Coast. Nor was it really in the long-term interests of the centralized trading system of the Dahoman kings themselves, for these had little but slaves to offer in return for the guns and other goods they needed to buy from the Europeans. From 1726 to 1748 the consequence was continual warfare on the Slave Coast, the ultimate results of which were that Dahomey was led by force of arms to recognize Oyo suzerainty and that the slave trade became firmly established, both under strict royal control in Dahomey's port of Whydah (Ouidah) and, increasingly, in ports beyond its control to the east, such as Badagri and Lagos.

The armies of Denkyera, Akwamu, Asante, and Dahomey all made use of firearms, and Akwamu seems to have been a pioneer in the development of tactics suited to the new weapons. The Portuguese had not imported guns into western Africa on any scale and as a matter of policy had sold them only to their allies. In the highly competitive trading situation that followed the Dutch breaking of the Portuguese monopoly, all the European trading nations vied with each other to sell guns, and they soon became an essential article of trade.

The muskets exported to western Africa were cheap varieties specially made for the African market. They were undoubtedly serviceable for hunting and for the protection of crops from the depredations of birds and wild animals, but it is debatable how significant their acquisition was in the rise of the new Guinea kingdoms of the 17th and 18th centuries. It is notable, for instance, that Oyo's military victories were attributed to its employment of cavalry and that it was consistently successful against Dahomey, whose soldiers were equipped with muskets.

Guns as symbols of prestige as well as power

From the African point of view, guns were expensive—though less perhaps to buy than to keep in working order and supplied with powder. In this respect, guns were not dissimilar to horses and, like horses, they became the particular prerogative of kings and their henchmen, symbols of prestige but also elements in the growth of royal power at the expense both of subjects and of unfortunate neighbouring peoples who did not evolve state systems. It followed that the trade in firearms—and also in those African exports, such as slaves and gold, which were most in demand by the European sellers of firearms—tended to be under strict royal control. Similarly, competition to secure supplies of guns and ammunition—as of horses in the Sudan—and to deny such supplies to rivals was a factor of some significance in the new era of power politics in Guinea.

By the 18th century the kingdoms and empires of Guinea, especially those of Lower Guinea, though commonly less extensive, were as powerful as those that had been established in the western Sudan. As with the latter, control of trade with the outside world had been an important element in the growth of these states. However, the relationships between the growth of trade and the growth of states in the two systems, those of Guinea and of the Sudan, were not entirely identical. The rise of the Guinea states, for instance, was associated with a growth of trade that was particularly dependent on the growth of a demand for one particular African export, slaves. Any cessation of this demand was therefore likely to create strains for the Guinea states; and Asante, Dahomey, and Oyo had hardly established paramountcies in their areas when Europeans began seriously to question both the morality and the economic value of slave labour.

It is also noteworthy that the Guinea peoples did not generally establish networks of itinerant traders extending far beyond their homelands, as the Mande and Hausa had done. The Asante, Dahomey, and Benin trading systems seem essentially to have been state corporations operating under royal control or license only within the boundaries of their political influence. The northern limits of the Guinea trading system hardly extended more than 300 miles from the coast. Conversely, Mande-Dyula and Hausa traders had ventured far beyond the political and military limits of their kings' authority (which, in the case of the Hausa, were always very restricted). Even after the rise of the Atlantic trade and of the Guinea kingdoms, traders from the north continued to be of prime importance to the commerce of Guinea and, sometimes, to its political life. The community of Muslim northerners settled in Kumasi, for example, sometimes played a significant role in Asante politics, and the northern trade also led to Islām gaining important ground among the Yoruba. Thus the Guinea states were also likely to be affected by any important changes in the political and economic life of the western Sudan.

THE ISLĀMIC REVOLUTION IN THE WESTERN SUDAN

The Moroccan occupation of the Niger Bend in 1591 meant that the domination of the western Sudan by Mande or Mande-inspired empires—Ghana, Mali, Songhai—which had persisted for at least five centuries, was at last ended. The Songhai kings were pushed southeast into their original homeland of Dendi, farther down the Niger close to Borgu, and Mande political power was limited to the so-called Bambara—*i.e.*, "pagan"—kingdoms of Segu (Ségou) and, later, of Kaarta, upstream and to the west of Macina. In and around the Niger Bend itself, the long-term effect of the Moroccan conquest was to open up the country to the Tuareg and Arabized Berber tribes of the Saharan fringes. By the middle of the 18th century the descendants of the Moroccan conquerors, who had settled down in the Niger Bend cities as a ruling caste, the Arma, had become tributary to the desert pastoralists.

Dominance of Tuareg and Berber tribes

The same tribes operated, or at least profited from, the trans-Saharan trade, and some of them had acquired leading positions in western African Islām. The Kunta tribe of Arabized Berbers had become preeminent in both these respects by the 18th century. It dominated the salt trade to Timbuktu, and in the person of Sīdī Mukhtār (died 1811) it had produced a spiritual leader so respected among the Muslims of the western Sudan that the Kunta were able to exercise on the quarrels between the pastoral tribes a mediating influence which was clearly to the general benefit of commerce and urban society.

Mukhtār's position was due to qualities of learning and holiness that were in part personal but also in large measure due to his leading role in the Qādirīyah, one of the Muslim brotherhoods (tariqas) in which particular traditions of both sanctity and learning were passed on from teacher to teacher. These brotherhoods or religious orders had arisen with the growth, from about the 11th century onward, of mystical currents of Muslim thought (especially in eastern Islām, where the Qādirīyah had begun). Mysticism proved to be congenial to Berber society in North Africa (where the Tijānīyah order evolved in the 18th century), and from here the tariqa entered the Sahara, arriving in western Africa by the beginning of the 16th century.

Hitherto Islām had been spread in western Africa essentially by merchants who, in order to secure their livelihood, chose to accommodate themselves and their religion within the pagan social and political framework that existed where they settled—which for the most part was only in the towns. But with the coming of the tariqa—of which the Qādirīyah was one of the first and, until the Tijānīyah began to advance in its tracks in the 19th century, certainly the foremost—western Africa began to experience the growth of organized groups of devout Muslims who were both specifically trained and morally compelled to work toward a true Islāmic society. Moreover, if the people and their rulers remained irresponsive or hostile, it was the Muslims' duty to preach the doctrine of conversion by force, through the jihad, divinely justified war or rebellion against rulers who were pagans or not true Muslims.

This doctrine was particularly attractive to the Fulani, who, as has been seen, were scattered in stranger communities between the agricultural settlements throughout the western African savannas. As the wealth, organization, and power of agricultural and urban society increased, so there was less scope available for the free movement of the Fulani cattle and less freedom for their herdsmen. The Fulani were subject to increased pressures to pay rents, taxes, and services to the rulers of the settled commu-

nities who, from the Fulani point of view, were aliens who had no natural right to these things. Although the bulk of the Fulani were pagans, they were, as pastoralists, naturally open to influence from the Saharan pastoralists who were Muslims and among whom the tariqa had been established. The Fulani also had ethnic links with the long Islāmized Tukulor of the far west, and they had a considerable and influential Muslim clerical class of their own. The Fulani clerics were thus particularly receptive to the doctrine of jihad and, throughout the Sudan, could ally themselves with considerable numbers of disgruntled and mobile pastoral kinsmen to make jihad a military reality.

The Fulani jihad

The earliest known Fulani jihad occurred in Bondu, close to the Islāmized Sénégal valley, where in the second half of the 17th century Fulani clerics succeeded in taking over political power from local Mande rulers. Early in the following century, considerable numbers of Fulani began to do the same in alliance with the local Muslim Mande traders in the nearby Fouta Djallon. By about 1750 a Muslim theocracy had been erected whose leaders were soon engaged in organizing trade to the Upper Guinea coast on which European traders were active. In the second half of the 18th century the same pattern was repeated in the Fouta Toro, the homeland of the Tukulor, for there, though the dispossessed rulers were Muslims, as a group they were too self-interested and exploitative to suit the clerics.

News of these developments in the westernmost Sudan naturally spread through the Fulani diaspora to more easterly territories influenced by the teaching of Sīdī Mukhtār and other like-minded tariqa divines. In 1804 the most famous of the western African jihads was launched in Hausaland by Usman dan Fodio.

Usman was the leading Fulani cleric in Gobir, the northernmost and most militant of the Hausa kingdoms. This was in a disturbed state in the 17th and 18th centuries. The growth of Tuareg power in Aïr on its northern frontiers had led the Gobir ruling class to seek compensation to the south and southwest, in the territories of Zamfara and Kebbi. There the breakup of the Songhai empire had led to a power vacuum, which had been an encouragement to Fulani settlement. The kings of Gobir, like other Hausa monarchs, were at least nominally Muslims, and for a time Usman had been employed at their court. He then used the influence he had gained to develop a Muslim community of his own, some miles away from the capital, governed according to the strict principles of law preached by the Qādirīyah. The kings of Gobir gradually came to the conclusion that they could not afford to tolerate this independent jurisdiction within their unsettled kingdom and began to take steps against the Muslim community. By 1804 the situation became such that Usman felt he had no alternative but to declare a jihad and to adopt the role of an independent Muslim ruler (*amīr al-mu'minīn* or, in Hausa, *sarkin musulmi*).

Both sides appealed for wider support. While the Hausa kings proved incapable of concerted action against the movement of Islāmic rebellion, discontented Fulani and oppressed Hausa peasantry throughout Hausaland welcomed the opportunity to rid themselves of vexatious overlords and arbitrary taxation. Within three years almost all the Hausa kings had been replaced by Fulani emirs who acknowledged the supreme authority of Usman. The most serious fighting was in and around Gobir itself, where the maintenance of large Fulani forces in the field alienated the local peasantry. Fortresses had to be established for the systematic reduction of the country, and in the process the old kingdom of Gobir was destroyed and two major military encampments, Sokoto and Gwandu, eventually emerged as the twin capitals of a new Fulani empire.

The rise of the Fulani empire

The core of this empire was composed of the three large former kingdoms of Katsina, Kano, and Zaria (Zegzeg), in which, together with the smaller former kingdom of Daura, a Fulani aristocracy had taken over the Hausa system of government and had brought it into line with the principles of Islām as stated by Usman. But the jihad had not stopped at their boundaries. Hausa clerics and adventurers joined with the Fulani in creating new Muslim emirates farther afield, among the pagan and hitherto largely stateless peoples of the Bauchi highlands, for example, and in the open grasslands of northern Cameroon, where there were large numbers of Fulani. There the vast new emirate of Adamawa was created. In the south Fulani and Hausa clerics intervened in a succession dispute in the old pagan kingdom of Nupe and by 1856 had converted it into a new emirate ruled from Bida. There had also been considerable Fulani and Muslim penetration into northern Yorubaland, and, in about 1817, its governor rashly invoked Fulani and Hausa aid in his rebellion against the king of Oyo. The governor's new allies took over, the new emirate of Ilorin was created, and the disintegration of the Oyo empire was accelerated.

The only serious check to Fulani conquest was in Bornu. By 1808 the forces of Fulani rebellion and invasion had reduced its ancient monarchy to impotence. Bornu and Kanem, however, had their own clerical class and tradition, and in the latter province arose a new leader, Muḥammad al-Kānemī, who asserted that the Fulani clerics did not have a unique right to interpret Muslim law for the government of humanity. Al-Kānemī was able to inspire a spirited national resistance, which by 1811 had turned the tide against the Fulani. By 1826 he was the effective master of a new Islāmic state, though the traditional kings were maintained in office until 1846, when the puppet of the time rebelled against al-Kānemī's son and successor, 'Umar, but was defeated and killed.

Usman dan Fodio was a scholar and theologian who had little inclination for the political and military direction of the movement he had inspired. His main role was to maintain the jihad's spiritual and moral force and direction, and he left a remarkable memorial of this in his innumerable writings. The practical commanders of the jihad were his brother, Abdullahi, and his son, Muḥammad Bello, who were men of action as well as considerable scholars. These two eventually became joint viceroys of the new empire, Bello ruling its eastern half from Sokoto and Abdullahi the western half from a seat of government at Gwandu. They oversaw the installations of the provincial emirs, received tribute from them, and endeavoured to ensure that their governments and systems of taxation followed the principles of Muslim law and were not arbitrary and extortionate. Gradually the original scholarly and clerical impulse of the jihad weakened (though it was never wholly forgotten), and the emirs tended to become more representative of the military Fulani aristocracy, which tended to intermarry into the old Hausa ruling class. Standards of scholarship decayed and Hausa, rather than Arabic, became the language of administration. But for half a century or more after the jihad, some 200,000 square miles of territory enjoyed a unified system of relatively impartial law and administration, and this was much to the advantage of its agriculture, industry, and trade.

Both Sokoto and Gwandu were in the extreme northwest of the empire, where the jihad had had its origins and where it continued longest, for Kebbi was never entirely subdued. It is possible also that it was in this direction, looking up the Niger toward the Kunta and to the considerable Fulani population of Macina, that it was thought that there might be further advances. Doubtless it was for these reasons that Abdullahi settled at Gwandu with responsibility for the western empire. The main Fulani successes, however, were to the southeast in Bello's sphere, and it was Bello who in 1817 succeeded to his father's titles of caliph and *sarkin musulmi*.

The jihad in Macina

When, about 1818, a jihad began in Macina, it was an independent movement led by a local Qādirī Fulani, Ahmadu ibn Hammadi. Ahmadu was certainly cognizant of Usman's jihad, and the circumstances in which his own movement was born were very similar to those that had occasioned the jihad in Hausaland. Ahmadu established an independent Muslim community that brought him into conflict with his local, pagan Fulani chief, who was unwise enough to call for help from his suzerain, the Bambara king of Segu. The result was a general rising under Ahmadu that established a theocratic Muslim Fulani state throughout Macina and extended to both the ancient Muslim centres of Jenne (Djenné) and Timbuktu.

The third major western African jihad of the 19th century

was that of al-Ḥājj ʿUmar Tal (*c.* 1797–1864), a Tukulor cleric from the Fouta Toro. As a young man, ʿUmar went on the pilgrimage (hajj) to Mecca (hence the honorific al-Ḥājj), and in all spent some 20 years away from his homeland. Twelve of these were spent at Sokoto, where he married a daughter of Bello's. He also spent some time with al-Kānemī in Bornu, and he shared with both men in the great revival of Muslim scholarship in the western Sudan. But ʿUmar had a wider experience of the Muslim world than either Bello or al-Kānemī, and he must have been acquainted with both the modernism of Muḥammad ʿAlī Pasha's regime in Egypt and the new puritanism of the Wahhābīyah in Arabia. Also while in Arabia he seems to have been appointed the western African caliph of the relatively new Tijānīyah brotherhood, which was appreciably more activist in its demand for reform than the Qādirīyah. About 1838 ʿUmar arrived home in the Fouta, where he quickly became estranged from the local clerics. In 1848 he moved away with such followers as he had to Dinguiraye, on the borders of the Fouta Djallon. There he built up a community of his own, attracting and training military and commercial adventurers as well as religious reformers. His community traded with the Upper Guinea coast for firearms and was consciously conceived as the nucleus for a new state. In 1852 the Dinguiraye community came into conflict with the adjacent Bambara chiefs. A jihad was launched northward through the gold-bearing valleys across the upper Sénégal, where in 1854 the Bambara kingdom of Kaarta fell. ʿUmar then turned west down the Sénégal toward his own homeland and the French trading posts. But he was repulsed by the French, and after 1859 he sought to join with the Fulani of Macina in the conquest of the more powerful Bambara kingdom of Segu. The Macina Fulani were opposed to the idea of a Tijānī power advancing into their own Qādirī zone in the Niger valley and even gave some aid to Segu. After ʿUmar's forces had conquered Segu in 1861, they continued eastward, and, finding that Ahmadu's somewhat autocratic and intolerant regime had estranged the longer established Muslim communities, they established ʿUmar's hegemony as far as Timbuktu (1863).

The empire of al-Ḥājj ʿUmar Tal

In less than 10 years al-Ḥājj ʿUmar's armies had conquered an empire almost as large as that of the Sokoto Fulani. It does not, however, appear to have been as well founded. Outside of the Niger valley and the major trading settlements, the majority of its inhabitants were basically pagans who had only accepted Islām because they had been subjected to the shock of conquest by comparatively small bodies of well-armed and well-led adventurers. This was a different situation from that in which relatively large numbers of Muslim Fulani and Hausa had poured out from the old Hausa states into territories already prepared for them by the infiltration of Islām and the presence of Hausa traders and Fulani settlers. In ʿUmar's empire individual captains, exempt from taxation themselves, settled down to exploit their conquests as virtually independent fiefs. Along the Niger axis of empire there were both old, established Muslim towns and Fulani communities whose inhabitants regarded the Tijānī Tukulor as upstarts. In 1864 ʿUmar was killed attempting to suppress a Fulani rebellion in Macina, and for many years his son and successor, Aḥmadu Seku (d. 1898), had to compete for his inheritance with his father's numerous other relations and captains.

The most important result of ʿUmar's conquests was that they established the Tijānīyah as the most powerful tariqa in western African Islām, and this, together with the earlier consolidation of Muslim power in the east under Sokoto, ultimately ensured that Islām became the dominant religion throughout the western Sudan, and one capable of peaceful expansion deep into Guinea. Already circumstances had changed, however, since the Fulani cavaliers had built up the Sokoto Muslim empire. Al-Ḥājj ʿUmar's empire builders relied on horses for their mobility, but they were also musketeers who knew the value of trade with the Europeans at the coast. Even more significantly, they had already come into conflict with, and had been worsted by, European military and political power advancing inland from the coast.

THE GUINEA COASTLANDS AND THE EUROPEANS (1807–79)

In addition to the Islāmic revolution in the Sudan, the major themes of western African history in the 19th century are the successful campaign against the export of slaves, the trade that for the previous 200 years had been the mainstay of Guinea commerce; the search by both Africans and Europeans for a stable new relationship in the absence of slave trading; and the failure of the major African kingdoms to adjust to the new economic and social circumstances swiftly enough to withstand growing European pressures.

The abolition of slavery

These three themes are closely interwoven in the course of events in Africa. It should be noted, however, that the major decisions regarding the abolition of the slave trade were taken outside Africa and were responses to economic and political changes and pressures in Europe and America. Many of the Christian churches had never accepted the morality of trading in human beings, and the 18th-century Evangelical movements in Protestant Europe led to open campaigning against the Atlantic slave trade and also against the institution of slavery itself. These things were equally condemned by new secular currents of thought associated with the French Revolution. Because plantation production in tropical America was no longer as profitable a field for investment by northern Europeans as industry, or as trade with other parts of the world, the propaganda against the slave trade began to take effect. Denmark outlawed slave trading by its citizens in 1803, Great Britain in 1807, the United States in 1808, Sweden in 1813, The Netherlands in 1814, and France (for the second time) in 1818.

The most significant of these actions against the slave trade was that of Britain. British ships had been by far the largest carriers of slaves at the end of the 18th century, and only Britain really possessed the naval resources necessary to secure enforcement of anti-slave-trade laws on the high seas. Furthermore, when Portugal, Spain, and some American countries expanded their slave trading to meet the deficiency caused by the British withdrawal, they met with strong opposition from Britain. The underlying reason for this was that Britain, more than any other European nation, had considerable amounts of capital, experience, and goodwill accumulated in trade with Africa. When British merchants tried to develop new lines in African trade to replace their old slave trade, however, they commonly found that, as long as their European or American rivals continued to buy slaves, African kings and merchants were generally not willing to organize alternative exports. Economic interest therefore combined with abstract morality to induce successive British governments to bring pressure on other governments to outlaw their slave trades and to permit the British navy to help enforce their laws on their ships at sea.

But these measures did not stop the export of slaves from Africa. Some nations, notably France and the United States, whose own naval controls were fitful, objected strongly to British warships stopping, searching, and, if need be, arresting their ships at sea. Furthermore, as long as there was a market for slaves in the Americas (*i.e.*, until all the American countries had abolished the institution of slavery), there were lawless individual traders who felt that the profits to be gained from running slaves across the Atlantic more than outweighed the risk of arrest. Except when actually embarking slaves on the African coast or unloading them in American waters, the chances of interception at sea were in fact quite small. Although the British navy maintained in western African waters an anti-slave-trade squadron of up to 20 ships, which between 1825 and 1865 arrested 1,287 slave ships and liberated about 130,000 slaves, during the same period about 1.8 million African slaves are believed to have been landed in the Americas.

The final cessation of the export of slaves from Africa to the Americas took place toward the end of the 1860s. The decisive factor was the abolition of slavery in the United States in 1865. Slavery was then legal only in Cuba and Brazil—and only to the 1880s—and the risks of transporting slaves to these two markets became too high. Before this, British governments had already embarked on

a policy of taking or supporting active steps in Africa to stop slaves from being offered for sale on its coasts and to encourage the production of alternative exports. The immediate results of these efforts were often not very great. For example, many African governments and merchants were no more inclined than many European or American governments or merchants to enforce or to observe the anti-slave-trade treaties that British officials wished upon them. They saw no reason why their economic interests, which were bound up with slavery and trade in slaves, should be subordinated to the new economic interests of British traders following what was to them the capricious decision that slavery and the slave trade were wrong.

The British naval presence

What was significant was that Britain, through its desire to stop the export of slaves from western Africa and to protect the interests of British merchants desiring to trade in other commodities, maintained a substantial naval presence in western Africa and was also acquiring new political, commercial, and missionary presences. These led to increasing interference in the domestic affairs of African societies and their governments.

This interference began with the British naval squadron's need of shore stations to serve as bases for its patrolling ships and as landing places for the appreciable numbers of slaves it was intercepting. The slave ships had to be taken to a European jurisdiction to be condemned, and the slaves they had carried could not simply be returned to the societies that had sold them, but needed to be maintained as wards for whom Britain accepted political and moral responsibilities. British political officers and missionaries therefore became established on African soil, and the area of their activities was continually increasing. Explorers were penetrating the hinterland in search of new avenues for trade. British traders were competing with the slave traders. The liberated Africans were branching out into trade for themselves or simply returning home with new Western and Christian attitudes. All these were apt to ask for political or missionary support, and, behind this, naval action could be called upon if there were difficulties with the local African rulers.

For the greater part of the 19th century the prime centre for British naval, political, and missionary activities on the western African coast was Sierra Leone. Toward the end of the 18th century the Sierra Leone peninsula had been chosen by British philanthropists as a suitable place to which Africans who had been taken to Britain as slaves and freed there, or who had fought on the British side in the U.S. War of Independence, might be repatriated. A first group was sent out and settled on the site of the future Freetown in 1787. Although many of the early settlers did not survive, others were brought, and by 1811 Freetown had a liberated African population of about 2,000.

After a first false start, the philanthropists hoped to find funds for the maintenance of their settlement by placing it under the control of a company they had floated to trade with the interior. But the only trade that prospered in the Sierra Leone region was the slave trade, in which the company naturally did not engage, and after 1799, their colony, whose indeterminate constitutional status caused many difficulties, was able to survive only with the help of annual grants-in-aid from the British government. Eventually, in 1808, the British government agreed to take over direct responsibility for the colony from the Sierra Leone Company.

British coastal settlements

Freetown was not the only British settlement on the western African coast. Officials of a descendant of the old African Company of slave trading days still occupied a number of forts on the Gold Coast and one in the Gambia, and these were now meant to provide support for British merchants engaged in other trades. In 1817, after the settlements on the Sénégal, which had been in British hands during the Napoleonic Wars, had been handed back to France, a considerable number of British traders and their African associates moved to the mouth of the Gambia River and established there the new settlement of Bathurst (Banjul). Neither the Gambia nor the Gold Coast were exclusive British spheres, however. The French were strong competitors on the former, and both the Dutch and the Danes still held forts on the latter, and where European interests were divided, there was no certainty that the British settlements would prosper in competition with slave traders, nor that they could be developed as effective bases in an active campaign against that trade.

From 1808, however, British policy required that such a base be maintained in western Africa, and Freetown was the obvious choice. It had one of the best natural harbours on the coast, and it was already experienced in the resettlement of liberated slaves. Christian missions had begun to establish themselves there since 1806, and it became the seat of a British governor and of anti-slave-trade courts and the headquarters of the navy's western African squadron. During the next 60 years this squadron was to swell the population under British rule by landing some 60,000 men and women, from all over western Africa, whom it had taken from arrested slave ships. Freetown's only serious disadvantage was that it was at one end of the slave-exporting coast. In 1827, therefore, the British navy also began to use the island of Fernando Po in the Gulf of Guinea as an alternative base and freed-slave settlement. But this activity aroused the interest of the Spanish government, which had had a legal claim to the island since 1778, and in 1834 the settlement was abandoned.

From 1814 to 1824 the British governor and commander in chief at Freetown was Sir Charles M'Carthy, an active military man who thought that the most effective means of achieving Britain's aims in western Africa was to extend its formal dominion over the most vexatious outlets for the slave trade. The home government for a time countenanced this policy and in 1821 transferred the forts on the Gambia and Gold Coast to M'Carthy's administration. During the 10 years of his government its expenditure quadrupled to nearly $400,000 a year. There was no corresponding increase in British trade (nor any diminution of the slave trade), however, with the result that the cost had to be met by British taxpayers who were antagonistic to spending money on colonies. In 1824 M'Carthy's forward policy led him to make common cause with the Fanti against Asante claims to overlordship on the Gold Coast. In the war that followed, however, he was defeated and killed, and the British government decided that it should withdraw from all formal commitments in western Africa except at Sierra Leone.

In fact the most prosperous British trade was developing on a part of the coast on which there was no British interference other than naval action to intercept slave ships and to secure anti-slave-trade treaties. That was the Niger delta. British shipping had been paramount there when the British slave trade had been abolished in 1807, and the merchants of the delta city-states had quickly adapted themselves to offering palm oil as an alternative export to slaves.

British trade with the Niger delta

Britain's Industrial Revolution had occasioned a growing demand for vegetable oils as lubricants and for the manufacture of soap, and the new Lancashire cotton industry was producing in quantity a commodity with which palm oil might readily be purchased. By the 1830s the British purchases of palm oil in western Africa were worth nearly $2 million a year. About nine-tenths of this trade was initially with the Niger delta. The oil palm grows throughout a belt just behind the western African coast, and the oil from its fruit was already widely consumed and traded locally. Africans of the delta were much quicker and more successful in developing an export trade in palm oil than were those of other coastal regions. One reason was simply that the oil was not easy to transport in quantity, and its value was not high in proportion to its bulk. Canoe transport was thus easier and cheaper than headloading or cask rolling, and the delta afforded a ready-made system of waterways. But its hinterland also had an unusually dense population in a relatively poor agricultural environment and therefore had both a greater need to exploit the semiwild palm trees than was usually the case and more labour with which to do this and to manufacture and transport the oil. Moreover, the collection of the fruit and the manufacture of the oil were traditional household activities, and to exploit these for export necessitated a commercial system that was both wide and intensive and, in addition, highly responsive and flexible.

The small, highly competitive city-states of the Niger delta, built up and ruled by merchants, could exploit the overseas demand for palm oil much more quickly and efficiently than was possible elsewhere. In Liberia and western Côte d'Ivoire, for example, the trading network, like the population, was thin and little advanced. Elsewhere export trading (for example, in slaves or gold) had been directed, or at least controlled, by large-scale organizations that were less flexible, politically motivated, and much less responsive to commercial changes; among these were the traditional political hierarchies of large kingdoms such as Benin, Asante, and Dahomey, or the new politico-religious administrations of the Fouta Djallon or of al-Ḥājj ʿUmar. It may be noted, incidentally, that the successful development of palm-oil exports from Yorubaland followed upon the collapse of the Oyo empire there. It was not until about the 1860s, when the total British purchases of palm oil were worth about $6 million a year, that exports from the rest of western Africa, with Yorubaland in the vanguard, began to equal those of the Niger delta.

British official policy toward western Africa remained one of minimum intervention until the 1870s. Indeed, the view that Britain should withdraw from all commitments other than in Sierra Leone was most forcefully asserted by a Parliamentary Select Committee as late as 1865. In fact, however, both positive and negative results of the active British campaign against the Atlantic slave trade made it impossible for the policy of nonintervention to be maintained in practice.

The growth of the spirit of European scientific inquiry during the 18th and 19th centuries combined with a practical interest in finding out what Africa produced besides slaves that could be of value to world trade and what political, economic, and transport systems existed to permit such products to be brought down to the coast, to lead to a great movement of European exploration of the interior of western Africa between 1788 and 1855. This movement was primarily directed from Britain, and from 1805 the British government sponsored many of the major expeditions.

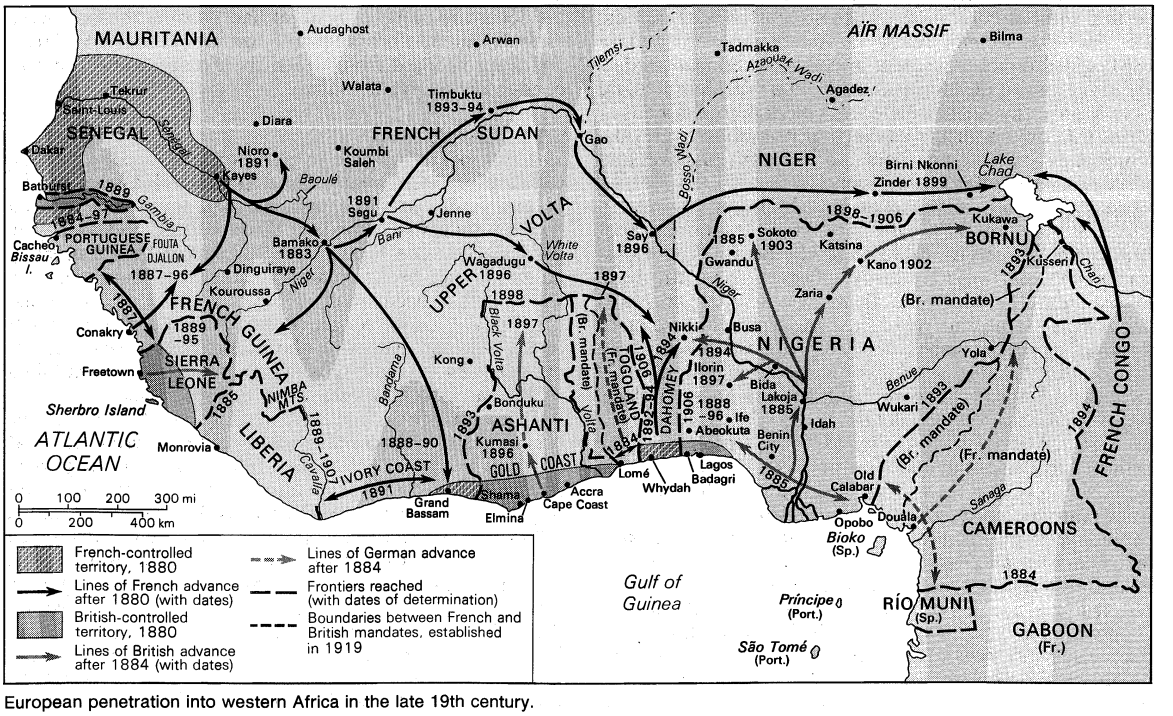
These explorations suggested a possible strategy of breaking through the barrier of the established slave-trading states at the coast by using the Niger River to trade directly with the interior. This seemed attractive after the rejection of the M'Carthy policy of positive coastal action, and from 1832 onward the British government sponsored or helped to sponsor a number of expeditions designed to develop navigation up the Niger. By 1854 quinine and the steamship had solved the technical problems of navigating the lower river, but a new political problem had been created in the objections of both black and white traders in the delta to their established trading system being bypassed in this way.

The exodus from Sierra Leone

By the middle of the century the development of the liberated African community in Sierra Leone under the tutelage of British administration, churches, and education meant that some of its members were providing a considerable reinforcement for the British interest in western Africa. Economic activities in Sierra Leone itself were limited, and Sierra Leoneans were soon finding their way along the coast as independent pioneers of trade and Westernization or as auxiliaries to British traders, officials, and missionaries. Their most significant influence was in Yorubaland. By the 1840s at least half the liberated Africans were of Yoruba extraction, and by this time their homeland afforded considerable scope both for independent traders and for people seeking to introduce Christian and Western ideas and ways into African life. Both these circumstances derived from the failure of the Oyo empire in the 18th century to establish a stable form of central government capable of maintaining a firm control over the provinces it had conquered. There remained a dangerously uncertain balance of power between the king and the traditional chiefs of the capital.

Such a situation was by no means unique in the history of the kingdoms of Guinea (or, for that matter, of western Africa). Dahomey seems to have avoided it only because its kings, initiating their kingdom through the conquest of peoples who were not of their own stock, had been able to build up an unusually authoritarian form of government. But in both Benin and Asante traditional kinship organizations imposed restraints on royal authority, and tensions could develop when it came to sharing the rewards of empire and trade. There was a near disastrous civil war at Benin at the end of the 17th century, though the king emerged from it with his authority strengthened, apparently because he was able to play off the town chiefs against the chiefs of his palace. Asante had come into being as an alliance of petty kingdoms against Denkyera, and its kings, desiring to be more than merely primus inter pares, began by entrusting the new kingdom's conquests mainly to the chiefs of their capital, Kumasi. But the chiefs then sought to control the monarch, and the latter had to turn for help from the provincial rulers to release him from this situation. Ultimately, however, from the time of Osei Kojo (*c.* 1764–77), the kings secured their

From J. Fage, *An Atlas of African History*; Edward Arnold (Publishers) Ltd.



European penetration into western Africa in the late 19th century.

preeminence throughout the kingdom by building up a new hierarchical military and civil administration, which was responsible uniquely to them and which limited the power of both sets of chiefs.

At Oyo the traditional town chiefs, who commanded the army of the capital, converted the kings into puppets during the 1750s and '60s. About 1774 they gave the throne to a king, Abiodun, who escaped from their control and used provincial forces to establish royal authority over the capital. After Abiodun's death (*c.* 1789), the provincial chiefs began to act with increasing independence. When (*c.* 1817) the viceroy of the north invited Fulani aid to help consolidate his rebellion, the result was not simply that the kings of Oyo lost their northern provinces to the Fulani; they also lost control over the northern trade routes on which they depended for their vital supplies of horses and slaves, and eventually (*c.* 1836) they had to evacuate their capital to the south. By this time there was no longer any central authority, and everywhere ambitious men were vying with each other to create personal dominions over as many clients and slaves as possible.

One consequence of this situation was a great increase in the number of slaves available for export from nearby Dahomey, which by 1818 had thrown off the last vestiges of Oyo suzerainty and was soon sending its armies deep into Yorubaland, and from independent traders at ports such as Lagos and Badagri. It was the close attention given by the British navy to these coasts that had led to the build-up of Yoruba former slaves in Sierra Leone. By the 1840s considerable numbers of these people were returning to Lagos and Badagri and, especially, the new inland town of Abeokuta, originally built up as a refuge where Egba (southern Yoruba) peoples could withstand pressures from Ibadan, the most powerful of the new Yoruba political units, and from Dahomey. The advent of the settlers from Sierra Leone soon brought British missionaries, and a new British-aligned influence was added to the tangled web of Yoruba politics.

The first British consulate

British officialdom soon followed. In 1848 a British consulate had been established for the Gulf of Guinea to maintain British interests in the complex situation arising from the splintered politics of the Niger delta and the beginnings of navigation on the river itself.

The consuls joined with naval officers in attempts to stop the king of Dahomey from exporting slaves, and, when repulsed, turned to Lagos, where they saw their opportunity in a split in the royal family. In 1851 the British navy restored to his throne a deposed monarch who had promised to stop the Lagos slave trade. He was in fact powerless to do this without continued British support. Lagos became the seat of a second British consulate in 1853, and in 1861 it was annexed. British and Sierra Leonean traders endeavouring to develop palm oil trade with Yorubaland were soon trying to persuade the colonial government at Lagos that only a further advance of its authority into the hinterland would stop its wars and its export of slaves and allow their own affairs to prosper.

The first serious advance of British power in western Africa occurred on the Gold Coast. After the withdrawal of British officials and troops in 1828, the British Gold Coast traders took on a young army officer, George Maclean, to represent their interests there. Maclean negotiated a peace with Asante and established an informal jurisdiction through the coastal states, which brought security for both British and Asante merchants. The consequent fourfold increase in British trade combined with the uncertain legal status of Maclean's jurisdiction to bring British officials back to the forts in 1843. In 1850 they took over the Danish forts also, but the continued Dutch presence on the coast prevented them from raising an effective revenue from customs duties, and they quarreled with the coastal peoples over the issue of direct taxation. They therefore failed to erect an effective coastal administration of their own on the foundation laid by Maclean, and they equally rejected alternatives proffered by educated Africans in cooperation with the coastal chiefs. Trade declined, and Asante's armies began to invade the coastlands to protect its interests there. Eventually the Dutch were led to withdraw altogether (1872) and the British to invade Asante and destroy its capital and to declare the whole coast a colony (1874).

Three-quarters of a century of turmoil following the British decision to campaign against the Atlantic slave trade and to foster the interests of legitimate trade and Christian civilization in western Africa had therefore resulted in the establishment of the new colony of Sierra Leone and direct British intervention in African affairs in much of the most prosperous area of the old trade—the Gold Coast, Lagos, and the delta and lower river of the Niger.

African sovereignty had also been infringed between Sierra Leone and the Ivory Coast where, inspired by the Sierra Leone example, private U.S organizations had settled freed slaves for whom there was no place in their own society prior to 1863. British and French merchants questioned the right of the settlers to control and to tax their trade and, since formal U.S. policy was anticolonial, the result, in 1847, was the proclamation of the Republic of Liberia. The settler government then embarked on a long struggle to assert control over the local Africans. Because, unlike a colonial government, it had no metropolitan resources or finance to help, this was a prolonged business.

The Republic of Liberia

The growth of British trade, and of British influence and power, in western Africa was by no means to the liking of the government, traders, and navy of France—Britain's principal competitors in the previous century. But France's mercantile interest in western Africa was not as strong as Britain's, and its traders there received less official and naval support than did the British. Not until the 1870s and the opening of the European scramble was any serious effort made to develop the trading footholds that were established on the coast between Senegal and Sierra Leone, on the Ivory Coast, and between the Gold Coast and Lagos.

France's main effort in western Africa was devoted to developing its old interests in Senegal following the British withdrawal in 1817. Initially an attempt was made to replace the former business of exporting labour to the West Indies by developing a local plantation economy. By the 1820s this was foundering, and matters then drifted until the arrival in 1854 of a new governor, Louis Faidherbe, a soldier with experience in the conquest of Algeria and in the government of its peoples. Faidherbe's concept was to secure control of the exports of the westernmost Sudan by extending French military and political control up the Sénégal River and to encourage local African production of the peanut (groundnut) to help meet the growing French and European demand for vegetable oils. By the time of his departure in 1865, Senegal had become the prototype for subsequent European colonization in western Africa and a springboard from which the French could think of conquering the whole Sudan.

Senegal as a colonial prototype

COLONIZATION

The European scramble to partition and occupy African territory is often treated as a peripheral aspect of the political and economic rivalries that developed between the new industrial nations in Europe itself and that were particularly acute from about 1870 to 1914. Its opening has commonly been taken to be either the French reaction to the British occupation of Egypt in 1882 or the Congo basin rivalry between agents of France and of Leopold II of the Belgians that led to the Berlin West Africa Conference of 1884–85, both of which are seen as being exploited by Bismarck for purposes of his European policy.

In western Africa, however, it seems fair to say that the beginnings of the scramble and partition were evident at least a generation before the 1880s and that they were determined by the local situation as much as or more than they were by European domestic rivalries. Already during 1854–74, the logic of the situation in western Africa had led France and Britain to take the political initiatives of creating formal European colonies in Senegal, in Lagos, and in the Gold Coast. All along the coast, in fact, the traditional African political order was becoming ineffective in the face of European economic and social pressures. For most of the 19th century these pressures had been predominantly British, but in the 1870s French compa-

nies began to offer effective competition to the British traders not only in Upper Guinea, where they had always been strong, but also on the Ivory Coast, in the ports immediately to the west of Lagos, and even in the lower river and delta of the Niger. An unstable situation was developing in which the European traders were likely to call for further intervention and support from their governments, and especially so if the terms of trade were to turn against them. Low world prices for primary produce during the depression years from the 1870s to the mid-1890s certainly caused difficulties for Europeans trading to western Africa and led them to think that an increase in European control there would enable them to secure its produce more cheaply.

The changing balance of power in western Africa was not confined to the coastlands. By the 1870s formal French and British armies had already ventured into the interior and had inflicted defeats on such major African powers as those of al-Ḥājj ʿUmar and Asante. In 1879 Faidherbe's heirs on the Sénégal River had launched the thrust that was to take French arms conquering eastward across the Sudan to Lake Chad and beyond.

By the end of the 1870s France and Britain, therefore, were already on the march in western Africa. The principal effect of the new forces stemming from domestic power rivalries in Europe itself—the most dramatic of which was the appearance in 1884 of the German flag on the Togoland coast, between the Gold Coast and Dahomey, and in the Cameroons—was to intensify and to accelerate existing French and British tendencies to exert their political and military authority at the expense of traditional African rulers.

French incentives to expand

There can be no question but that, by the end of the 1870s, the advance of the British interest in western Africa had been more rewarding than the advance of the French interest. Devoting their attention primarily to the active economies of the Niger delta, the Lagos hinterland, and the Gold Coast, British traders had secured $24 million of business a year, compared with the French merchants' trade of $8 million, three-quarters of which was concentrated on the Sénégal River. Initially, therefore, the French had much more incentive for expansion than the British.

Britain was already in political control of the Gold Coast, and the arrival of the German treaty makers in Togo and in the Cameroons in 1884 hastened it to declare its protectorate over most of the intervening coastline on which British traders were active. The gap left between Lagos and Togo was swiftly filled by the French, and from 1886 they also established formal authority over all other parts of the coastline that were not already claimed by the governments of Liberia, Portugal, or Britain. In this way the baselines were established from which France subsequently developed the colonies of Dahomey, the Ivory Coast, and French Guinea.

France's advance inland from these southern coasts was subsidiary, however, to the main thrust, which was eastward from the Sénégal region through the Sudan. The glamour of its past had persuaded the French that the Sudan was the most advanced, most populated, and most productive zone of western Africa. Once they had reached the upper Niger from the Sénégal (1879–83), the French forces had a highway permitting them further rapid advances. By 1896 they had linked up with the troops that had conquered Dahomey (1893–94) to threaten the lower Niger territories which British traders had penetrated from the delta.

The rapid French advance across western Africa from the Sénégal River had denied the British any chance of exploiting the commercial hinterland of the Gambia River and had severely restricted their opportunities from Sierra Leone. Government and mercantile interests nonetheless were able to agree on the need for British action to keep the French (and also the Germans from Togo and from the Cameroons) out of the hinterlands of the Gold Coast, Lagos, and the Niger delta. Asante submitted to an ultimatum in 1896 (the real war of conquest was delayed until 1900–01, when the British had to suppress a widespread rebellion against their authority), and a British protectorate was extended northward to the limits of Asante influence.

From company territory to protectorate

On the Niger, British interests were first maintained by an amalgamation of trading companies formed in 1879 by Sir George Goldie to combat French commercial competition. In 1897 the British government agreed to support Goldie's Royal Niger Company in the development of military forces. Three years later, however, it recognized the foolishness of allowing the company's servants and soldiers to compete for African territory with French government officials and troops and to enforce its monopolistic policies on all other traders within its sphere. The company was divested of its political role, and the British government itself took over direct responsibility for the conquest of most of the Sokoto empire. Thus, although the French eventually reached Lake Chad, they were kept to the southern edges of the Sahara, and most of the well-populated Hausa agricultural territory became the British protectorate of Northern Nigeria. In 1914 this was merged with the Yoruba territories, which had been entered from Lagos during the 1890s, and with the protectorate over the Niger delta region to constitute a single Colony and Protectorate of Nigeria.

As early as 1898 Europeans had staked out colonies over all western Africa except for some 40,000 square miles of territory left to the Republic of Liberia. Portugal had taken virtually no active part in the scramble, and its once extensive influence was now confined within the 14,000 square miles that became the colony of Portuguese Guinea. Germany, the latecomer, had claimed the 33,000 square miles of Togo (together with the much larger Cameroon territory on the eastern borders of what is usually accepted as western Africa). France and Britain remained, as before, the main imperial powers.

France claimed by far the larger amount of territory, nearly 1.8 million square miles compared with some 450,000 square miles in the four enclaves secured by Britain. In other terms, however, France had done less well. Its territory included a large part of the Sahara, and the three inland colonies of French Sudan (modern Mali), Upper Volta (modern Burkina Faso), and Niger were by and large scantily populated and, because of their remoteness from the coast, were contributing little or nothing to the world economy. In 1897 the trade of the four British colonies was worth about $24 million, compared with about $14 million for the seven French territories, and their combined population of more than 20 million was more than twice as great.

The political boundaries established by the Europeans by 1898 (though usually not surveyed or demarcated on the ground until much later) largely determine the political map of western Africa today. The only subsequent change of significance followed the British and French conquests of the German colonies during World War I (1914–18). While the larger parts of both Togo and Cameroon were entrusted by the League of Nations to the French to administer as separate colonies, in each case a smaller western part was entrusted to Britain to be administered together with the Gold Coast and Nigeria respectively. Ultimately British Togo chose to join with the Gold Coast and so became part of the new independent Ghana. The northern part of British Cameroon similarly joined with Nigeria, but the southern part chose instead to federate with the former French Cameroon.

If 20 years had sufficed for the European powers to partition western African lands, at least a further 20 years were needed to establish colonial regimes that were effective throughout all the vast territories claimed by Europe and that were accepted by all the Africans involved. The first problem was a military one.

Feudal levies versus modern technological warfare

Small and mobile columns of African soldiers, led and trained by European officers and noncommissioned officers and equipped with precision rifles, machine guns, and artillery, rarely experienced much difficulty in defeating the great empires created by the 19th-century jihadists. These chose to meet the invaders in pitched battles in which their massed feudal levies, with few modern weapons and limited skill in their use, served only as targets for the superior firepower and discipline of their opponents. Once these battles had been lost, the surviving leaders were usually ready to acknowledge the Europeans

as new overlords. The main problems were really ones of distance and logistics. Thus it was not until 1900–03 that Sir Frederick Lugard's forces were sufficiently established in northern Nigeria to defeat the Sokoto Fulani, while the French "pacification" of the even more remote territory further north, which eventually became their colony of Niger, was not really completed until the 1920s.

A much more serious military problem was often presented by smaller political units, which were ethnically more homogeneous and often more densely populated than the jihad empires. Their subjugation was often a protracted business in which the Europeans had to fight virtually for each settlement. This was the case with the British campaign against Asante in 1900–01, with the subjugation of the Sierra Leone protectorate in 1898–99, and above all, perhaps, with the advance of British power into the densely populated Igbo and Tiv territories, which was hardly complete until as late as 1918. Similarly, the most formidable resistance faced by the French came not from the Tukolor, but from the more southerly empire established from the 1860s onward from the hinterland of Sierra Leone to western Gonja by the Mande leader Samory Touré. Though Samory was a Muslim whose activities did much to consolidate the hold of Islām in his territories, he was not a cleric like Usman dan Fodio or al-Ḥājj ʿUmar. He came from a family of Dyula traders and soldiers, and the principles of his government recalled those of ancient Mali rather than of the jihad empires. Samory established his network of military and political control over territories long subject to Mande commercial penetration and settlement, and a number of campaigns had to be fought against him until he was finally captured and exiled by the French in 1898.

Once the superior firepower and organization of the Europeans had secured their military supremacy, they were faced with an even larger problem; namely, how the small forces they commanded were to maintain a permanent occupation and effective control over the vast territories they had overrun. Lugard, for instance, had conquered the Sokoto empire with only about 3,000 soldiers, only 150 of whom were Europeans, and to administer his northern Nigerian colony of some 250,000 square miles and 10 million people he had a civil establishment of only 200 Europeans. This kind of situation persisted almost throughout the colonial period. At the end of the 1930s, for example, the European establishment available to the British governor of the Gold Coast to control nearly four million people was only 842. It is obvious, then, that the conquerors were often very slow to extend effective rule throughout their empires, and particularly to those parts of them that were most remote, presented serious political problems, or seemed least profitable.

Initial difficulty of administration

No European control could be exercised without the cooperation of large numbers of Africans. This was secured in two ways. First, just as the Europeans had relied on Africans for the rank and file of their armies and police, so their administrations and economic enterprises could not function without a host of Africans employed as clerks, messengers, craftsmen of all kinds, and labourers. All of this employment offered new opportunities to Africans, and to ensure an efficient labour force all European administrations began to supplement and develop the schools begun by the missionaries.

As well as recruiting and training large numbers of Africans as auxiliaries in all spheres of European activity, the colonial powers also came to rely on African chiefs as essential intermediaries in the chain of authority between the colonial governments and their subjects at large. Both the French and the British colonial regimes were essentially hierarchical. The administration of each colony was entrusted to a governor who was responsible to a colonial minister in the government in Europe (in the French case, via a governor-general at Dakar). These governors were assisted by senior officials and a secretariat in the colonial capital, and their decisions and orders were transmitted for implementation to provincial and district commissioners. A district officer, however, could not deal directly with each of the tens, or even hundreds, of thousands of Africans in his care. He therefore gave orders either to the traditional chiefs or to Africans who had been recognized as local rulers by his government, and these intermediaries passed them on to the people at large.

In this connection a difference of theory began to be discernible between French and British policy. The French regarded the local African chiefs as the lowest elements in a single administrative machine. This administration was to be conducted on entirely French lines. The British, on the other hand, came to believe more and more in "indirect rule." British authority was not to reach directly down to each individual African subject. While the British retained overall control of a colony's administration, it was to be made effective at the district level by cultivating and by molding the governments of the traditional African rulers.

The policy of indirect rule

Indirect rule was neither a new nor a specifically British expedient. Maclean had been an indirect ruler on the Gold Coast in the 1830s; Goldie had proposed indirect rule for the empire his Royal Niger Company had hoped to conquer; and, in the early days of their expansion, the French had often had no alternative but to seek to control their newly won territories through the agency of the African governments they had conquered. Once they were firmly established, however, the French almost invariably moved away from the practice. The British, on the other hand, evolved a theory of indirect rule that they tried to apply systematically to their colonies during the first half of the 20th century. This was largely due to the influence of Lugard. In 1900–06 he had seen no other way to control the vast population in northern Nigeria, whose rulers he had defeated, and he had subsequently been promoted governor-general (1912–19) of a united Nigeria, which was by far the most important British colony in Africa. After his retirement to Britain, he became a dominating influence on the formation of colonial administrative policy, so that indirect rule became accepted as the ideal philosophy of government for British tropical Africa.

Not all areas of western Africa were as suitable for Lugardian indirect rule as northern Nigeria. Lugard himself experienced considerable problems in trying to apply it to the largely chiefless societies of eastern Nigeria and to the Yoruba of the southwest, where authority and law were not as clear-cut. In the Gold Coast indirect rule proved more acceptable to the Asante than the direct rule imposed after the conquest of 1900–01. Farther south, however, the Western-style economy and modes of thought had made such inroads that there were endless problems in the implementation of indirect rule, and the full constitutional apparatus for it was hardly installed until the 1940s.

The development of indirect rule also implied a contradiction with an earlier tradition of British colonial government, that of the colonial legislative council. The governors of British colonies were allowed more initiative than French governors and were supposed to exercise this in the interests of their individual territories insofar as these did not contradict the overriding British interest. To help them in this, each colony was equipped with a legislative council that included representatives of local opinion, and this council's consent was normally required before laws were enacted or the colonial government's budget was approved.

The institution of the legislative council had evolved from experience with settler colonies outside Africa; when such councils were introduced into tropical Africa from the 1840s onward, most of their members were colonial officials. A minority of "unofficial" members represented trade and the professions rather than the traditional communities, and these were not elected but were nominated by the governor. However, 19th-century colonial officials, traders, and professionals were almost as likely to be black as white, and the early legislative councils were by no means ineffective vehicles for the expression of African interests and of criticisms of British policy. It was thus possible both for the British and for the educated African elite in their colonies to view the legislative councils as embryo parliaments that would eventually become composed of elected African members who would control the executive governments, which would themselves, through the growth of education in the colonies, become more and more composed of African officials.

The legislative councils as embryo parliaments

Although very little thought was given to the matter, because it was supposed that the development might take centuries, it was supposed that the British colonies in Africa would follow the example of Canada and Australia and ultimately emerge as self-governing members of the empire. The equally remote future for the French colonies, on the other hand, was thought to be the acculturation (*assimilation*) of their people, so that ultimately they would all become full French citizens, the colonies would be integrated with metropolitan France, and the African citizens would share equally with the French-born in its institutions.

Both of these ideals were more appropriate to the colonial situations in western Africa before the great scramble for territory that began in 1879, when the colonies were comparatively small territories in which European influence had been slowly but steadily gaining ground for a considerable period. They were effectively shelved when it came to grappling with the problem of governing the enormously greater numbers of Africans without any real previous contacts with European ways who were quickly brought under colonial rule in the years after 1879. Thus, on the French side, though those born in the four major communes (Saint-Louis, Gorée, Rufisque, and Dakar) of the old colony of Senegal continued to enjoy the French citizenship that they had been granted prior to 1879, other Africans became French subjects (possessing the obligations of citizens but not their rights), who could only qualify for citizenship after stringent tests. By 1937, out of an estimated 15 million people under French rule in western Africa, only some 80,500 were citizens, and only 2,500 of these had acquired their citizenship by means other than the accident of birth in one of the four communes.

In the British colonies, however, where the legislative councils were already a reality, there was a dichotomy between them and the institution of indirect rule. Initially, insofar as this was resolved at all, it was at the expense of the development of the legislative councils. Thus the competence of the council in the Gold Coast was not extended to Asante before 1946, while in Nigeria until 1922 the council's competence was restricted to the small territory of Lagos. It was not until 1922 that any elected members appeared in the councils, and they remained for a generation a small proportion of the total unofficial membership, chosen only by tiny electorates in a few coastal towns. For the rest, the African population remained firmly under British control through the mechanism of indirect rule. The implication was not only that the norms of African society and political behaviour were far removed from those of western Europe but also that the British had by no means accepted that African society and politics would or should evolve in that direction. Those few Africans who had become educated and acculturated in Western ways were not thought to be representative of the mass. There was a move to exclude local Africans from the colonial administration, which became regarded as a professional service, liable to serve anywhere in Africa, with the role of holding the ring until, in some unexplained fashion, the native administrations under indirect rule had developed sufficiently to make British control superfluous.

The colonialists as harbingers of material change

Colonial rule. In fact, of course, the very existence of colonial rule meant that the fabric of African societies was exposed to alien forces of change of an intensity and on a scale unparalleled in the previous history of western Africa. Hitherto remote territories like Niger and Mauritania, where there had been very little change since the introduction of Islām, were from about 1900 suddenly caught up in the same tide of aggressive material changes that had for some time been affecting the coastal societies in Senegal or in the southern Gold Coast and Nigeria. From the African point of view, there was little to choose between the European colonial powers. Portugal, despite the fact that it was virtually bankrupt at the onset of the colonial period, was as significant a bringer of change as France, Germany, and Britain. In fact, in the long run, a strange combination of its poverty with memories of its older colonial tradition were to make Portugal's sense of a *mission civilisatrice* even more pervasive than that of its stronger rivals.

Liberia's formal status as an independent republic did not mean that the forces of change associated with the colonial period were excluded from its territory. Its Afro-American ruling elite were orphaned members of a very rapidly changing Western society, who felt it essential to impose its ethos on black Africa. While colonial administrators often had a narrow, 19th-century concept of government as an arbiter, rather than as an active protagonist of change, the Liberians felt a need actively to enlist the support of Western capital and enterprise if they were to consolidate their rule over African peoples and to maintain the independence of their republic.

Up to 1912 the inexperience and relative weakness of Liberia's ruling elite meant that it achieved little except to run up a dangerous indebtedness to ingenuous and potentially rapacious European investors. In 1925–26, however, the tide began to turn for them when the American Firestone Tire & Rubber Company, worried lest its supplies of raw material should become a British colonial monopoly, secured a new American loan for Liberia and began to operate a one-million-acre plantation concession in the hinterland of Monrovia. The country was now supplied with a sure access to world trade, and its government with the means to achieve a stable revenue. Within 25 years Liberia's foreign trade grew from less than $3 million a year to some $45 million, and government revenue from a mere $500,000 a year to nearly $10 million. The evident dangers that Liberia might become too dependent on a single export crop, and that it and its administration might become sole fiefs of the American company, began to disappear when during World War II U.S. strategic interests caused its government to begin to give aid to Liberia and to develop its first modern port, and when in the 1950s both American and European interests began to exploit Liberia's large-scale deposits of high-grade iron ore. By the 1960s Liberia was on the way to becoming one of the richer western African countries, and the ruling elite began to feel sufficiently secure to share both some of its political power and some of its prosperity with the native peoples.

A cardinal rule for all colonial administrations in Africa before the 1930s was that colonies ought not to be a financial burden on the metropolitan governments and their taxpayers: the cost of colonial administration and development should be covered by the local revenues they could raise. So long as such a doctrine was maintained, it was impossible for any but the richest colonial administrations to devise coherent plans for the economic development of their territories; indeed, prior to the 1940s, the colonial government of the Gold Coast was virtually unique in putting forward such a plan, and then only in the 1920s, which were by and large exceptionally prosperous years.

Sources of revenue

The principal sources of revenue were (1) duties on the trade entering and leaving the territory and (2) direct taxation (usually a poll tax or hut tax). But only those coastal colonies that had already entered the world economy prior to about 1880 had much in the way of trade on which customs duties might be levied or a sufficient internal production of commodities and circulation of money to produce any significant income from direct taxation. Other territories—such as British northern Nigeria, or the French colonies of the Sudan (Mali) and Niger—could not really provide enough revenue to support even the most essential administrative services, such as policing or—for that matter—tax gathering. For some time, therefore, these administrations were in receipt of grants-in-aid from some central source, and it was an attempt to shift this burden from metropolitan resources that as much as anything led the French in 1895 to bring together their western African colonies under a government general and that led Lugard to argue for the unification of the Nigerian colonies, which he eventually achieved in 1912–14. In each case it seemed advisable to use some of the comparatively buoyant revenues of the coastal territories to subsidize the administrations of those in the interior.

It was obvious enough that what was needed was to increase the European commercial penetration of western Africa. But only the prospect of the most lucrative prizes could induce private European investors to place substan-

tial amounts of capital in Africa in advance of adequate European administrations that could guarantee the safety and security of their investments and in advance of the economic infrastructures that would ensure their efficient deployment. The only lure that really operated to attract European investment in advance of the provision of such services was the prospect of rich mineral deposits. The greater part of western Africa's mineral wealth lies in ores such as those of iron, aluminum, and manganese, which are extremely bulky in relation to their value and require very large investments in transport and other facilities before they can be economically worked, and for which there was relatively little overseas demand before the 1930s. The possibilities of diamond mining in Sierra Leone and the Gold Coast were not really recognized until the 1930s.

Early investment incentives

In effect then, it was only the gold of the Gold Coast and Asante forests and, to a lesser extent, the tin of the Bauchi plateau in central Nigeria, that attracted the early attention of European investors.

Modern methods of gold mining first began to be employed on the Gold Coast as early as 1878, but the industry could not make much headway before 1902. By that time the colonial government had taken the decisive steps of defeating Asante, beginning to build a railway system, and establishing an effective civil administration in the relevant areas, which could ensure proper land surveys and some means of controlling and adjudicating disputes over the ownership of land and the validity of concessions of it. Bauchi tin mining began much later, in 1903, but similar, if less acute, difficulties prevented much progress before 1914.

Despite their poverty, and despite the risk of saddling the home governments and taxpayers with unwanted expenditure, colonial governments found that there was no alternative to their providing the basic infrastructures needed by the vast territories they claimed to rule. It was impossible to wait for private European enterprise to provide railways, harbours, telegraph lines, roads, medical services, schools, and all the other things that were needed to support an effective government, let alone to provide some possibility of economic growth sufficient to pay for better government.

French territories. The problems facing the French were much more formidable than those facing the British. The British colonies were essentially based on territories close to the sea, in which European trade had been long established and whose African peoples were already accustomed to producing for the world market. The French had such a colony in Senegal, but from this they had expanded over vast, remote, and thinly populated territories that required very considerable investment before they could be efficiently administered or developed. By and large the French public had appreciably less capital to invest overseas than the British public had. By 1936 it was estimated that, whereas the British colonies in western Africa had attracted about $560 million of capital, the total outside investment in French West Africa amounted only to some $155 million.

French strategy was initially to open up and develop its western African empire from a base in Senegal on the same Sénégal–Niger river axis along which it had been conquered. As early as 1882 work was begun on a railway to link the heads of navigation of the two rivers at Kayes and at Bamako (which became the capital of the French Sudan). But this line was not completed until 1906, by which time it had become evident that Saint-Louis, at the mouth of the Sénégal River, was not capable of development into a modern port, and that the Sénégal was really suitable for navigation for only three months in the year. So first a railway was completed from Saint-Louis to the new harbour of Dakar in the lee of Cape Verde (1885), and then during 1907–24 a line was built directly from Dakar (since 1902 the federal capital for French West Africa) to Kayes to bypass the Sénégal River altogether.

The construction of an effective west–east transport system from the coast to the upper Niger thus took some 42 years to complete, and the only part of it that was profitable was that serving the peanut-growing areas of Senegal. There was a lag of some 20 years after 1924 before the thinly populated and impoverished French Sudan could respond to the stimulus of its improved communications with the outside world. Indeed the only major crop developed for the world market that could withstand the high costs of transport to the coast—over some 700 miles of railway—was cotton, and that only after considerable further investment in irrigation. Ultimately the main economic role of the Sudan was to provide foodstuffs for Senegal, whose peasant farmers found it more profitable to concentrate on growing peanuts for export.

Changes in French economic strategy

By 1914 French economic strategy had shifted from the concept of opening up the inland territories of the French Sudan, Upper Volta, and Niger, to the encouragement of agricultural production in the coastal colonies. To a limited extent, the way was pioneered by European plantations, more especially perhaps in the Ivory Coast. Generally these colonies were made remunerative by administrative pressures to induce African farmers to produce for export. Ultimately, just as the economy of the Senegal had become largely dependent on the export of peanuts, so that of French Guinea became dependent on bananas (though at the very end of the colonial period, European and American capital began the successful exploitation of considerable deposits of bauxite and iron ore), and the economies of Dahomey and of Togo (after its conquest from Germany) became dependent on palm produce. The most dramatic successes were achieved in the Ivory Coast, where considerable exports were developed of coffee, cocoa, bananas, and lumber. Railways were built from suitable points on the coast to facilitate the export of these crops.

In the 45 years from 1912–13 to 1956–57, the French had boosted the foreign trade of their western African empire from about $58 million a year to about $600 million a year, with the result that the revenues available to their colonial administrations increased from about $8.5 million a year to as much as $315 million. (These figures exclude the part of Togo that was incorporated in the French empire only after 1914–18, and the trade and revenue of which by the mid-1950s were worth some $24 million and $4 million a year respectively.) In absolute terms in relation to the total population, which in the same period is thought to have doubled to an estimated 19 million, the results were not so spectacular; in 1956–57 foreign trade per capita overall amounted to about $32 and government revenue to about $17. The significance of the figures is also obscured by the federal system to which all the colonies except Togo were subject and which was deliberately used to enable the richer colonies to help the poorer. The trade and revenue figures cannot be easily broken down between the individual colonies. Whereas the estimated gross national products (GNPs) for Senegal and the Ivory Coast were in the order of $180 and $160 per capita respectively (the former considerably inflated by the colony's possession of the federal capital), only Togo (about $73) and French Guinea and Sudan (about $58 and $53, respectively) were thought to have GNPs per capita higher than $40.

British territories. Each of the four British colonies must necessarily be treated as an independent unit, as each was so treated in British policy. The Gambia was merely a strip of land, averaging only seven miles in width, on either side of 292 miles of navigable waterway penetrating into what otherwise was French Senegal. Even in the 1950s its population did not exceed 300,000, and the possibilities for any sort of development were limited. In fact the colony achieved a fair degree of prosperity by concentrating on the production of peanuts, grown in part by farmers who migrated annually from Senegal for the purpose. By 1956–57 foreign trade was some $60 per capita and government revenue $14.

The Sierra Leone situation was one of a relatively dense population exploiting or even overexploiting a poor environment for its subsistence, and initially the most that was achieved was to develop some palm produce for export. During the 1930s the situation began to change when European companies began to exploit extensive diamond-bearing gravels and to mine high-grade iron ore. By the mid-1950s foreign trade, which had been $14 million ($9

per capita) in 1913–14, had risen to $101 million ($44). About half of this was based on the activities of the foreign-owned mining companies. These provided little local employment; and furthermore, large numbers of people had been led to abandon farming to dig for diamonds on their own account. This gave rise to numerous social, economic, and political problems, because legally the diamond-bearing grounds had been conceded to the European companies. These factors may explain why the increase in government revenue, and hence the capacity of the government to sponsor further development, was low in comparison with other western African territories. It rose from $3.6 million ($2.40 per capita) in 1913–14 to $27 million ($11.70) in 1956–57, a factor of increase of 4.9, which compares unfavourably with a factor of 21.1 for French West Africa as a whole, 11.4 for the Gold Coast, 6.1 for Nigeria, or even 5.9 for the Gambia.

Comparative increases in revenue

The Gold Coast was a complete contrast, indeed one of the most successful examples of colonial development anywhere in British tropical Africa. The people of its coastlands were long accustomed to world trade, and indeed to British rule, with the result that the Gold Coast entered the colonial period with a very high level of economic activity. In 1912–13 its foreign trade was worth $42.5 million ($28.30 per capita) while government revenue was $6.5 million ($4.30 per capita). Subsequent development was facilitated by the possession, within a manageable area that was adequately but not too densely populated, of a considerable variety of resources.

The first railway was built inland from Sekondi in the southeastern Gold Coast between 1898 and 1903 with the dual purpose of supporting gold mining and ensuring political control of Asante. This railway subsequently was used for the removal of manganese ore and bauxite. Extensive diamond diggings, worked equally by individual Africans and by European companies, began to be developed from 1919 onward. But the mainstay of the economy became cocoa, which local farmers began to produce on small plots in the forest toward the end of the 19th century. They found a reliable market for their produce. Cocoa became the most valuable export when it outranked gold in 1913, and thereafter went on to contribute more than four-fifths of exports and to constitute something between a third and a half of the world's supply.

The prosperity derived from cocoa in the 1920s enabled the governor, Sir Frederick Gordon Guggisberg, to pledge the country's revenues for loans to finance a coherent program of economic and social development. The Gold Coast's first deep-water port was built at Takoradi, the cocoa-producing forestlands were equipped with a comprehensive railway and road system, and the foundations were laid for educational and medical services as good as any in tropical Africa. Subsequent development was severely checked by the Great Depression of the 1930s and by events of World War II, but by the mid-1950s the postwar demand for tropical produce generated trade for the Gold Coast, estimated to have fewer than five million people, of about $500 million a year, not far short of that generated by all the 19 million people living in French West Africa. Government revenue reached the high level of $27.50 per person, by far the highest in western Africa, while the GNP of about $200 per person was probably higher than that of any tropical African country.

Nigeria provides yet another contrast. The people of its southern territories, like those of the southern Gold Coast or of Senegal, had long been in touch with the world economy. In 1912–13 the country's trade, at some $65 million a year, was 50 percent higher than the Gold Coast's and greater even than the combined total for the eight French colonies, including Senegal. But Nigeria was a giant territory, three times as large as the other three British colonies put together, and though compared with the French federation it was relatively small and compact (373,000 square miles), it had the same problem of extending over a considerable area of the remote western Sudan. This could not be ignored—as the much smaller northern Gold Coast or such northern French colonies as Niger were effectively ignored—because the Nigerian Sudan contained more than half the country's enormous population. By the mid-1950s the Nigerian population was more than 32 million, more than half that of western Africa.

Two things were clearly needed: first, to develop a transport system to make it possible to control and open up the populous north; and, second, to use some of the wealth generated from the growth of foreign trade in the south to stimulate development in the north. No coherent policy was possible, however, before the amalgamation of the separate colonial administrations, which was achieved under Lugard in 1912–14. Initially, even railway building tended to provoke disunion. The first line was built inland from Lagos in 1898–1901 to open up Yorubaland. Before this line was extended to the north across the Niger, the northern government had begun its own railway, from the highest point of navigation on the river, through its new administrative capital of Kaduna, to Kano. In 1912 this was intercepted by an extension of the Lagos line, and subsequently branches were built to areas active in tin mining and the cultivation of peanuts. Finally, another line was built from a new eastern port, Port Harcourt, to the coal mines around Enugu (1916), and this was subsequently extended to Kaduna (1927). By the 1930s Nigeria had 1,900 miles of railway, nearly as many as those possessed by all the French territories together (2,160 miles) but built at nearly twice the cost.

While southern Nigerian development, based essentially on cocoa production in the west and processing of palm oil and kernels in the east, followed much the same pattern as that of the southern Gold Coast, and with essentially similar social consequences, the development of peanuts as the prime export crop of the north did not produce comparable results for its appreciably larger population. By the mid-1950s the trade of Nigeria, at some $800 million a year, was still greater than that of all French West Africa in total, but it was appreciably less per capita, $25.30 compared with $32.20, and the annual revenue available to government, at $173 million, was small in proportion to the total population, only about $5.50 per capita. Inevitably a serious gap had developed between the economic and social progress of the south and that of the north.

Nigerian trade

DECOLONIZATION AND THE REGAINING OF INDEPENDENCE

The end of the colonial period and the establishment during 1957–76 of all the former colonies as independent states was attributable both to a change in European attitudes toward Africa and the possession of colonies and to an African reaction to colonial rule born of the economic and social changes it had produced.

Europeans had colonized western Africa in the later 19th and early 20th centuries confident that their civilization was immensely superior to anything Africa had produced or could produce. Yet hardly had their colonies been established than these convictions began to be challenged. World War I, and the immense misery and loss of life it caused, led some Europeans to doubt whether nations who could so brutally mismanage their own affairs had any moral right to dictate to other peoples. Some reflection of this view was seen in the League of Nations and the system of mandates applied to the former German colonies. Although in western Africa these were entrusted to either French or British administration, the mandated territories did not become the absolute possessions of the conquerors, and the role of the new rulers was declared to be to equip the mandated territories and their peoples for self-government.

A second shock to European self-confidence came with the Great Depression of the 1930s, when trade and production shrank and millions of Europeans had no work. It began to be argued that a remedy lay in more active development of the overseas territories controlled by Europe. If more European capital and skills were directed to the colonies, so that they could produce more raw materials for European industry more efficiently, both Europe and the colonies would gain; as the colonies became wealthier through the exploitation of their resources, the people of the colonies would buy more from Europe.

In 1929 Britain had enacted the first Colonial Development Act, providing that small amounts of British

First Colonial Development Act

government money could be used for colonial economic development, thus breaking the deadlock by which the only colonial governments that could embark on development programs to increase the wealth of their subjects, and to improve their own revenues, were those that already commanded sufficient revenue to pay for the programs or to service the loans the programs required. The idea that the colonies should be actively developed, in the European as much as in the African interest, was broadened during and after World War II. Transport and currency problems made it urgent for Britain and France to exploit strategic raw materials in their colonies. Furthermore, during 1940–44, when France itself was in German hands, it was only from the colonies and with their resources that General Charles de Gaulle and his associates could continue the fight.

The British funding policy, initiated in 1929, of providing the funds needed for colonial development was greatly expanded in the 1940s and extended to social as well as economic plans. After the war the governments of both Britain and France required their colonial administrations to draw up comprehensive development plans and in effect offered to provide the funds for those that could not be funded from local resources.

In view of past history, the need for such plans was probably greater in the French colonies than in the British, and the French West African program for 1946–55 envisaged the investment of $1,108,000,000, compared with programs totaling $549 million for the four British colonies. Virtually all of the financing for the French program came from France itself. But some of the British colonies had built up considerable reserves from the high prices commanded by their produce during the war and immediate postwar years, and they themselves were able to provide much of the money needed. This tended to accentuate already existing disparities. In the extreme case the Gold Coast plan envisaged spending $300 million, only 4 percent of which was British money. This was the same level of expenditure, roughly $60 per capita, as envisaged for French West Africa. Nigeria's program, with a contribution from Britain of 42 percent, proposed to spend $220 million—only about $7 per capita. The figures for Sierra Leone were $21 million, 45 percent from the United Kingdom, and $10 per capita; and for the tiny Gambia $8 million, 35 percent, and $27 per capita.

Gold Coast development

The accompanying political changes were more cautious and turned out to be inadequate to accommodate African aspirations—which had been derived from social changes occasioned during the classical period of colonial rule and further whetted by the policies of active economic development. On the British side, during 1945–48 the legislative councils were reformed so that African representatives outnumbered the European officials. Many of these African members, however, were still government nominees, and, because of the British attachment to indirect rule, those who were elected were mainly representative of the traditional chiefs.

Political advance for the French colonies was naturally seen in terms of increased African participation in French political life. In 1944 it was proposed that the colonies become overseas territories of France. Delegates from the colonies in fact participated in the making of the new postwar French constitution, but this was subject to referenda in which metropolitan French votes predominated. The constitution eventually adopted in 1946 was less liberal to Africans than they had been led to expect.

The emergence of African leaders. By the later 1940s, however, there were appreciable numbers of Africans in both the French and the British colonies who had emerged from traditional society through the new opportunities for economic advancement and education. In coastal areas Christian missionaries and their schools had advanced with the European administrations. The colonial governments, requiring African subordinates for their system, commonly aided and developed the elementary and vocational education initiated by the Christian missions and often themselves provided some sort of higher education for the chiefly classes whose cooperation they required. If rather little of this education had penetrated to the Sudan by the 1940s, in some coastal areas Africans had become eager to invest some of their increasing wealth in education, which was seen as the key to European strength.

Relatively few Africans started up the French educational ladder—school attendance by the mid-1950s was some 340,000, about 1.7 percent of the total population—but those who did found themselves in a system identical with that in France. In British West Africa schools had got a footing before there was much administration to control them, and their subsequent development was more independent. The British educational system therefore developed into a pyramid with a much broader base than the French one. By the mid-1950s there were more than two million schoolchildren in Nigeria, about 6 percent of the total population and a much higher proportion of the population of the south, in which the schools were concentrated; in the Gold Coast there were nearly 600,000, some 12 percent of the population. Many more people in the British than in the French territories thus got some education, and appreciably more were able to attend universities. In 1948 universities were established in the Gold Coast and Nigeria; by 1960 the former territory had about 4,500 university graduates and the latter more than 5,000. The first French African university was a federal institution at Dakar opened in 1950; by 1960 the total number of graduates in French West Africa was about 1,800.

Universities

By the 1940s there was enough education to make European-style political activity possible in all the coastal colonies. Such activity may be traced back to at least the 1890s, when Gold Coast professionals and some chiefs founded the Aborigines' Rights Protection Society (ARPS) to prevent the wholesale expropriation of African lands by European entrepreneurs or officials. The ARPS went on to campaign against the exclusion of qualified Africans from the colonial administration. Following this, in 1918–20, a National Congress of British West Africa was formed by professionals to press for the development of the legislative councils in all the British colonies into elective assemblies controlling the colonial administrations.

In French West Africa early political activity was concentrated in the four towns of Senegal whose people possessed political rights before 1946. Because the seat of power was very clearly in France, with Senegalese electors sending a deputy to the French National Assembly, the result by the 1930s was the emergence of a Senegalese Socialist party allied to the Socialists in France.

An African elite

By the late 1940s both the French and the British territories possessed an educated, politicized class, which felt frustrated in its legitimate expectations; it had made no appreciable progress in securing any real participation in the system of political control. In fact, anything approaching effective African participation seemed more remote than ever. Implementation of the development programs led to a noticeable increase in the number of Europeans employed by the colonial regimes and their associated economic enterprises. On the other hand, because many Africans had served with, and received educational and technical training with, the British and French armies, the war had led to a great widening of both African experience and skills. Furthermore, the postwar economic situation was one in which African farmers were receiving high prices for their produce but could find little to spend their money on, and in which the eagerly awaited development plans were slow to mature because European capital goods were in short supply.

The formation of African independence movements. There thus developed a general feeling among the intelligentsia that the colonies were being deliberately exploited by ever more firmly entrenched European political and economic systems and that there had developed a new, wider, and mobilizable public to appeal to for support. In 1946 politicians in French West Africa organized a federation-wide political association, the African Democratic Rally (RDA). The RDA and its members in the French National Assembly aligned themselves with the French Communist Party, the only effective opposition to the governments of the Fourth Republic. The result, during 1948–50, was the virtual suppression of the RDA in Africa by the colonial administrations.

In British West Africa the tensions were greatest in the Gold Coast. In 1947 the established politicians brought in Kwame Nkrumah, who had studied in the United States and Britain and had been active in the Pan-African movement, to organize a nationalist party with mass support. In 1948 European trading houses were boycotted, and some rioting took place in the larger towns. An official inquiry concluded that the underlying problem was political frustration and that African participation in government should be increased until the colony became self-governing. In 1951, therefore, a new constitution was introduced in which the legislative council gave way to an assembly dominated by African elected members, to which African ministers were responsible for the conduct of much government business. By this time Nkrumah had organized his own mass political party, able to win any general election, and during the following years he negotiated with the British a series of concessions that resulted in 1957 in the Gold Coast becoming the independent state of Ghana.

Kwame Nkrumah

Once the British had accepted the principle of cooperating with nationalist politicians, their other western African colonies began to follow the example set by the Gold Coast. But Nkrumah had been greatly aided by the high price for cocoa in the 1950s (which meant that by 1960 Ghana's trade was worth $630 million a year and that government revenue, at more than $280 million, was broadly adequate to give the people what they wanted in the way of modernizing programs) and by the comparatively high level and generally wide spread of education in a sizable yet compact territory that was without too serious ethnic divisions. The other colonies were not so well placed.

The small size of The Gambia was the principal factor contributing to the delay of its independence until 1965. Sierra Leone was a densely populated country that was appreciably poorer than Ghana (its GNP per capita, at about $70, being approximately one-third of Ghana's) and in which there was a wide disparity in levels of education and wealth between the Creoles—the descendants of liberated slaves who lived in and around Freetown—and the rest of the people. When independence was achieved in 1961, these deeply rooted problems had been papered over rather than solved.

Nigeria presented the greatest challenge to British and African policymakers alike. In the south two nationalist parties emerged, the Action Group (AG), supported primarily by the Yoruba of the west, and the National Convention of Nigerian Citizens (NCNC), whose prime support came from the Igbo of the east. These parties expected the whole country quickly to follow the Ghanaian pattern of constitutional change. But any elective central assembly was bound to be dominated by the north, which had some 57 percent of the population and whose economic and social development had lagged far behind. The north's political leaders—most of whom were conservative Muslim aristocrats closely allied with the British through indirect rule—were not at all eager to see their traditional paramountcy invaded by aggressive and better-educated leaders from the south.

AG and NCNC in Nigeria

The first political expedient was to convert Nigeria into a federation of three regions. In 1957 this allowed the east and the west to achieve internal self-government without waiting for the north, but it left open the questions of how politics were to be conducted at the centre and how Nigerian independence was to be secured. At this juncture it occurred to the northern leaders that by allying themselves to one of the southern parties they might maintain their local monopoly of power and gain prestige in the country as a whole by asking for its independence. The problem of central politics was thus resolved when the northern leaders entered a coalition federal government with the NCNC, and in 1960 Nigeria became independent.

Meanwhile, in French West Africa the RDA, led by Félix Houphouët-Boigny, broke with the Communist Party. The votes of a small bloc of African deputies in the French National Assembly were of considerable value to the shifting coalitions of non-Communist parties that made up the unstable French governments of the 1950s, and the RDA began to seek to influence these governments to allow greater freedom to the colonies.

By 1956 Houphouët-Boigny's policy had secured a widening of the colonial franchises and the beginnings of a system by which each colony was on the way to becoming a separate unit in which African ministers would be responsible for some of the conduct of government. The implications of this approach, however, did not meet with the approval of some other African leaders, most notable among them Léopold Sédar Senghor in Senegal and Ahmed Sékou Touré in Guinea. Senghor had stood outside the RDA since the days of its alliance with the Communists, which he had thought could only bring disaster. Together with Sékou, who had remained within the RDA, he argued that Houphouët's policy would split up the western African federation into units that would be too small and poor to resist continued French domination.

In 1958 the French Fourth Republic collapsed and de Gaulle was returned to power. On Sept. 28, 1958, in a referendum, the colonies were offered full internal self-government as fellow members with France of a French Community that would deal with supranational affairs. All of the colonies voted for this scheme except Guinea, where Sékou Touré led the people to vote for complete independence. Senegal and the French Sudan were then emboldened in 1959 to come together in a Federation of Mali and to ask for and to receive complete independence within the community. These two territories separated in the following year, but all the others now asked for independence before negotiating conditions for association with France, and by 1960 all the former French colonies were de jure independent states.

Intimations of imperial decline

By that time only the excessively conservative regimes of Portugal and Spain sought to maintain the colonial principle in western Africa. Encouraged and aided by independent neighbours, Guinean nationalists took up arms in 1962 and after 10 years of fighting expelled the Portuguese from three-quarters of Portuguese Guinea. In 1974 the strain of this war and of wars in Mozambique and Angola caused the Portuguese people and army to overthrow their dictatorship. Independence was quickly recognized for Guinea-Bissau in 1974 and for the Cape Verde Islands and São Tomé e Príncipe in 1975.

Spain concluded in 1968 that the best way to preserve its interests in Equatorial Africa was to grant independence to its people without preparing them for it. The result was chaos. Potential phosphate riches led Spain to hold on to the Spanish Sahara, however, until the death of Francisco Franco in 1975, and growing pressure from Morocco led to agreements by which, in the following year, despite an armed nationalist revolt sponsored by Algeria, the territory was partitioned between Morocco and Mauritania. In 1976, under Algerian sponsorship, the nationalists proclaimed the Sahrawi Arab Democratic Republic. Mauritania signed an agreement with the nationalists' armed wing, the Popular Front for the Liberation of Saguia el Hamra and Rio de Oro (Polisario) in 1979 and renounced all claims to its sector, but Morocco refused to yield its sector and engaged in an armed struggle with the Polisario over the territory. (J.D.F.)

WESTERN AFRICA SINCE INDEPENDENCE

Western Africa's political history since independence has been dominated by two factors. First, the hopes raised by political independence in many countries were not realized. National leaders clung to personal power and frustrated expectations of political evolution along democratic lines, with corruption, self-seeking, and a loss of political direction as a consequence. Dissatisfaction was commonly focused on the military, which in most states was the only institutionalized authority apart from the government. Since the 1960s western Africa has been characterized by a pattern of military coups. In some states successive coups have taken place as one military faction replaced another; in others, the army has intervened as periodically restored civilian regimes have failed.

Second, the relative economic position of western Africa has seriously declined since independence. Government mismanagement has played a part, but significant underlying factors have played a larger role. At independence few states had the economic resources to satisfy popular

Regional economic decline

expectation. Internationally approved, ambitious public- and private-sector industrial modernization programs were often ill-judged and harmful to the agricultural sector. A declining capacity to feed themselves led in many states to costly import substitution. At the same time, export crops were dependent on declining world market prices, and the price of imported oil rose spectacularly in the 1970s. Many states experienced chronic balance-of-trade deficits and incurred crippling international debts. In the 1970s and '80s drought and famine exacerbated the situation, mainly but not exclusively in the Sahel. A deepening indebtedness and difficult and prolonged negotiations with lending and rescheduling agencies such as the International Monetary Fund (IMF) and the World Bank resulted. The parlous economic situation is the chief political problem facing western African governments.

Francophone countries. In Senegal, Léopold Senghor, with strong French support, maintained power until his resignation in 1980 by balancing conflicting factions and promising controlled political liberalization. His nominated successor, Abdou Diouf, continued these general policies together with the link to France. In the 1980s Senegal experienced budget deficits and agreed to a program of fiscal restraint with the IMF and the World Bank.

Côte d'Ivoire has been ruled since independence by the strongly pro-French Félix Houphouët-Boigny, who was reelected to a sixth term in office in 1985. A period of sustained economic growth, marked by significant foreign investment, ended in the 1980s. Côte d'Ivoire is sensitive to international commodity prices, and in 1987 it too agreed to a short-term policy of fiscal restraint with the IMF.

In Guinea, Ahmed Sékou Touré held power from independence in 1958 to his death in 1984. Initially he repudiated any connection with France and the Western powers. Guinea adopted a variant of international Marxism, experienced a number of internal political crises, and embraced economic policies that ran into trouble in the 1970s. Following Sékou Touré's death, the army seized power. President Lansana Conté survived a coup attempt in 1985 and, with a ruined economy, accepted a stringent IMF and World Bank retrenchment program. French investment returned, but continuing economic hardship led to ongoing unrest within the army.

In Cameroon, Ahmadou Ahidjo successfully created a unified state from French and British trusteeship territories, but his government became increasingly authoritarian. In 1979 there was a regional uprising, an attempted coup, and an upsurge of covert opposition from Anglophone elements. In 1982 Ahidjo resigned the presidency in favour of the prime minister, Paul Biya. In 1984 another coup attempt by the army was suppressed with considerable violence. Biya's reassertion of presidential authority thereafter was greatly assisted by access to oil revenues. Cameroon has experienced significant if sometimes fragile economic growth and has good economic prospects.

Chad became independent in 1960. In 1975 the historic division between the Muslim north and non-Muslim south had erupted in violence. Prolonged warfare followed, in which France and Libya intervened. Since 1982 President Hissen Habré has led an unstable country with great economic problems. In Mali, Modibo Keita was ousted by the army in 1968. The successor regime of General Moussa Traoré restored constitutional rule in 1979. This was widely seen as only a pretense at democracy, and there ensued widespread unrest and plots that were countered by repression. The 1980s were dominated by chronic deficits in foreign trade and an economic crisis that was deepened by severe droughts.

In Niger, Hamani Diori was removed by the army in 1974 owing to the economic hardships that followed severe drought. His successor, Seyni Kountché, used uranium revenues to consolidate his rule, but by the time Kountché died in 1987, the uranium boom had ended. His cousin and successor, Colonel (later Brigadier General) Ali Saibou, inherited an economy that was distressed by debt.

Upper Volta experienced a series of coups as successive civilian and military regimes unsuccessfully grappled with a disabled economy. In 1983 Captain Thomas Sankara assumed power. He changed the name of the country to Burkina Faso in 1984. Sankara was a charismatic figure who undertook socialist agricultural initiatives. These foundered, but Sankara refused to approach the IMF. In 1987 he was assassinated and replaced by Captain Blaise Compaoré. State capitalism replaced Sankara's socialism, and overtures were made to the IMF in an attempt to solve the chronic trade deficit.

Socialism in Burkina Faso and Benin

Dahomey, which changed its name to Benin in 1975, suffered five military coups until, in 1972, Major Mathieu Kérékou seized power. Kérékou's Marxist sympathies aroused widespread apathy and disillusionment. In the 1980s socialism was cautiously abandoned, and Benin was opened to foreign investment. In 1984 public-sector cuts were instituted under IMF oversight, but retrenchment alienated Benin's large educated elite and those with socialist leanings. In 1988, amid rumours of debt restructuring, there was an attempted coup.

In Mauritania, the government of Moktar Ould Daddah was unable to cope with drought, poverty, and military confrontation with Algerian-supported Polisario Front nationalist guerrillas in Western Sahara (formerly Spanish Sahara). In 1978 it was replaced by the military, which concluded peace with the Polisario Front but became embroiled in hostilities with Morocco. In 1984 Colonel Maaouya Ould Sidi Ahmed Taya replaced Colonel Mohamed Khouna Ould Haidalla in a bloodless coup and restored diplomatic relations with Morocco. Throughout the 1980s the government was plagued by internal ethnic conflicts between black Africans and Muslims of Arab extraction. Despite valuable mineral and fish resources and aid from Saudi Arabia, Taya was forced to accept a debt rescheduling program from the IMF in 1985.

In Togo, the army intervened twice in the 1960s, and since 1967 the country has been ruled by General Gnassingbe Eyadema. Togo benefited from phosphate revenues until this boom collapsed in the mid-1970s. From 1979 Togo periodically resorted to the IMF for help with economic stabilization and debt rescheduling. General Eyadema's tentative democratization was offset by a semiofficial personality cult that grew up around him.

Anglophone countries. Nigeria, among the former British colonies, is the demographic giant of western Africa. It has a wealth of resources, but its federal structure has been threatened by regional and ethnic rivalries. In 1964 political arrangements broke down under these strains, and in 1966 the army intervened in an attempt to create a unitary government. Regional rivalries deepened, and in 1967 Igbo officers from eastern Nigeria declared a secessionist republic of Biafra. After three years of warfare the federal government of General Yakubu Gowon liquidated Biafran independence. Gowon tried to prevent future secession by enlarging the number of regional states to 12 (increased to 19 in 1976), but his government lost support because of its perceived reluctance to surrender power. Gowon was toppled in 1975 by a coup led by Brigadier Murtala Ramat Mohammed, who was himself assassinated in 1976. The army continued in power under General Olusegun Obasanjo, who instituted measures to restore Nigeria to civilian rule. In 1979 multiparty elections led to the formation of the Second Republic under President Alhaji Shehu Shagari, who in mid-1983 won a second term. By then, however, the revenues generated by Nigeria's oil boom in the 1970s were drying up. There was a widespread belief that Nigeria's oil wealth had been corruptly squandered, and at the end of 1983 the Second Republic was replaced by a military regime under Major General Mohammed Buhari. In 1985 Buhari was himself replaced in a military coup by General Ibrahim Babangida, who promised to return Nigeria to civilian rule by 1992. Meanwhile, he had to grapple with Nigeria's deepening economic malaise, for which IMF and World Bank assistance was sought after protracted and difficult negotiations.

In 1957 Ghana achieved independence under the charismatic leadership of Kwame Nkrumah. His prominence in the international arena was offset by corruption and an increasingly autocratic political machine. In 1966 Nkrumah was overthrown by the army. In 1969, under military

tutelage, the Second Republic was instituted under Kofi Busia. In 1972 the army again intervened, but the successive regimes of Generals Ignatius Kutu Acheampong and Frederick W.K. Akuffo failed to resolve Ghana's deepening economic crisis. The military reluctantly promised a return to civilian rule in 1979. In June of that year, however, a group of junior officers under Flight Lieutenant J.J. Rawlings carried out a coup, purged senior figures in the army, and held elections. But the Third Republic, headed by Hilla Limann, proved incapable of reversing economic catastrophe, and in 1981 Rawlings led another coup. In the 1980s the government headed by Rawlings pursued broadly socialist programs of democratization. At the same time it tackled Ghana's long-term economic decline and concluded a wide-ranging structural adjustment program with the IMF and the World Bank.

In Sierra Leone, there were military coups in 1967 and 1968, and then the army installed Siaka Stevens at the head of a civilian government. In 1985 Stevens retired and handed over power to General Joseph Saidu Momoh. Momoh's regime was marked by economic difficulties. Relations with the IMF fluctuated, chiefly because IMF-recommended austerity measures produced widespread unrest in Sierra Leone and, in 1987, an alleged plot against Momoh's government.

In Anglophone western Africa parliamentary government has survived only in the tiny Gambia. In 1980 and 1981, however, the presidency of Sir Dawda Jawara was threatened by unrest. In 1981 neighbouring Senegal intervened militarily in support of Jawara. This led in 1982 to the creation of the confederation of Senegambia, in which Senegal and The Gambia retained their sovereign status but agreed to military integration and economic and political cooperation. Steady but slow progress has been made under this arrangement. In 1985 Jawara sought IMF advice on public service cuts and external debt restructuring. In 1988 there was an obscure coup attempt, during which it became clear that Jawara still relied on Senegal for the maintenance of order.

Creation of Senegambia

Lusophone countries. The Lusophone states of western Africa have also experienced a number of problems. In Guinea-Bissau in 1980 President Luis Cabral was replaced in a coup by Prime Minister João Bernardo Vieira. In 1987 Vieira agreed to a stringent World Bank structural adjustment package, by which Guinea-Bissau's currency was devalued by 60 percent.

Cape Verde has remained politically stable. It is a one-party state, but President Aristides Pereira was a pragmatist who has been reelected to office since 1975. The issue of unity with Guinea-Bissau remains difficult, but the two states developed increasingly close ties in the 1980s.

Liberia. Liberia, the only western African state not formally colonized by a European power, has also been beset by political and economic problems. During President William V.S. Tubman's long tenure in office (1944–71), the ruling Afro-American True Whig Party concluded that its maintenance of power depended on an economic and political partnership with the indigenous African Liberians. Tubman's successor, President William R. Tolbert (1971–80), sought ways to implement this policy. In 1979, however, long-term inequalities, rural poverty, and economic mismanagement produced serious food riots. In 1980 government repression precipitated a military coup led by Master Sergeant (later Commander in Chief) Samuel K. Doe. Doe executed former government ministers and promised a return to civilian rule in 1985. Doe survived a coup attempt in 1981, but his clear intention of holding onto power in any successor government made him the target of an assassination attempt in 1985. In the same year he duly won the presidential election, but another army coup attempt had to be bloodily crushed. Doe was unable to resolve the country's economic problems. In 1986 the IMF and World Bank withdrew from Liberia, and in 1987 Doe invited American financial experts to manage Liberia's distressed economy. (T.C.McC.)

For coverage of related topics in the *Macropædia* and *Micropædia,* see the *Propædia,* sections 941, 96/11, and 978, and the *Index.*

COUNTRIES OF THE WESTERN SUDAN

Burkina Faso

Burkina Faso is a landlocked state in western Africa with an area of 105,869 square miles (274,200 square kilometres). A former French colony, it gained independence as Upper Volta (République de Haute-Volta) in 1960; the name Burkina Faso was adopted in 1984. The country is bounded to the north and west by Mali, to the south by Côte d'Ivoire, Ghana, and Togo, and to the east by Benin and Niger. The capital, Ouagadougou, is about 500 miles (800 kilometres) by road from the sea.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Burkina Faso consists of an extensive plateau, which is slightly inclined toward the south. The lateritic (red, leached, iron-bearing) layer of rock that covers the underlying crystalline rocks is deeply incised by the country's three principal rivers—the Black Volta, Red Volta, and White Volta—all of which converge in Ghana to the south to form the Volta River. The Oti, another tributary of the Volta, rises in southeastern Burkina Faso. In the southwest there are sandstone plateaus bordered by the Banfora Escarpment, which is about 500 feet (150 metres) high and faces southeast. The country is generally dry and the soil infertile. Great seasonal variation occurs in the flow of the rivers, and some become dry beds in the dry season.

The river system

Climate. The climate is generally sunny, hot, and dry. In the north the climate is semiarid steppe, known locally as the Sahelian type and characterized by three to five months of rainfall, which is often erratic. To the south it becomes increasingly of the tropical wet-dry type sometimes called Sudanic, characterized by greater variability of temperature and rainfall and greater total rainfall.

Four seasons may be distinguished in Burkina Faso: a dry and cool season from mid-November to mid-February, with temperatures dropping to about 60° F (16° C) at night; a hot season from mid-February to June, when maximum temperatures rise to about 104° F (40° C) in the shade and the harmattan—a hot, dry, dust-laden wind blowing off the Sahara—is prevalent; a rainy season, which lasts from June to September; and an intermediate season, which lasts from September until mid-November. Annual rainfall varies from about 40 inches (1,000 millimetres) in the south to less than 10 inches in the north.

Plant and animal life. The northern part of the country consists of savanna, with prickly shrubs and stunted trees that come to life during the rainy season. In the south the prickly shrubs give way to scattered forests, which become more dense along the banks of the perennial rivers. While tree growth in the north is discouraged by the climate, farmers in the south often permit only useful trees, such as the karite (shea tree) or the baobab, to survive.

Animal life in the eastern region includes buffalo, antelope, lions, hippopotamuses, elephants, and crocodiles. Elephants, buffalo, and antelope are also found in the southeast and on the banks of the Black Volta, while herds of hippopotamuses are to be seen some 40 miles from the city of Bobo Dioulasso. Animal life also includes monkeys. Bird and insect life is rich and varied, and there are many fish in the rivers.

Settlement patterns. The population as a whole is unevenly distributed among the different regions. The Mossi country is densely settled. Situated in the eastern and central regions, it contains about two-thirds of the total population. In the remaining regions the population is scattered.

About 90 percent of the population is rural—the highest percentage in western Africa—and lives in some 7,700 villages. Villages tend to be grouped toward the centre

of the country at higher elevations away from the Volta valleys. For several miles on either side of the Volta rivers the land is mostly uninhabited because of the prevalence of the deadly tsetse fly, which carries sleeping sickness, and the simulium fly, which carries onchocerciasis, or river blindness.

Ouagadougou, the administrative capital and the seat of government, is a modern town, in which several commercial companies have their headquarters. It is also the residence of the *morho naba,* emperor of the Mossi, and an important regional centre for international aid programs.

The principal towns

Apart from Ouagadougou, the principal towns are Bobo Dioulasso, Koudougou, Ouahigouya, Kaya, Fada Ngourma, and Banfora. Bobo Dioulasso, in the west, was the economic and business capital of the country when it formed the terminus of the railroad running to Abidjan, Côte d'Ivoire, on the coast; since 1955, however, when the railroad was extended to Ouagadougou, it has lost some of its former importance, although it remains a commercial centre.

The people. Two principal ethnic groups live in Burkina Faso. The first of these is the Voltaic group, which may be further divided into four subgroups—the Mossi, which include the Gurma and the Yarse, the Gurunsi, the Bobo, and the Lobi. The second group is the Mande family, which is divided into five subgroups: the Samo, the Marka, the Busansi, the Senufo, and the Dyula. In addition, there are Hausa traders, Fulani herders, and the Tuareg, or rather their settled servants, the Bella.

Each of the ethnic groups found in Burkina Faso has its own language, although Moré, the language of the Mossi, is spoken by a great majority of the population, and Dyula and Hausa are widely used in commerce. French, the official language, is used for all communication with other countries. About one-half of the population are animists, attaching great importance to ancestor worship. Islām exerts an increasing influence upon customs, and Muslims account for approximately two-fifths of the population. The seat of the Roman Catholic archbishopric is Ouagadougou, and there are eight bishoprics. There are few Protestants in the country.

In the late 20th century, yearly population growth averaged more than 2 percent; nearly one-half of the population is below the age of 15. Average life expectancy is 47 years for women and 44 years for men.

The economy. Most of the population is engaged in subsistence agriculture or stock raising. Difficult economic conditions, made worse by severe intermittent droughts, have provoked considerable migration from rural to urban areas within Burkina Faso and to neighbouring countries such as Côte d'Ivoire and Ghana. As many as 1.5 million people, or almost one-third of the country's labour force, are abroad at any given time.

The development of industry in Burkina Faso is hampered by the small size of the market economy and by the absence of a direct outlet to the sea.

Resources. Minerals, especially manganese and gold, represent potential wealth for this otherwise poorly endowed nation. Gold mines at Poura, southwest of Koudougou, were reopened in late 1984, and smaller gold deposits near Sebba and Dori-Yalogo in the north are known to exist. Reserves of nickel, bauxite, zinc, lead, and silver are being studied. The country's substantial manganese deposits at Tambao in the northeast potentially represent Burkina's most important resource and one of the world's richest sources of this mineral. Exploitation is limited by existing transport inadequacies.

Agriculture. Agricultural production consists of subsistence foodstuffs, with the surplus being sold as cash crops. Surplus cotton, shea nuts, sesame, and sugarcane are exported, while sorghum, millet, corn (maize), peanuts (groundnuts), and rice are grown for local consumption. Fonio (a crabgrass with seeds that are used as cereal), cassava, sweet potatoes, and beans are also grown. Stock raising, one of the principal sources of revenue, includes cattle, sheep, goats, pigs, donkeys, horses, and camels. Chickens, ducks, and guinea fowl are also raised.

Industry. Industry is limited to a number of plants, mainly in the cities and larger towns, that produce processed rice, beer, soft drinks, and flour, manufacture textiles and shoes, and assemble bicycles.

Finance. Burkina Faso, along with six other French-speaking states in western Africa, is a member of the West African Monetary Union. These states share a common central bank, with headquarters in Dakar, Senegal, and a common currency, the CFA (Communauté Financière Africaine) franc. Branches of the central bank in Burkina Faso are located in Ouagadougou and Bobo Dioulasso. Among the partially or wholly state-owned commercial banks, the most important is the Banque Internationale du Burkina in Ouagadougou. Burkina Faso is also a member of the Economic Community of West African States (ECOWAS), a body encompassing most states in western Africa, which attempts to integrate and harmonize the economic interests of the region. One of the poorest countries in the world, Burkina Faso relies heavily on international aid and on remittances from migrants to help offset its current account deficit.

Trade. External commerce, both in imports and in exports, is primarily with the Franc Zone and with neighbouring African countries in particular. Many cattle are exported to Côte d'Ivoire and to Ghana. There is a deficit in the balance of payments, largely due to the relatively small amounts of exports, which are not of sufficient value to equal the value of imported materials required for promoting further development.

Transportation. In addition to the rail line that links Ouagadougou to the port of Abidjan in Côte d'Ivoire, the capital is also linked by road to the principal administrative centres in the country and to the capitals of neighbouring countries. The railroad to Abidjan is 712 miles long, of which 321 miles run through Burkina Faso. Running from east to west before crossing the border, the line serves the towns of Koudougou, Bobo Dioulasso, and Banfora.

Burkina Faso has one of the most poorly developed road networks in proportion to its size among the western African states. Only about a quarter of the network is usable year-round. The remainder consists mostly of unpaved rural roads. Three road-building projects completed in the late 1960s and early '70s were financed by the European Development Fund. The first of these roads runs from Bobo Dioulasso to Faramana to the Mali frontier. The second runs from Ouagadougou to Pô to the Ghanaian frontier. The third runs from Ouagadougou to Koupéla. Additional internationally aided road maintenance and improvement programs, particularly in the country's northeast, were carried out in the 1980s.

International airports are located at Ouagadougou and Bobo Dioulasso. Internal air service, linking about 50 smaller airstrips, is supplied by the national airline.

Administration and social conditions. *Government.* A constitution, adopted by referendum in 1977, allowed for multiparty elections and a parliamentary democracy with a president as head of state. The military, which has had great influence in political life, staged a coup in 1980 and suspended the constitution and all political activity. Additional military coups were carried out in 1982, 1983, and 1987.

Local government

Burkina Faso is divided into 30 *provinces,* which, in turn, are divided into 250 *départements.* In addition, there are *communes* administered by mayors and elected municipal councillors.

Education. School enrollment is one of the lowest in Africa, even though the government devotes a large portion of the national budget to education. French is the language of instruction in primary and secondary schools. Higher education is sought at Ouagadougou University (established 1974). Other institutes in Ouagadougou sponsored by neighbouring francophone states offer degrees in rural engineering and hydrology. Some students seek higher education in France; in Dakar, Senegal; or in Abidjan, Côte d'Ivoire.

Health and welfare. The state of health of the Burkinabé is very poor. Periodic droughts have contributed to malnutrition and related diseases, especially among young children and pregnant women. Only about one-third of the people have access to safe drinking water.

Gastrointestinal diseases and malaria are the main causes of death. Onchocerciasis, sleeping sickness, leprosy, yellow fever, and bilharzia are also endemic. The government carried out a successful child-immunization program in 1985, as a result of which infant and child mortality have decreased, though adult mortality has not. Hospitals exist only in the leading towns, but the government improved primary health care by raising the number of village clinics to 7,000.

Cultural life. Folklore is rich, reflecting the country's ethnic diversity. On national occasions each region is represented in the capital by its own folkloric group.

Ouagadougou draws large numbers of visitors to the biennial Pan-African Film Festival (FESPACO). The International Crafts Fair, which is held in alternate years, celebrates the rich and diverse craft production of the nation's artisans. Several daily newspapers are published, including the government-sponsored *Sidwaya,* as well as a number of weeklies. There are three national parks—those of Po, Arly, and in the east, straddling the border with Benin and Niger, the great "W" National Park.

For statistical data on the land and people of Burkina Faso, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (P.H.Gu./M.Ec.)

HISTORY

Early history. Axes belonging to a Neolithic culture have been found in the north of Burkina Faso. The Bobo, Lobi, and Gurunsi are the earliest known inhabitants of the country. In about the 15th century AD conquering horsemen invaded the region from the south to found the Gurma and the Mossi kingdoms, in the eastern and central areas, respectively. Several Mossi kingdoms developed, the most powerful of which was that of Ouagadougou, in the centre of the country. Headed by an emperor titled the *morho naba* ("great lord"), the Ouagadougou Mossi state defeated attempted invasions by Muslim Songhai and Fulani neighbours yet maintained valuable commercial links with major western African trading powers such as the Dyula, the Hausa, and the Asante (Ashanti).

Arrival of European explorers

European exploration and colonization. The German explorer Gottlob Adolf Krause traversed the Mossi country in 1886; and the French army officer Louis-Gustave Binger visited the *morho naba* in 1888. France obtained a protectorate over Yatenga in 1895; and Paul Voulet and Charles-Paul-Louis Chanoine defeated the *morho naba* Boukari-Koutou (Wobogo) in 1896 and then proceeded to overrun the Gurunsi lands. The Gurma accepted a French protectorate in 1897; and in 1897 likewise the lands of the Bobo and of the Lobi were annexed by the French (though the Lobi, armed with poisoned arrows, were not effectively subdued until 1903). An Anglo-French convention of 1898 fixed the frontier between France's new acquisitions and the northern territories of the Gold Coast.

The French divided the country into administrative *cercles* ("circles") but maintained the chiefs, including the *morho naba,* in their traditional seats. At first attached to French Sudan (or Upper Senegal-Niger, as that colony was called from 1904 to 1920), the country was organized as a separate colony, Upper Volta (Haute-Volta), in 1919. In 1932 it was partitioned between Côte d'Ivoire, Niger, and French Sudan. In 1947, however, Upper Volta was reestablished to become an overseas territory of the French Union, with a territorial assembly of its own. The assembly in 1957 received the right to elect an executive council of government for the territory, which at the end of 1958 was transformed into an autonomous republic within the French Community. When independence was proclaimed on Aug. 5, 1960, the new constitution provided for an executive president elected by universal adult suffrage for a five-year term and an elected Legislative Assembly. (H.J.D./Je.D./M.Ec.)

Independence. Since Burkina Faso became an independent nation, the military has on several occasions intervened during times of crisis. In 1966 the military, led by Lieutenant-Colonel (later General) Sangoulé Lamizana, ousted the elected government of Maurice Yaméogo. General Lamizana dominated the nation's politics until November 1980, when a series of strikes launched by workers, teachers, and civil servants led to another coup, this time headed by Colonel Saye Zerbo.

Colonel Zerbo's short-lived rule ended in November 1982, when noncommissioned army officers rebelled and installed Major Jean-Baptiste Ouedraogo as president. The Ouedraogo government soon split into conservative and radical factions, with the radicals seizing power on Aug. 4, 1983. They set up a National Revolutionary Council (CNR) with Captain Thomas Sankara as head of state.

The Sankara regime

A year after taking power, Sankara renamed the country Burkina Faso, meaning "Land of Incorruptible People," and ordered all officials, including himself, to open their bank accounts to public scrutiny. His government lasted until 1987 and could take credit for several concrete achievements: vaccination and housing projects, tree planting to hold back the Sahel, promotion of women's rights, and curbing of waste in government.

During Sankara's rule, tensions with Mali over the mineral-rich Agacher Strip erupted in a brief border war in December 1985. The dispute was settled in the International Court at The Hague a year later, apparently to the satisfaction of both states.

Initially a coalition of radical groups that included army officers, trade unionists, and members of small opposition groups, the Sankara regime gradually lost most of its popular support as power became concentrated in the hands of a few military officers—the most important of which were Sankara, Captain Blaise Compaoré, Major Jean-Baptiste Boukari Lingani, and Captain Henri Zongo. Urban dwellers objected to lower standards of living, while sections of the peasantry expressed unhappiness with ill-devised development schemes. The left accused the government of softening its anti-imperialist stance in order to secure foreign aid. Indeed, in spite of initial tensions between the United States and Burkina as a result of close links between Sankara and Colonel Muammar al-Qaddafi of Libya, by the end of Sankara's rule the United States had replaced France as Burkina's main aid donor. As support continued to decline, on Oct. 15, 1987, a military coup overthrew Sankara, killing him and eight others.

Compaoré took power at the head of a triumvirate that also included Captain Zongo and Major Lingani. Calling themselves the Popular Front (FP) and promising to adhere to the original principles of the Sankara revolution, the FP has stressed self-sufficiency in food and less strident relations with conservative states in the region, Togo and Côte d'Ivoire in particular. On the other hand, Libyan relations with Burkina remain friendly since the Compaoré government has taken power.

For later developments in the history of Burkina Faso, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (M.Ec.)

Cape Verde

Cape Verde (República de Cabo Verde) is a republic comprising a group of islands that lie 385 miles (620 kilometres) off the west coast of Africa, between 14°30′ and 17°30′ N and between 22°30′ and 25°30′ W, with a total land area of 1,557 square miles (4,033 square kilometres). Praia on São Tiago is the capital.

PHYSICAL AND HUMAN GEOGRAPHY

Cape Verde is named after the westernmost cape of Africa, which is the nearest point on the continent. The country consists of 10 islands and five islets, which are divided into the Windward (Barlavento) and Leeward (Sotavento) groups. The Windward Islands consist of Santo Antão, São Vicente, Santa Luzia, São Nicolau, Boa Vista, and Sal, together with the islets of Raso and Branco. The Leeward Islands include Maio, São Tiago (Santiago), Fogo, and Brava and the three islets called the Rombos—Grande, Luís Carneiro, and Cima.

The islands of the archipelago

The largest port in the islands is Porto Grande at Mindelo, on São Vicente. Its deepwater harbour accommodates sizable vessels and is used primarily as a fueling station.

The land. *Relief, drainage, and soils.* The islands are mountainous and are volcanic in origin. Only the three oldest—Boa Vista, Maio, and Sal, the so-called Rasas

("Flat") islands—have suffered enough erosion to have much level ground. Fogo (meaning "Fire") has an active volcano, Mount Cano, whose last major eruption was in 1951. Its cone rises 9,281 feet (2,829 metres) above sea level. The peak of Mount Coroa on Santo Antão is 6,493 feet. São Tiago and São Nicolau both have mountains more than 4,200 feet high. All the islands, especially the Windwards, have been eroded by sand carried by high winds, so that the outline of the landscape appears jagged.

There are few watercourses that run all year, and even these do not reach their ends during the dry season. Dry watercourses fill up for several days during the short, intense rainy season. Rains tend to occur as torrential storms, causing severe soil erosion and great damage to agriculture. Groundwater is the primary source of domestic water supply. Some groundwater sources, however, are sulfurous; others, mainly on São Vicente and Boa Vista, are slightly salty due to the low water tables.

Almost a quarter of the land area is rock of volcanic origin; basalt is a common type. More than 60 percent of the land is arid and lacking in humus and thus is suitable only for rough grazing. Sand and limestone outcrops are common in these areas. The remaining 15 percent is fertile; it contains alluvial deposits and is suitable for irrigation.

Soil erosion has been one of Cape Verde's greatest problems. It began to have a serious effect in the early 19th century, attributable to overgrazing by goats. Since independence a nationwide campaign to prevent erosion has been under way, involving planting drought-resistant varieties of acacia trees (which now cover 7.5 percent of the land surface), building small dikes, and improving farming techniques.

Climate. Moderate, stable temperatures and extreme aridity characterize the climate. February is the coolest month, with an average temperature of 71° F (22° C), and September is the warmest, having an average of 80° F (27° C). The islands are almost constantly under the influence of a dry northeast wind. Consequently there is almost no rainfall, except for a period from August through October, when an average of 1.6 inches (39 millimetres) a month is recorded. These rains can fail for years at a time.

Plant and animal life. On islands higher than 1,000 feet, which includes most of the larger islands, elevations are great enough to generate rainfall on the windward slopes. Grasses and some pine plantations are found in these relatively moist locations. The leeward slopes, however, exhibit a characteristic rain shadow effect that produces desert conditions, and the sparse shrub cover almost disappears. The shrubs remaining in these areas are mostly thorny or bitter; some are toxic. A sea mist on the higher hills permits some agriculture, and irrigated valley bottoms are densely cultivated. Salt areas on Maio and Sal have interesting xerophilous plants.

The scarcity of water limits the number of land turtles in the archipelago, but two species of sea turtles lay their eggs on the sandy shores of the uninhabited islets. There are many geckos, lizards, and several species of skinks. A species of giant skink is protected by law, but it may be extinct. There are 19 known species of butterflies, but none is endemic, and all the species are of African origin.

Birds

There are 105 known species of birds, of which only 38 breed regularly, including four species of petrels and two of shearwaters. Other bird species include the greater flamingo, the frigate bird and the buzzard (both nearly exterminated), the Egyptian vulture, the Cape Verde Islands kite, and the red-billed tropic bird. Several other birds are represented by local species, of which the kingfisher is among the most conspicuous. The only truly endemic species, however, are the cane warbler and the Raso lark, which is restricted to Raso, one of the smallest uninhabited islets. The rest of the birds are overseas migrants. Remarkably, gulls and terns do not breed on the islands.

Mammals of Cape Verde include the feral goats found on Fogo, the descendants of domestic goats that were brought to the islands. The islands' rodent population probably originated with rodents brought on early ships. The long-eared bat is the only indigenous mammal.

Settlement patterns. The majority of the population is rural, living in small villages in fertile valleys or on the coast in fishing communities. Proximity to a water supply is the determining feature in settlement location. There are three urban centres: Praia, Mindelo, and the island of Sal (centred on the international airport at Espargos). Their continuing growth and development are based on good access to transportation networks.

The people. The overwhelming majority of the islands' population is Creole (mulatto), the descendants of early contacts between Portuguese settlers and Africans brought as slaves to work on the plantations in the 16th century. Among the latter, Fulani (Fula) and Mandingo people from the region of Senegal, The Gambia, and Guinea-Bissau predominated.

Although Portuguese is the official language and is used in formal situations and for most written material, Crioulo is the mother tongue of most people. It is one of the oldest of the Portuguese creole languages. Different dialects of Crioulo exist on the different islands.

Most of the population is Roman Catholic, but a flourishing Protestant mission is based in Praia with a publishing venture in Fogo. Particularly in Sotavento, the celebration of Roman Catholic saints' days with drumming, dancing, and singing reveals the underlying African culture.

The population growth rate of more than 2 percent per year is high by world standards but is relatively low for western Africa. A steady emigration of young males seeking employment abroad and one of the lowest birth rates in sub-Saharan Africa have been responsible for dampening Cape Verde's population growth. Life expectancy, at 60 years for males and 64 years for females, is the highest in western Africa.

About half of the people live abroad as emigrant workers, often returning upon retirement. This trend began in the early 19th century with the arrival of American whaling ships that offered the opportunity for work and travel to the United States. Cape Verdeans still work on foreign merchant ships, and Boston, Mass., has a large immigrant community. During Portuguese rule, Cape Verdeans worked throughout the Portuguese empire as government officials and labourers. More recently, Portugal, The Netherlands, and France have become important destinations.

The economy. The economy is mixed, with approximately half the national production generated by state-owned concerns. Agriculture absorbs most of the labour force, although small-scale industry and services together generate a larger share of the gross national product (GNP). The mainstays of the economy are the revenue from Amílcar Cabral International Airport on Sal, foreign aid, and emigrants' remittances. These enable the balance of payments to stay generally positive despite imports far exceeding exports. As a nonaligned nation, Cape Verde receives aid from a variety of sources. There is one state bank, the Banco de Cabo Verde. The currency is the Cape Verdean escudo.

Resources. Cape Verde has few natural resources. The lack of fresh water is a problem. On São Vicente and Sal it is provided by desalination plants (which also generate electricity). Water is also obtained from springs, wells, and rainwater stored in cisterns. The country relies on imported oil for fuel. Fish are the major natural resource. Salt from Boa Vista, Maio, and Sal (meaning "Salt") was once an important export. Pozzolana, a volcanic rock that is used in making cement, is exported.

Agriculture. Bananas and coffee have been the chief agricultural exports. Crops grown for local consumption include corn (maize), sugarcane, castor beans, broad beans, potatoes, and peanuts (groundnuts). Severe and recurrent droughts affect the islands, causing unemployment and a dramatic fall in crop output. Deaths from starvation are a thing of the past, but there is a heavy reliance on imported foodstuffs. During the decade-long drought of the 1970s, 95 percent of food needs were met by imports.

Industry. Small-scale industries such as textiles and pharmaceuticals play an increasingly important role in the economy. Fish processing is becoming well established. In 1981 a cold-storage plant was opened in Mindelo, and canning facilities exist. Tuna, shark, and lobster are important exports.

Transportation. There are international air services from Sal to Lisbon, Boston, Moscow, Senegal, Guinea-Bissau, and Brazil. Within the islands, regular ferries and planes provide local service. All the islands except Brava have airports. With the exception of São Tiago the road networks are limited. There is a small national shipping line and a national airline, Transportes Aéreos de Cabo Verde (TACV). Ship repairing is carried out in Mindelo, where dry docks accommodate vessels of up to 2,800 tons deadweight.

Administration and social conditions. *Government.* According to the 1980 constitution (amended in 1981), the head of state, the president, is elected to a five-year term by the National People's Assembly. He is assisted by an appointed Cabinet whose ministers are drawn from among nominees of the Assembly. The Assembly has 83 deputies elected by universal adult suffrage to five-year terms. The African Party for the Independence of Cape Verde (PAICV) is the only legal political party. In 1980 local assemblies headed by elected executive committees were formed. At the middle level there are island committees, also elected by universal adult suffrage. The Supreme Tribunal of Justice, located at Praia, is the highest court and oversees a network of popular tribunals at the local level.

Education. According to official policy, education is compulsory at the primary level for children aged seven to 14 years, but in the mid-1980s only two-thirds of all school-age children were enrolled in either primary or secondary schools. Secondary schooling is provided by four *liceus* and one industrial and commercial school. There are three teacher-training schools as well. The islands have no university. Only about half the adult population is literate, but the government has launched a major effort to build schools and train teachers to combat this problem.

Health and welfare. The major health problem is diarrhea in infants, caused by poor hygiene. There are isolated cases of malaria on São Tiago and leprosy on Fogo. The Ministry of Health is responsible for a network of locally run hospitals and health centres. A mother and child protection program has operated since 1977 with preventive medicine as a priority. Successful vaccination campaigns have reached 90 percent of the children. The infant mortality rate, while still high by world standards, is significantly lower than in other western African countries. Social welfare depends on the extended family.

Cultural life. Portuguese customs and culture have influenced the islands, but they are blended with African traditions as well. Popular culture demonstrates the African heritage. There is a rich body of oral narratives. Popular characters in these stories are Ti Lobo and Chibinho ("Uncle Wolf and Nephew"). Improvised singing is a feature of social gatherings and festivities. The melancholic *morna,* expressing the sorrows of emigration and love, is a song form unique to Cape Verde.

Cape Verdean literature

Since the late 19th century, Cape Verde has produced some outstanding writers and poets. Between 1936 and 1960 the cultural magazine *Claridade* ("Clarity") was the focus for an artistic movement that marked a break with Portuguese literary traditions and established a Cape Verdean identity. Baltasar Lopes da Silva, who used the pseudonym Osvaldo Alcântara for his poetry, is a key figure from this period. Later writers have extended the movement's interest in the Creole culture to use Crioulo as well as Portuguese. Corsino Fortes is the best-known poet of this later generation.

One television channel and two government radio stations broadcast in Portuguese and Crioulo. A weekly government newspaper, *Voz di Povo* ("Voice of the People"), is published in Praia. Foreign publications circulate freely, as do local magazines. There is a publishing house, the Cape Verdean Institute of Books, which specializes in works on Cape Verdean history and culture.

For statistical data on the land and people of Cape Verde, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

Prior to the arrival of the Portuguese there is no evidence of the islands having been inhabited, but it is thought that the Moors had visited Sal to collect salt supplies in previous centuries. In 1460 the Portuguese navigators Diogo Gomes and António de Noli sighted Maio and São Tiago. In 1462 the first settlers from Portugal landed on São Tiago, subsequently founding there the oldest European city in the tropics—Ribeira Grande. Sugar was planted in an attempt to emulate the success of the earlier settlement of Madeira. Cape Verde's dry climate was less favourable, but, with the development of transatlantic slave trade, the importance and the wealth of the islands increased. In 1532 the first bishop was consecrated. The prosperity of Ribeira Grande, however, attracted pirates, who attacked the city in 1541. The English later attacked it twice—in 1585 and 1592—the first time under the command of Sir Francis Drake. After a French attack in 1712, the city was abandoned. Portugal attempted to administer its possessions and commerce on the African coast through the islands. Until the 19th century, trade was controlled through the crown-issued monopoly contracts. English, French, and Dutch activity in the area meant, however, that the crown was never really able to enforce its edicts. Smuggling was rife.

The founding of Ribeira Grande

From the 17th to the 19th century Cape Verde was famous for its woven cotton cloth (*panos*). Cotton grew easily, and indigo produced a rich blue dye. The skill of narrow-loom weaving had come with the slaves from the western African coast. The cloths were a valuable form of currency for the slave trade on the mainland. American machine-woven cloth ended this industry in the mid-19th century.

With the decline of the slave trade (which was finally abolished in 1876) and with increasing drought, the prosperity of the islands slowly vanished. In the early 1800s they experienced recurrent drought and famine as well as government corruption and maladministration. Conditions improved toward the end of the 1800s, with the establishment at Mindelo of a coaling station and a submarine cable station. After World War I, prosperity again declined as fewer ships visited Mindelo. The colonial administration encouraged emigration to the cocoa plantations of São Tomé and Príncipe.

Portuguese administration

The Portuguese administration of Cape Verde was unified under a governor in 1587. The islands' status was changed in 1951 from that of a colony to an overseas province. In 1961 all citizens were given full Portuguese citizenship.

During the war for independence from Portugal (1961–75) fought by its colonies in Africa, Cape Verde was used as a garrison by the Portuguese army. Some Cape Verdeans fled to Guinea-Bissau to join the African Party for the Independence of Guinea-Bissau and Cape Verde (PAIGC), under the leadership of Amílcar Cabral.

On July 5, 1975, Cape Verde became an independent republic. The first president, Aristides Pereira, had been secretary general of the PAIGC since 1973. The island republic is a member of the Organization of African Unity and of the United Nations.

Disapproval of the coup in Guinea-Bissau in 1980 prompted the dissolution of the Cape Verde branch of PAIGC and resulted in the formation in 1981 of the PAICV. President Pereira was reelected in February of the same year. Cape Verde affirmed its nonaligned status and sought foreign aid to salvage its economy and to fund development. Cape Verde plays an active diplomatic role in its relationship with the other former Portuguese colonies in Africa. (W.M.B./C.S.Sh.)

For later developments in the history of Cape Verde, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

Chad

The Republic of Chad (République du Tchad) is an independent landlocked state in north central Africa. It has an area of 495,755 square miles (1,284,000 square kilometres). It is bounded on the north by Libya, on the east by The Sudan, on the south by the Central African Republic, and on the west by Cameroon, Nigeria, and Niger. Chad obtained independence from France on Aug. 11, 1960. The capital, N'Djamena (formerly Fort-Lamy), is almost

1,000 miles (1,600 kilometres) by road from the western African coastal ports.

PHYSICAL AND HUMAN GEOGRAPHY

Although it is the fifth largest country on the continent, Chad—much of the northern part of which lies in the Sahara—has a population density of only about 10 persons per square mile (four persons per square kilometre). Most of the population lives by agriculture; cotton is grown in the south, and cattle are raised in the central region. Chad is one of the world's poorest countries.

The land. The frontiers of Chad, which constitute a heritage from the colonial era, do not coincide with either natural or ethnic boundaries.

Relief and drainage. In its physical structure Chad consists of a large basin bounded on the north, east, and south by mountains. Lake Chad, which represents all that remains of a much larger lake that covered much of the region in earlier geologic periods, is situated in the centre of the western frontier; it is 922 feet (281 metres) above sea level. The lowest altitude of the basin is the Djourab Depression, which is 573 feet above sea level.

In the early Holocene, possibly until as recently as 7,000 years ago, the lake stood at a level of about 1,100 feet above sea level, or some 180 feet higher than today, and was as much as 550 feet deep. At that stage Mega-Chad, as it has been called, occupied an area of some 130,000 square miles and overflowed southward via the present-day Kébi River and then over the Gauthiot Falls westward to the Benue River and the Atlantic Ocean. Older dune systems, flooded by Mega-Chad, form linear islands in the present lake and extend hundreds of miles to the east, the interdunal hollows being occupied by diatomites and other lake sediments.

The mountains

The mountains that rim the basin include the volcanic Tibesti Massif to the north (of which the highest point is Mount Koussi, with an altitude of 11,204 feet [3,415 metres]), the sandstone peaks of the Ennedi Plateau to the northeast, the crystalline rock mountains of the Ouaddaï (Wadai) region to the east, and the Oubangui Plateau to the south. The semicircle is completed to the southwest by the mountains of Adamawa and Mandara, which lie mostly beyond the frontier in Cameroon and Nigeria.

Chad's river network is virtually limited to the Chari and Logone rivers and their tributaries, which flow from the southeast to feed Lake Chad. The remaining Chad waterways are either seasonal or are of insignificant size. The Chari, which arises from headstreams in the Central African Republic to the south, is later joined from the east by the Salamat Wadi and from the west by the Ouham River, its largest tributary. After entering an ill-defined area of swampland between Niellim and Dourbali, it flows through a large delta into Lake Chad. The Chari is about 750 miles in length and has a flow that normally varies between 600 and 12,000 cubic feet (17,000 to 340,000 litres) per second, according to the season. The Logone, which for some of its course runs along the Cameroon frontier, is formed by the junction of the Pendé and Mbéré rivers; its flow varies between 170 and 3,000 cubic feet per second, and its course is more than 600 miles long before it joins the Chari at N'Djamena. The level of Lake Chad fluctuates according to the flow of these rivers, as well as according to the degree of precipitation, evaporation, and seepage. The droughts of the 1970s and early '80s in the Sahel region of western Africa reduced the lake to record low levels. By 1985 it had been reduced to a pool, immediately to the north of the Chari–Logone mouth, occupying about 1,000 square miles.

Soils. Several types of soil formation occur in Chad, apart from the sand of the desert zone and the sheer rock of the mountainous areas. On the south side of Lake Chad the soils are derived from clayey deposits that accumulated on the floor of Mega-Chad. Along the seasonally flooded banks of the Chari and Logone rivers and the Salamat Wadi, hydromorphic (waterlogged) soils occur. Tropical iron-bearing soils, red in colour, are found on the exposed folds and mounds of the Ouaddaï region's upland slopes. In the Kanem region (area north of Lake Chad) subarid soils are characteristic, except in the depressions that occur between the dunes on the shores of Lake Chad, where hydromorphic soils liable to salinization are found.

Climate. Chad's wide range in latitudes (that extend southward from the tropic of Cancer for more than 15°) is matched by a climatic range that varies from wet and dry tropical to hot arid. At the towns of Moundou and Sarh, in the wet and dry tropical zone, between 32 and 48 inches (800 and 1,200 millimetres) of rain falls annually between May and October. In the central semiarid tropical (Sahel) zone, where N'Djamena is situated, between 12 and 32 inches of rain falls between June and September. In the north rains are infrequent, with an annual average of less than one inch being recorded at Largeau.

Chad thus has one relatively short rainy season. The dry season, which lasts from December to February everywhere in the country, is relatively cool, with daytime temperatures of about 85° to 95° F (29° to 35° C) and nighttime temperatures that drop to about 55° F (13° C). From March onward it becomes very hot until the first heavy rains fall. At N'Djamena, for example, daytime temperatures average more than 100° F (38° C) between March and June. Heavy rains begin at N'Djamena in July, and average daytime temperatures drop to the low 90s F (mid-30s C), but nighttime temperatures remain in the 70s F (20s C) until the onset of N'Djamena's dry, cool season in November.

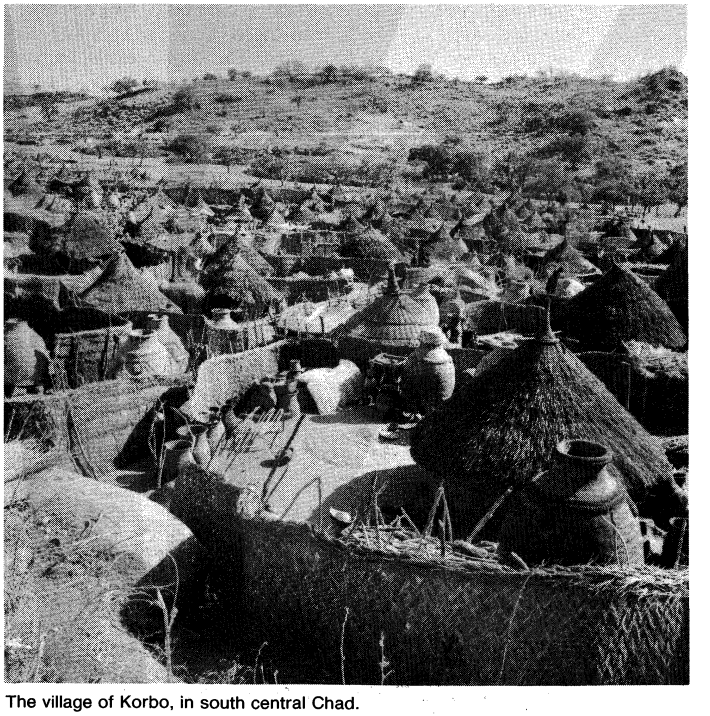
Three vegetation zones

Plant and animal life. Three vegetation zones, correlated with the rainfall, may be distinguished. These are a wet and dry tropical zone in the south, characterized by shrubs, tall grasses, and scattered broad-leaved deciduous trees; a semiarid tropical (Sahel) zone, in which savanna vegetation gradually merges into a region of thorn bushes and open steppe country; and a hot arid zone, composed of dunes and plateaus in which vegetation is scarce and occasional palm oases are to be found.

The tall grasses and the extensive marshes of the savanna zone have an abundant wildlife. There large mammals—such as the elephant, hippopotamus, rhinoceros, warthog, giraffe, antelope, lion, leopard, and cheetah—coexist with a wide assortment of birds and reptiles. The rivers and the lake are among the richest in fish of all African waters. The humid regions also contain swarms of insects, some of which are dangerous.

Settlement patterns. Conditioned by soil and climate, land is put to different uses in the three vegetation zones. In the wet and dry tropical zone, farmers cultivate rice and sorghum in the clay soils and peanuts (groundnuts) and millet in the sandier areas. Manioc, recently introduced,

Shostal Associates/Superstock



The village of Korbo, in south central Chad.

is also cultivated. Between the latitudes of 11° and 15° N, the retreat of the rivers in the dry season leaves behind flooded depressions called *yaere,* allowing a second crop of "dry season" sorghum, or *berbere,* to be cultivated. Since 1928 the cultivation of cotton in the area between the Logone and Chari rivers has been encouraged, first by the colonial administration and since 1960 by the national government. Cotton cultivation, while tending to upset the ecological balance by exhausting the soil, has nevertheless resulted in the introduction of a cash economy in place of a barter economy. The cultivation of rice, begun in 1958 in irrigated plots in the Bongor region, south of N'Djamena, has proved successful. A joint venture with France introduced sugarcane cultivation in the 1970s. Improved strains of both cotton and rice have produced higher yields.

The intermediate semiarid tropical zone is inhabited by both sedentary cultivators and nomadic pastoralists. The northern limit of the bloodsucking tsetse fly, deadly to cattle and the carrier of sleeping sickness to humans, is latitude 10° N; beyond this limit, extensive stock raising begins, occasionally in association with agriculture, as for example in the Kanem region. The inhabitants raise millet and grow peanuts wherever the mean annual rainfall exceeds 15 inches. Cotton is grown where and when rainfall exceeds 30 inches. Large herds of cattle migrate over the semiarid tropical zone in search of pasture and water. In very limited areas bordering Lake Chad, the presence of water allows three harvests of wheat and corn (maize) to be grown in some years on irrigated plots called polders. Elsewhere the seminomadic inhabitants are almost completely dependent upon rainfall. Drought has had serious repercussions, affecting both the livestock and the pastoralists, whose livelihood depends on milk products.

In the hot arid zone, nomads live among their herds of camels, frequenting palm groves growing in such oases as that at Largeau. Farther north, in the Tibesti Mountains, tiny plots of millet, tomatoes, peppers, and other minor crops are grown for local consumption, often in the shade of date palms. These garden crops depend on irrigation, the sources of water being springs breaking out from the sandstones and volcanic rocks at widely separated points and shallow wells in the sandy sediments flooring steep-sided valleys.

Urban life

Urban life in Chad is virtually restricted to the capital, N'Djamena. Founded in the early years of the 20th century, the city has undergone a dramatic growth in population. This growth is due not to a high degree of industrialization but to the other attractions of urban life. The majority of the population is engaged in commerce. Other major towns, such as Sarh (formerly Fort-Archambault), Moundou, and Abéché, are less urbanized than is the capital.

The people. The population of Chad presents a tapestry composed of different languages, peoples, and religions that is remarkable even amid the variety of Africa. The degree of variety encountered in Chad underscores the significance of the region as a crossroads of linguistic, social, and cultural interchange.

Linguistic groups. More than 100 different languages and dialects are spoken in the country. Although many of these languages are imperfectly recorded, they may be divided into the following 12 groupings: (1) the Sara-Bongo-Bagirmi group, representing languages spoken by about one million people in southern and central Chad, (2) the Mundang-Tuburi-Mbum languages, which are spoken by several hundred thousand people in southwestern Chad, (3) the Chado-Hamitic group, which is related to the Hausa spoken in Nigeria, (4) the Kanembu-Zaghawa languages, spoken in the north, mostly by nomads, (5) the Maba group, spoken in the vicinity of Abéché and throughout the Ouaddaï region of eastern Chad, (6) the Tama languages, spoken in the Abéché, Adré, Goz Béïda, and Am Dam regions, (7) Daju, spoken in the area of Goz Béïda and Am Dam, (8) some languages of the Central African groups, particularly Sango (also the lingua franca of the Central African Republic), which are spoken in the south, (9) the Bua group, spoken in southern and central Chad, (10) the Somrai group, spoken in western and central Chad, and (11) Mimi and (12) Fur, both spoken in the extreme east.

In addition to this rich assortment, Arabic is also spoken in various forms and is one of the two official languages of the country. The dialects spoken by the nomadic Arabs differ from the tongue spoken by settled Arabs. A simplified Arabic is spoken in towns and markets; its diffusion is linked to that of Islām.

French is the other official language, and it is used in communications and in instruction as well, although the national radio network also broadcasts in Arabic, Sara Madjingay, Tuburi, and Mundang. While a regional form of French, showing local linguistic and environmental peculiarities, is spoken widely in the towns, its penetration into the countryside is uneven. Its use is closely linked to the development of education.

Ethnic groups. As might be expected, the linguistic variety reflects an ethnic composition of great complexity. A general classification may nevertheless be made, again in terms of the three regions of Chad.

In the wet and dry tropical zone, the Sara group forms a significant element of the population in the central parts of the Chari and Logone river basins. The Laka and Mbum peoples live to the west of the Sara groups and, like the Gula and Tumak of the Goundi area, are culturally distinct from their Sara neighbours. Along the banks of the Chari and Logone rivers, and in the region between the two rivers, are found the Tangale peoples.

Peoples of the semiarid tropical zone

Among the inhabitants of the semiarid tropical zone are the Barma of Bagirmi, the founders of the kingdom of the same name; they are surrounded by groups of Kanuri, Fulani, Hausa, and Arabs, many of whom have come from outside Chad itself. Along the lower courses of the Logone and Chari rivers are the Kotoko, who are supposedly descended from the ancient Sao population that formerly lived in the region. The Buduma and Kuri inhabit the Lake Chad region and, in the Kanem area, are associated with the Kanembu and Tunjur, who are of Arabic origin. All of these groups are sedentary and coexist with Daza, Kreda, and Arab nomads. The Hadjeray (of the Guera Massif) and Abou Telfân are composed of refugee populations who, living on their mountainous terrain, have resisted various invasions. On the plains surrounding the Hadjeray are the Bulala, Kuka, and the Midogo, who are sedentary peoples. In the eastern region of Ouaddaï live the Maba, among whom the Kado once formed an aristocracy. They constitute a nucleus surrounded by a host of other groups who, while possessing their own languages, nevertheless constitute a distinct cultural unit. The Tama to the north and the Daju to the south have formed their own separate sultanates. Throughout the Ouaddaï region are found groups of nomadic Arabs, who are also found in other parts of south central Chad. Despite their widespread diffusion, these Arabs represent a single ethnic group composed of a multitude of tribes. In Kanem other Arabs, mostly of Libyan origin, are also found.

In the northern Chad regions of Tibesti, Borkou, and Ennedi the population is composed of black nomads. Their dialects are related to those of the Kanembu and Kanuri.

Religious groups. The great majority of Muslims are found in the north and east of Chad. Islāmization in Kanem came very early and was followed by the conversion to Islām of the major political entities of the region, such as the sultanates of Wadai, Bagirmi, and Fitri, and—more recently—the Saharan region. Islām is well established in most major towns and wherever Arab populations are found. It has attracted a wide variety of ethnic groups and has forged a certain unity which, however, has not resulted in the complete elimination of various local practices and customs.

Animism flourishes in the southern and most populous part of the country and in the mountainous regions of Guera. The various traditional religions provide a strong basis for cohesion in the villages where they are practiced. Despite a diversity of beliefs, a widespread common feature of traditional religion is the socioreligious initiation of young people into adult society.

In Chad, as elsewhere, Christian missionary work has not affected the Muslim population; it has been directed

toward the animist populations in the cities in the western regions south of the Chari River and in parts of the central uplands area. There are three Roman Catholic dioceses, with an archbishop at N'Djamena. There are some Protestant mission groups, and an effort has been made to form a Chad Evangelical church.

Demographic trends. More than two-fifths of the population of Chad are under the age of 15. Nearly one-fourth of the people are considered to be urban dwellers, the majority living in N'Djamena. The population is increasing at a comparatively low rate for an African country. Emigration—especially to The Sudan, Nigeria, and northern Cameroon—resulting from drought, conflict, and famine, may help to account for this.

The economy. *Resources.* Chad's principal mineral resource is natron (a complex sodium carbonate), which is dug up in the Lake Chad and Borkou areas and is used as salt and in the preparation of soap and medicines. Annual production is a few thousand tons, part of which is marketed in Nigeria. There are indications of deposits of gold in the Ouaddaï area, uranium in the Ennedi Plateau area, uranium and wolframite in the Aozou Strip in the far north, and bauxite near Laï. Oil has been found north of Lake Chad.

Agriculture and fishing. Cotton is Chad's primary product. Although it is basically an export crop, the processing of raw cotton provides employment for a majority of those in industry and accounts for most of Chad's export earnings. Most of the cotton fibre ginned in Chad's processing plants is exported to Europe and the United States.

Importance of livestock

Chad's livestock population constitutes its second most important economic resource and is primarily distributed across central Chad. Much of this wealth of livestock is not reflected in the national cash economy, however, and livestock products form less than one-tenth of exports. There is a refrigerated meat-processing plant at Sarh, which exports meat to the Congo region and Gabon. The government has tried to improve livestock by introducing stronger breeds and production by building new slaughterhouses.

Rice is produced in the Chari valley and in southwestern Chad, and wheat is grown along the shores of Lake Chad; little of either crop is processed commercially.

About half the fish caught is salted and dried for export. Most fish are caught in the Lake Chad, Chari, and Logone basins.

Industry. The development of industry is hampered by a shortage of power. Energy is generated by using oil products that have to be imported from Nigeria.

The primary industries, such as cotton ginning, slaughtering, and the milling of wheat and rice, are all associated with agriculture. Secondary industries are few and rely on imported materials.

Finance and trade. The country relies heavily on foreign financial assistance. The sums received exceed export earnings and in many years constitute as much as a quarter of the gross national product. The main imports are petroleum products, partly from Nigeria, plus cereals and manufactured articles, above all from France. Raw cotton is by far the main export, the amount varying annually from about 80,000 to 150,000 tons. In addition, live cattle, meat, and fish are exported to Nigeria.

Transportation. Chad's economic development is primarily contingent upon the establishment of an effective transportation network. There are three access routes to the sea—by road, river, or rail, through neighbouring countries. Most of the country's roads and trails are impractical for travel during part of the rainy season. Year-round traffic is possible on gravel-surfaced roads and on a paved section between N'Djamena and Guélendeng. Three major road axes, forming a triangle joining N'Djamena, Sarh, and Abéché, were completed but have fallen into disrepair. In 1985 a bridge across the Chari River to Kousseri, Cameroon, ended N'Djamena's dependence on an unreliable ferry for its road connection through Cameroon to the railhead at Ngaoundéré and the sea.

Rivers are of secondary importance, due to great seasonal fluctuations in water levels, with only about half of the total river length navigable year-round. The Chari is navigable between Sarh and N'Djamena between August and December, and the Logone is navigable between Mondou and N'Djamena in September and October. Two railways have their terminals near the Chad border. Across the Nigerian frontier to the west there is a railhead at Maiduguri, which links up with the Nigerian ports of Lagos and Port Harcourt. Across the Sudanese frontier to the east is the railhead at Nyala, which leads eventually to Port Sudan on the Red Sea. Air traffic plays an important role in the Chad economy, in view of the paucity of alternative means. N'Djamena's airport can accommodate large jets, and there are more than 40 secondary airports.

Administration and social conditions. *Government.* The constitution of 1975 was replaced by the *charte fondamentale* in 1978, and this in turn was abolished in 1979. In 1982, pending the adoption of a new constitution, the provisional *acte fondamental* was instituted. It assigned full powers to the president. The government consists of a Council of Ministers subordinate to the president and the National Consultative Assembly, an advisory body consisting of two representatives from each of the country's 14 *préfectures,* plus two representatives from the capital, all appointed by the president.

Civil war broke out in the mid-1960s when two guerrilla groups began a struggle to overthrow the government and create closer ties with Arab North Africa. Internal struggles and conflict with Libya continued in the 1980s under President Hissen Habré, a northerner, strengthening his grip on the state apparatus in N'Djamena.

Education. The size of the country, the dispersion of populations, and the occasional reluctance to send children to school all constitute educational problems that the government is endeavouring to overcome. Less than one-half of the school-age population is enrolled. Missions and public education services are responsible for primary education. Secondary and technical education is also available. The University of Chad, founded in 1971, offers higher education, and some Chad students study abroad.

Health and welfare. There are major hospitals at N'Djamena, Sarh, Moundou, Bongor, and Abéché. Other health facilities include dispensaries and infirmaries dispersed throughout the country. The government, in cooperation with the World Health Organization, has developed a health education and training program. Campaigns have been conducted against malaria, sleeping sickness, leprosy, and other diseases.

Cultural life. With its rich variety of peoples and languages, Chad possesses a valuable cultural heritage. The government has in the past encouraged cultural activities and institutions. There is a national museum, with a rich collection of prehistoric and traditional artifacts. The Chad Cultural Centre seeks to awaken a conscious interest in national traditions. The lives of the people have been so dislocated by war and famine since the 1960s, however, that Chad is more impoverished than ever, and the main efforts of the government and people are now directed toward survival.

For statistical data on the land and people of Chad, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (A.T.G./Ed.)

HISTORY

The region of the eastern Sahara and Sudan from Fezzan, Bilma, and Chad in the west to the Nile valley in the east was well peopled in Neolithic times, as discovered sites attest. Probably typical of the earliest populations were the Negroid cave dwellers described by Herodotus as inhabiting the country south of Fezzan. The ethnographic history of the region is that of gradual modification of this basic stock by the continual infiltration of nomadic and increasingly Arabicized white African elements, entering from the north via Fezzan and Tibesti and, especially after the 14th century, from the Nile valley via Darfur. According to legend, the country around Lake Chad was originally occupied by the Negroid Sao. This vanished people is probably represented today by the Kotoko, in whose country, along the banks of the Logone and Chari, was unearthed in the 1950s a medieval culture notable for work in terra-cotta and bronze.

The relatively large and politically sophisticated kingdoms of the central Sudan were the creation of Saharan Berbers, drawn southward by their continuous search for pasturage and easily able to impose their hegemony on the fragmentary indigenous societies of Negroid agriculturalists. This process was intensified by the expansion of Islām. There are indications of a large immigration of pagan Berbers into the central Sudan early in the 8th century.

The Kanem-Bornu state

From the 16th to the 19th century. The most important of these states, Kanem-Bornu, which was at the height of its power in the later 16th century, owed its preeminence to its command of the southern terminus of the trans-Saharan trade route to Tripoli.

Products of the Islāmized Sudanic culture diffused from Kanem were the kingdoms of Bagirmi and Ouaddaï, which emerged in the early years of the 17th century out of the process of conversion to Islām. In the 18th century the Arab dynasty of Ouaddaï was able to throw off the suzerainty of Darfur and extend its territories by the conquest of eastern Kanem. Slave raiding at the expense of animist populations to the south constituted an important element in the prosperity of all these Muslim states. In the 19th century, however, they were in full decline, torn by wars and internecine feuds. In the years 1883–93 they all fell to the Sudanese adventurer Rābiḥ az-Zubayr.

The 20th century. By this time the partition of Africa among the European powers was entering its final phase. Rābiḥ was overthrown in 1900, and the traditional Kanembu dynasty was reestablished under French protection. Chad became part of the federation of French Equatorial Africa in 1910. The pacification of the whole area of the present republic was barely completed by 1914, and between the wars French rule was unprogressive. A pact between Italy and France that would have ceded the Aozou Strip to Italian-ruled Libya was never ratified by the French National Assembly, but it provided a pretext for Libya to seize the territory in 1973. During World War II Chad gave unhesitating support to the Free French cause. After 1945 the territory shared in the constitutional advance of French Equatorial Africa. In 1946 it became an overseas territory of the French Republic.

Independence. A large measure of autonomy was conceded under the constitutional law of 1957, when the first territorial government was formed by Gabriel Lisette, a West Indian who had become the leader of the Chad Progressive Party (PPT). An autonomous republic within the French Community was proclaimed in November 1958, and complete independence in the restructured community was attained on Aug. 11, 1960. The country's stability was endangered by tensions between the black and often Christian populations of the more economically progressive southwest and the conservative, Muslim, non-black leadership of the old feudal states of the north, and its problems were further complicated by Libyan involvement.

Lisette was shouldered aside by an associate more acceptable to some of the opposition, N'Garta (François) Tombalbaye, a southern trade union leader, who became the first president of the republic. In March 1961 Tombalbaye was able to achieve a fusion of the PPT with the principal opposition party, the National African Party (PNA), to form a new Union for the Progress of Chad. An alleged conspiracy by Muslim elements, however, led in 1963 to the dissolution of the National Assembly, a brief state of emergency, and the arrest of the leading ministers formerly associated with the PNA. Only government candidates presented themselves at the new elections in December 1963, ushering in the one-party state.

(D.H.J./A.T.G.)

Civil war. In the mid-1960s two guerrilla movements emerged. The Front for the National Liberation of Chad (Frolinat) was established in 1966 and operated primarily in the north from its headquarters at the southern Libyan oasis of al-Kufrah, while the smaller Chad National Front (FNT) operated in the east central region. Both groups aimed at the overthrow of the existing government, the reduction of French influence in Chad, and closer association with the Arab states of North Africa. Heavy fighting occurred in 1969 and 1970, and French military forces were brought in to suppress the revolts.

By the end of the 1970s, civil war had become not so much a conflict between Chad's Muslim northern region and the black southern region as a struggle between northern political factions. Libyan troops were brought in at President Goukouni Oueddei's request in December 1980 and were withdrawn, again at his request, in November 1981. In a reverse movement the Armed Forces of the North (FAN) of Hissen Habré, which had retreated into The Sudan in December 1980, reoccupied all the important towns in eastern Chad in November 1981. Peacekeeping forces of the Organization of African Unity withdrew in 1982, and Habré formed a new government in October of the same year. Simultaneously, an opposition government under the leadership of Goukouni was established, with Libyan military support, at Bardaï in the north. After heavy fighting in 1983–84 Habré's FAN prevailed, aided by French troops. France withdrew its troops in 1984 but Libya refused to do so. Libya launched incursions deeper into Chad in 1986, and they were turned back by government forces with help from France and the United States.

Government of Habré

In early 1987 Habre's forces recovered the territory in northern Chad that had been under Libyan control and for a few weeks reoccupied Aozou. When this oasis was retaken by Muammar al-Qaddafi's Libyan forces, Habré retaliated by raiding Maaten es Sarra, which is well inside Libya. A truce was called in September 1987 that lasted into 1988.

For later developments in the history of Chad, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (A.T.G./Ed.)

The Gambia

The Gambia is a small republic with an area of 4,127 square miles (10,689 square kilometres). Essentially, The Gambia is a strip of land, from 15 to 30 miles (25 to 50 kilometres) wide and 295 miles long, on either bank of the Gambia River; except for a short coastline, it is surrounded by Senegal. Its unusual shape and size are attributable to territorial compromises arising from 19th-century Anglo-French rivalry in western Africa. The capital is Banjul.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The Gambia River is the country's dominant feature, flowing across a plateau of Miocene-Pliocene sandstone, compacted sediment composed predominantly of quartz grains formed from 23.7 to 1.6 million years ago. In the east, narrow valleys are separated by broad interfluves or flattish hills. In the west, lower and smaller sand hills alternate with depressions filled in with sand so that a flat plain is formed.

Climate. The climate is of the drier tropical type, characterized by a short and intense rainy season occurring between June and October and by a longer dry season. Near the coast the rainy season lasts longer and the rainfall is heavier, diminishing eastward. At Yundum the average annual rainfall is about 51 inches (1,300 millimetres) and the mean monthly temperature is 77° F (25° C), while at Basse Santa Su, about 270 miles inland, the comparable figures are 43 inches and 82° F (28° C). The relative humidity is high but drops from December to April, when the dry northeastern wind known as the harmattan is dominant.

Plant and animal life. The vegetation cover of The Gambia is savanna on the uplands, various kinds of inland swamp in the low-lying areas, and mangrove swamp along the brackish lower Gambia River. Few wild animals are native to the region, and those that survive are under pressure from the human and domestic animal populations. In the middle and upper river areas there are warthogs, monkeys, baboons, antelope, pygmy hippopotamuses, and crocodiles.

Settlement patterns. Patterns of settlement in The Gambia reflect three regions, expressed on both banks of the river, consisting of (1) the swamps adjacent to the river (and not extending above Kau-Ur), (2) the riverine flats, known as *banto faros* (from a Malinke word meaning "beyond the swamp"), and (3) the sandstone uplands. Most

rural settlement is concentrated on the uplands, which have the best-drained soils. A number of settlements are located in the *banto faros* on the middle course of the river, where there is less danger of flooding than in the swamps. Many villages are built on the boundary between the uplands and the riverine flats.

The people. The river basin was a focal point for migrating groups of people escaping the turmoil of western Sudanic wars dating from the 12th century. The Diola (Jola) are the people longest resident in the country; they are now located mostly in western Gambia. The largest group is the Malinke (Mandingo), comprising about two-fifths of the population. The Wolof are the largest population group in Banjul. Nomadic Fulani (Fula) settled the extreme upriver areas, and their kingdom, Fuladu, became a major power. The Soninke (Serahuli), an admixture of Malinke and Fulani, are also concentrated in the upriver areas. The population is more than 90 percent Muslim.

Ethnic composition

The economy. *Agriculture.* The Gambia can be described as a classic monoculture; peanuts are the only valuable crop. Land is cleared by the slash-and-burn technique, but farmers practice conservation. Most land is held in common by the villagers. There is a sharp division of labour, with men involved in planting, cultivating, and harvesting cash crops. Women cultivate subsistence crops such as cassava, yams, eggplant, tomatoes, and lentils. There are citrus orchards in the western area near Banjul.

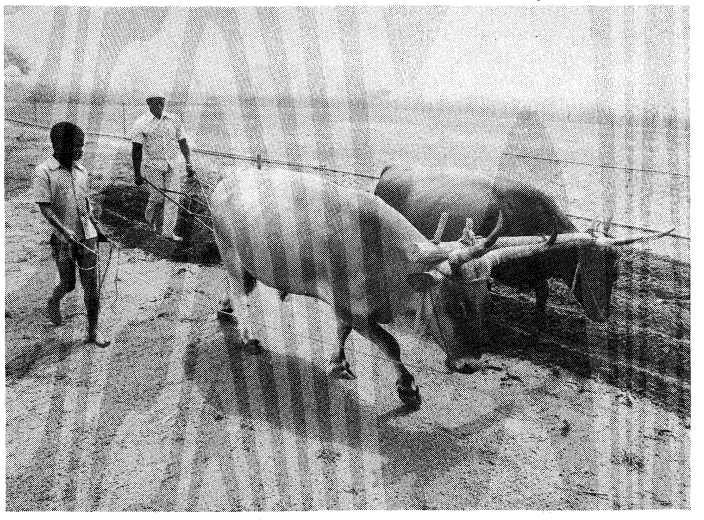
Importance of peanuts in the economy

The Gambian economy is based upon peanuts, the main cash crop and the source of most governmental revenue. Production has increased steadily with the wider use of fertilizers and ox-drawn equipment and the introduction of better seeds. In order to diversify the economy the government has encouraged the production of rice. A pilot scheme was begun in the mid-1960s to introduce plantation oil palm production. Stock farming, always a factor in the Fulani culture, has also received government support. The drought years of the 1970s and '80s seriously damaged agricultural production, particularly upriver.

Fishing. There is some potential for commercial fishing offshore and in the river. Most Gambians are not fishermen, and those who are have been handicapped by inadequate equipment. The government has loaned small amounts for the purchase of motorized fishing boats and the construction of smoke huts for the processing of bonga (shad, or West African herring), which is exported to other western African states.

Industry. The only significant industry is peanut processing. The crop is sold to agents of the Gambia Produce Marketing Board, which fixes the season's price in advance, pays the producers in cash, and sells the crop overseas. The agents arrange for transportation of the peanuts to Banjul or Kuntaur, where the nuts are shelled before being shipped. After shelling, a large part of the crop is pressed into oil at pressing mills. The residue is used as cattle cake.

© Georg Gerster—Comstock, Inc.



Ox-drawn equipment used for plowing in The Gambia.

Trade and finance. The Gambia has a relatively large volume of trade for a small country. In the early 1980s, however, the country had a yearly adverse balance of trade reflecting the losses caused by the drought. Besides peanuts, exports include cotton, rice, and cattle. All manufactured items must be imported; other imports include petroleum products, lumber, and cement. The chief trading partners are France, Britain, and Senegal.

Transportation. The Gambia River has historically been the chief route from the interior to the coast, but a modern all-weather road now reaches the eastern border and parallels the river on both sides. There are no railways and no domestic air services. The main port is Banjul, and there is an international airport nearby at Yundum.

Administration and social conditions. *Government.* The 50-member House of Representatives holds legislative power. There are 36 elected members, five elected by the chiefs in assembly, and eight nonvoting nominees; the attorney, who is nominated but has a vote, is also a member. The People's Progressive Party (PPP) has dominated the legislature since independence. The government is headed by a president, who is elected by universal suffrage to a five-year term. The vice president and Cabinet members are appointed by the president from among the 36 elected members of the house. The highest judicial body is the Supreme Court, whose members are appointed by the president.

Local administration in each of the 35 districts is theoretically the responsibility of the local chief, who is assisted by councillors. The districts are grouped into seven divisions, six of which are administered by councils consisting of a majority of elected members. Actual power, however, is in the hands of district officers appointed by the central government. Banjul is administered by a city council.

Education. Education at the primary level is free but not compulsory. There are secondary and post-secondary schools, including a teacher-training college at Brikama. Gambian students seeking higher education also travel to Sierra Leone, Britain, or the United States.

Health and welfare. There are several general hospitals and a number of health centres, dispensaries, and maternity and child care clinics. The Medical Research Council at Fajara investigates tropical diseases. Nonetheless, health care in The Gambia is poor. The infant mortality rate at about 170 per 1,000 live births (in the 1980s) is the highest in western Africa, and life expectancy at about 43 years is below average for the region.

Cultural life. The peoples of the Senegambia have not produced any significant art, though there were blacksmiths in all societies and a few drum and *kora* (a complex stringed instrument) makers. Dance and music were tied to village activities, and some songs were part of the repertoire of the praise singers called *griots.*

The only newspaper, the government-operated *News Bulletin,* circulates mainly in Banjul. Radio Gambia, also government-run, broadcasts in English, French, and various Gambian languages. In Banjul is a small museum whose holdings are mainly anthropological.

For statistical data on the land and people of The Gambia, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (E.R.A.F./H.A.G.)

HISTORY

Precolonial history. Gambian history before the arrival of Europeans is speculative. The Malinke and Wolof kingdoms, fully established by the 19th century, were still in the formative stages when the Portuguese explorer Alvise Ca' da Mosto (Cadamosto) arrived in 1455. The Malinke were the westernmost peoples of the old Mali empire. The Wolof probably migrated from the Songhai regions, and the Fulani pastoralists were part of a migration from the Futa Toro. Although locally powerful, none of the small Gambian kingdoms were ever strong enough to dominate Senegambia. Continuing internecine warfare made it easy for the French and British to dominate the territory.

European colonization. The first Europeans in the Gambia, the Portuguese, established trading stations in the late 1400s but abandoned them within a century. Trade possibilities in the next two centuries drew English, French,

Dutch, Swedish, and Courlander trading companies to western Africa.

The 18th century witnessed a struggle for prestige in Senegambia between France and England. Trade was minimal, and no chartered company made a profit. French and British outposts changed hands a number of times. Britain controlled the entire area between 1763 and 1783, but the Treaty of Versailles returned all French possessions. British presence on the river was maintained until 1823 by individual traders.

The founding of Bathurst

Captain Alexander Grant was sent to the Gambia in 1816 to reestablish a base from which the navy could operate to control the slave trade. He purchased Banjul Island (St. Mary's) from the king of Kombo, built barracks, laid out a town, and set up an artillery battery to control access to the river. The town, Bathurst (now Banjul), grew rapidly with the arrival of traders and workers from Gorée and upriver. The Gambia was administered as a part of the British West African Federation from 1821 to 1843. It was a separate colony with its own governor until 1866, when control was returned to the governor-general at Freetown until 1889.

British domination of the riverine areas seemed assured after 1857, but the increasing importance of peanut cultivation in Senegal prompted a new imperialism. By 1880 France controlled Senegal; in the 1870s the British attempted twice to trade the Gambia to France, but opposition at home and in the Gambia foiled these plans. Complicating matters was the series of religious conflicts, called the Soninke–Marabout Wars, lasting a half century. Only one Muslim leader, Maba, emerged who could have unified the various kingdoms, but he was killed in 1864. By 1880 the religious aspect had all but disappeared, and the conflicts were carried on by war chiefs such as Musa Mollah, Fodi Silla, and Fodi Kabba.

British protectorate. A Paris conference in 1889 obtained French agreement to British control of the Gambia River and the drawing of the present-day boundaries. In 1900 Britain imposed indirect rule on the interior, or protectorate (established in 1894), areas, dividing the Gambia into 35 chiefdoms, each with its own chief. The real power was concentrated in the governor and his staff at Bathurst.

Except for some trouble with slave-raiding chiefs, the Gambia enjoyed peace after its separation from Sierra Leone. In 1906 an ordinance was passed abolishing slavery throughout the protectorate. During World War II the Gambia contributed soldiers for the Burmese campaign and was used as an air staging post.

Independence. Political parties were late in developing, but by 1960 there were a number whose leaders demanded independence. Britain, believing that eventually the Gambia would merge with Senegal, gave the territory revised constitutions in 1954, 1960, and 1962 and finally granted it independence within the Commonwealth in February 1965. The Gambia became a republic on April 24, 1970. The first president, Sir Dawda Jawara, head of the PPP, was returned in all elections after 1972. In 1981 an attempt to overthrow the government was put down with the aid of Senegalese troops after heavy fighting in Banjul.

Senegambia

In the aftermath, leaders of both countries created the confederation of Senegambia. This plan called for each state to retain independence of action in most areas, but military and economic resources were to be integrated. A Senegambian executive and legislature were also established, but the confederation still exists mainly on paper.

For later developments in the history of The Gambia, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (H.A.G.)

Mali

The Republic of Mali (République du Mali) is a landlocked state in central western Africa. Bounded on the north by Algeria, on the east by Niger and Burkina Faso, on the south by Côte d'Ivoire and Guinea, and on the west by Senegal and Mauritania, it covers an area of 478,841 square miles (1,240,192 square kilometres). Bamako is the national capital. As a part of French West Africa from 1898 to its independence in 1960, it was known as the French Sudan. Its current name, taken at the time of independence, is derived from the Mali empire of the upper and middle Niger, which was ruled by the Malinke (Mandingo) from the 13th to the 16th century.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Mali's landscape is largely flat and monotonous. Two basic relief features can be distinguished—plateaus and plains. The highland regions are localized and discontinuous.

The plateaus and the plains

The plateaus of the south and southwest (extensions of the Fouta Djallon highlands of Guinea and the Guinea Highlands of Guinea and Côte d'Ivoire) lie between about 1,000 and 1,600 feet (300 and 500 metres) above sea level but attain heights approaching 2,000 feet in the Mandingue Plateau near Bamako and more than 2,100 feet near Satadougou.

The plateaus of the southeast and east, also extensions of the Guinea Highlands, are a series of small, broken hills. Altitudes in the southeast range between almost 1,000 feet in the region of Sikasso and 1,739 feet at Mount Mina. East of the Niger River the Dogon Plateau descends gently westward to the river valley but ends in abrupt cliffs on the southeast. These cliffs reach an altitude approaching 3,300 feet at Bandiagara.

The only marked relief feature in the north is the Iforas Massif. An extension of the mountainous Hoggar region of the Sahara, this heavily eroded sandstone plateau rises to altitudes of more than 2,000 feet. Northern and central Mali are composed of the plains of the Niger River basin and of the Sahara.

Drainage. The drainage system is composed of the Sénégal and Niger rivers and their tributaries. The Sénégal River flows in a northwesterly direction across Mali for 420 miles (670 kilometres) on its course to the Atlantic Ocean. Its main headwaters—the Bafing and the Bakoye (Bakhoy) rivers—rise in the Fouta Djallon and join at Bafoulabé to form the Sénégal. The river then flows to the west across the plateau region, where it is broken by falls at Gouina and Félou.

For 1,100 miles, more than one-third of its total length, the Niger River flows through Mali. Rising in the Fouta Djallon, the river is of significant size by the time it enters Mali near Kangaba. It flows to the northeast across the Mandingue Plateau, where it is broken by falls at Sotuba. Reaching Koulikoro, it spreads out in a wide valley and flows majestically to its confluence with the Bani River at Mopti. The Niger then forms an interior delta because the land is flat and the river's descent almost nonexistent. The river breaks down into a network of branches and lakes as it continues northward. At Bourem the Niger makes a great bend to the south, known as the Niger Bend, and flows past Gao and Ansongo to the Niger border at Labbezanga.

The flow of the Niger varies seasonally. High waters occur on the upper Niger from July to October, at the delta from September to November, and at the bend from December to January. Periodic floods and the rich alluvial soils in the central delta make the Niger valley an important agricultural region.

Soils. The soils in Mali are generally poor. In the south, ferruginous (iron-bearing) soils are shallow and form a hard, red crust because of intense evaporation. The desert region is composed of sand, rock, and gravel.

Climate. Mali lies within the intertropical zone and has a hot, dry climate. The sun is near its zenith throughout most of the year. In general, there are two distinct seasons, the dry and the wet. The dry season, from November to June, is marked by low humidity and high temperatures and is influenced by the *alize* and harmattan winds. The *alize,* from the northeast from December to February, causes a relatively cold spell, with temperatures averaging 77° F (25° C). From March to June the dry, hot harmattan blows from the east out of the Sahara and sweeps the soil into dusty whirlwinds. The wind is accompanied by daytime temperatures of 104° to 113° F (40° to 45° C).

During the rainy season from June to October, the monsoon wind blows from the southwest. Preceded by large, black clouds, the heavy rainstorms often include gusty

winds and much lightning and thunder. Temperatures lessen somewhat in August, when most of the rainfall occurs.

Climatic zones

The country can be divided into three climatic zones—the Sudanic, the Sahelian, and the desert zones. Sudanic climate occurs in almost one-third of the country, from the southern border to latitude 15° N. It is characterized by an annual rainfall of 20 to 55 inches (510 to 1,400 millimetres) and average temperatures of 75° to 86° F (24° to 30° C). The Sahel, or the area bordering the Sahara, receives between eight and 20 inches of rain a year and has average temperatures between 73° and 97° F (23° and 36° C). In the Sahara, temperatures during the day range from 117° to nearly 140° F (47° to 60° C), while at night the temperature drops to 39° to 41° F (4° to 5° C).

Plant and animal life. There are two main vegetation zones that correspond to the climatic regions of the Sudan and the Sahel. In the Sudanic zone there are localized forest corridors along the Guinean border and the river valleys. The rest of the area is covered with savanna. The trees include the nere, or twoball nitta tree (*Parkia biglobosa*), the karite (*Butyrospermum parkii*), the cailcedra (Senegal khaya; *Khaya senegalensis*), and the kapioka. The incidence of trees decreases to the north as the Sudanic zone merges with the Sahel. The Sahel is characterized by steppe vegetation; drought-resistant trees such as the baobab, the doum palm, and palmyra are found. These trees also disappear to the north, however, where short, thorny plants such as the mimosa, acacia, and *cram-cram* (*Cenchrus biflorus,* a member of the grass family) occur. All vegetation gradually disappears as one enters the Sahara region.

The animal life of the Sudan and of the Sahel is rich and varied. Large, herbivorous mammals include gazelles, antelopes, giraffes, and elephants. The main carnivores are lions, panthers, and hyenas. Crocodiles and hippopotamuses inhabit the rivers, and there is a wide variety of monkeys, snakes, and birds (including the ostrich). There is a national park along the Baoulé River in the west and an animal reserve between Ansongo and Ménaka in the east.

Settlement patterns. Mali is traditionally divided into the nomadic region of the Sahel and the Sahara and the agricultural region of the Sudanic zone. More than four-fifths of the population is rural. The rural population lives in thatched dwellings grouped together in villages of between 150 and 600 inhabitants. The villages are surrounded by cultivated fields and grazing lands. The older towns, such as Djenné, Timbuktu (Tombouctou), Gao, and Ségou, are built in the characteristic Sudanese style of architecture. The newer towns, such as Bamako, Kayes, San, and Kati, consist of a central business district, around which African residential districts are grouped. The houses are built of a mixture of earth and cement.

The people. Mali contains two distinct, stratified societies. Most of the population belong to traditional social groups, which have inherited hierarchical social structures. These groups consist of nobles, vassals, and members of various castes, all of whom acquired their status by birth.

Rural and urban society

The second Malian society is formed by the urban population. Privileged groups are the educated government officials and the traders. The middle socioeconomic group is composed of civil servants and industrial workers. The lowest group is made up of the unemployed.

Ethnic groups. What is known as the "white" population includes nomadic groups of Berbers (including the important Berber subgroup of the Tuareg) and the Arab-Berber group known as the Moors. These groups live in the Sahelian zone and north of the Niger Bend.

The black population is composed of numerous agricultural groups, some of whom are descended from the peoples of the ancient empires of Ghana, Mali, and Songhai. The largest group are the Bambara, who live along the upper Niger River. The Soninke, or Sarakole, are descended from the founders of the Ghana empire and live in the western Sahelian zone. The Malinke, bearers of the heritage of the Mali empire, live in the southwest, while the Songhai are settled in the Niger valley from Djenné to Ansongo. The Dogon live in the plateau region around Bandiagara. The Voltaic group includes the Bwa, or Bobo, the Senufo, and the Minianka; they occupy the east and southeast.

The Fulani, or Peul, are nomadic pastoralists of the Sahel and Macina. Other ethnic groups of note include the Tukulor (Tokolor), the Khasonke, the Bozo, and the Somono.

Linguistic groups. French is the official national language. There are several indigenous languages and dialects, which roughly correspond to either ethnic groups or regions. The most important is the Mande group, which includes Bambara (spoken by almost two-thirds of the population), Malinke, Khasonke, and Wasulunka (Ouassoulou). Soninke and Dogon are also related to Bambara; Dogon includes many dialects. The related but autonomous languages of the Voltaic peoples are Bwa, Senufo, and Minianka. The Fulani and Tukulor speak Ful (Fulah, or Peul), and Songhaic is spoken all along the Niger Bend. The Tuareg have retained their ancient Berber language and written script, *tifinagh,* which is related to that of ancient Libya. The Moors speak Arabic.

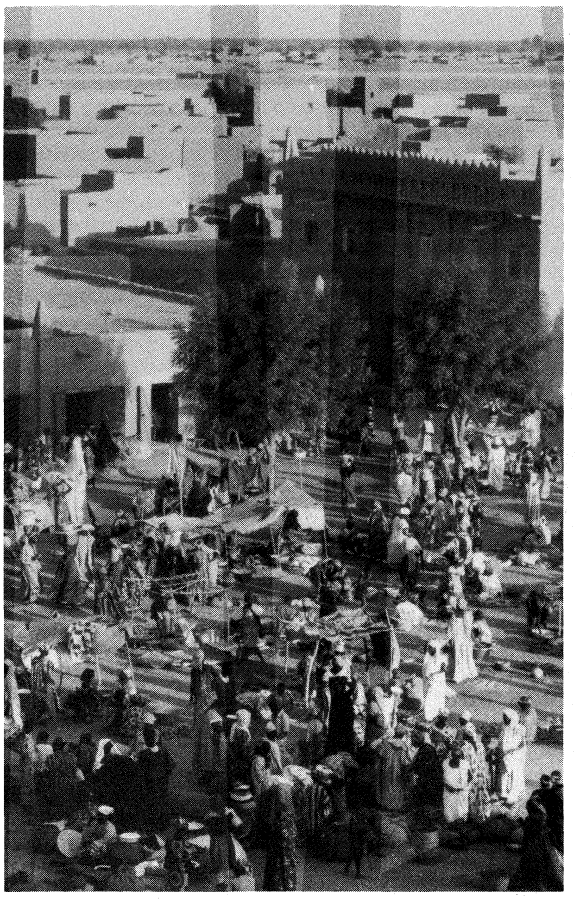
Religious groups. There are three main religions. Islām is practiced by about nine-tenths of the population, animism by most of the rest, and Christianity by a small number. Islāmization dates to the 11th century and has eclipsed traditional animism among the Soninke, Songhai, Tukulor, Moors, Tuareg, and Fulani. Animism continues as the religion of the Voltaic peoples, the Malinke, and the Bambara.

Demographic trends. The population of Mali continues to grow, though more slowly since 1980 than in the previous two decades. Life expectancy at birth has risen to 43 years for males and 46 years for females, and there has been a slight decline in both the birth and death rates, though both remain very high by world and African standards. The population is heavily weighted toward the young, the median age being about 17 years. Population densities throughout Mali are low, and in the more remote Gao and Timbuktu regions densities are only about three persons per square mile. These have long been regions of sparse population, but the droughts of the 1970s and '80s have led many of Mali's Tuareg to migrate either to the towns or, if their herds have managed to survive, to find new grazing lands farther south in Mali or in neighbouring Burkina Faso. Predictably there has been a major increase in the permanent urban population, which now stands at about 20 percent of the entire population. Unemployment in the urban areas is high, however. Less than half of the labour force is employed in industry, and many people are involved in small-scale commerce. Where opportunities exist, Malians migrate to France and other European nations, mainly for educational purposes.

The economy. Mali's is basically an agricultural economy. Because the northern half of the country is occupied by the Sahara, most human activity is concentrated in the more southerly regions, in particular in the valleys of the Niger and Sénégal rivers and their tributaries. Most agriculture is at the subsistence level, and, for many people, cash crops provide a valuable income supplement. Pastoralism is also important. Progress in the rural sector has been limited by an unfavourable climate, by periodic droughts since the late 1960s, and by low levels of technology. Other sectors are no further advanced: the development of Mali's extensive mineral and water resources is limited, and the country's industrial sector, which is still in its infancy, concentrates heavily on food processing. Foreign exchange is obtained chiefly from the export of primary commodities that have suffered from volatile world markets and foreign currency fluctuations. The revenue is insufficient to cover the cost of Mali's highly processed imports from France and other Western nations. Added to its problems, Mali has suffered severely from resource mismanagement, and the national debt continues to grow.

At the time of independence in 1960, the government adopted a policy of socialism. State companies and rural cooperative societies were organized to regulate both the production and the distribution of goods. Since the military coup d'état in 1968, socialist policy has been mitigated by the encouragement of private business.

Mali's form of socialism



Marketplace at Djenné, Mali.
© Brian A. Vikander—West Light

Bilateral external aid is provided largely by France, the United States, the nations of the European Economic Community (EEC), and the nations of the Organization of Petroleum Exporting Countries (OPEC). International aid is granted by such organizations as the United Nations, the European Development Fund (FED), and the United Nations Development Programme (UNDP).

Since 1981 the Mali government has responded to pressures from the World Bank, the International Monetary Fund (IMF), and aid donors to introduce measures to encourage private investment and enterprise, liberalization of domestic markets, and the general reduction of state control. Planning is based on five-year periods, and the plan of 1987–91 included in its objectives self-sufficiency in food, economic reforms, a campaign against desertification, and a scheme for protection and improvement of grazing land.

Resources. Mali's natural resources remain relatively undeveloped. Iron is the most widespread mineral resource, but it is not currently exploited because of Mali's limited infrastructure. Deposits are found in the west near the Senegal and Guinea borders. Bauxite deposits are located near Kayes and on the Mandingue Plateau. Manganese is also found, and there are phosphate deposits in the Ansongo region.

Important deposits of gold are at Kalana near Bougouni, on the Mandingue Plateau, and in the Iforas Massif. Lithium has been discovered near Kayes and Bougouni, and there are uranium deposits in the Iforas. There are also traces of tungsten, tin, lead, copper, and zinc, as well as deposits of salt, marble, kaolin (china clay), and limestone.

Agriculture and fishing. Subsistence and commercial agriculture are the bases of the Malian economy. More than 80 percent of the working population is in subsistence agriculture, and the government supports the development of commercial products. Areas of cultivation are located in the Sudanese and Sahelian zones; an agricultural area of major importance is the inland Niger delta. Crops such as millet, rice, wheat, and corn (maize), as well as potatoes, yams, and cassava, are the main subsistence crops. Cotton and peanuts (groundnuts) are the important commercial crops; sugarcane, tobacco, and tea are also grown for market. Market gardens produce a variety of vegetables and fruits, including cabbages, turnips, carrots, beans, tomatoes, bananas, mangoes, and oranges. Irrigation projects have been developed on the Niger near the towns of Ségou and Mopti. Agricultural products

The major areas for the raising of livestock are the Sahel and the area around Macina. There are cattle, sheep, and goats in the country.

Mali is one of the largest producers of fish in western Africa. The inland delta is a particularly important fishing ground, though the drought has caused a major setback to the industry.

Industry. Most industrial enterprises engage in the processing of food and other agricultural products. There are several rice mills, flour mills, and cotton gins. The oil and soap factory of Koulikoro satisfies domestic demands for its products. There are breweries, a tannery, and a refrigerated slaughterhouse at Bamako; fruit-preservation plants at Baguinéda; a sugar refinery at Dougabougou; and cigarette and match factories at Djoliba.

The Malian Company of Textiles (Comatex) produces cotton fibre and cloth, while the Textile Industry of Mali (Itema) manufactures printed cloth and blankets. There are also shops for the construction of motorcycles, the repair of machinery, and the assembly of radios. Handicrafts are important, and the Malians are noted for their clothing, pottery, shoes, baskets, and wood carvings. Handicrafts

Although mineral resources are extensive, the mining industry is minimal. Exploited deposits are those of salt (at Taoudenni), marble and kaolin (at Bafoulabé), gold (at Kalana), and limestone (at Diamou).

Electricity is largely produced in thermal power stations, but the role of hydroelectric power is growing. Thermal stations are located in Bamako and other large towns. Hydroelectric power is produced at the Sotuba and Markala dams on the Niger River, at the Felou Dam on the Sénégal River, and at the Sélingué Dam on the Niger near Bamako. Construction of the Manantali Dam on the Sénégal is a joint venture with Senegal and Mauritania that will add substantially to power capacity. Mali has also begun to exploit solar energy; solar-powered pumps provide electricity to some 30 villages, and the world's first commercial solar power station was established at Diré.

Finance and trade. The Central Bank of Mali, managed equally by Mali and France, controls the nation's credit and the exchange rate between the CFA (Communauté Financière Africaine) franc (the monetary unit of the country) and the French franc. The Development Bank of Mali finances development projects, while the Malian Bank of Credit and Deposits and the French-owned West African International Bank carry out credit and depository functions. Several French insurance companies maintain offices in Bamako.

The most important export items are cotton, peanuts, live animals, and dried and smoked fish. Imports consist largely of textiles, food products, automobiles, iron and steel, and petroleum products. Mali is a member of the Organization for the Development of the Sénégal River, which also includes Senegal and Mauritania. Despite strict customs controls, smuggling—especially of cattle and fish—is considerable.

Transportation. Mali's transportation systems are concentrated in the Sudanic and Sahelian regions. Because Mali is landlocked, its major transport routes connect with those of neighbouring countries and their ports to provide it with outlets to the sea.

There are several main axes of paved roads radiating from Bamako. The road from Bamako to Bougouni and Sikasso connects with the Côte d'Ivoire road running to the port of Abidjan. A road links Bamako with Kankan in Guinea and Monrovia in Liberia. Another main road runs through Bamako, Ségou, San, Mopti, Gao, and Ansongo to the border with Niger. Construction of a trans-Saharan road connecting Gao with North Africa was begun in 1979.

The one railroad track runs from Koulikoro, a short distance northeast of Bamako, northwestward to Kayes and to Kidira, on the Senegal border, where it connects with the Senegalese railway to Dakar. Another railway line is planned to link Bamako with the Guinean railway, which runs to Conakry.

Inland waterways

The Sénégal River is navigable from July to October from Kayes in the west of the country to Saint-Louis, Senegal, at the river's mouth. The Niger River is navigable throughout its length in Mali from July to January.

A national airline, Air Mali, used to operate both domestic and international flights. Mismanaged and with large debts, Air Mali ceased operation following an accident in 1986. Early in 1988 the government announced that Air Mali would be restarted. Mali's main airport is at Bamako, with others at Ségou, Mopti, Goundam, Timbuktu, Gao, Kayes, Kéniéba, Nara, and Niono. Bamako is also linked by air to France, Morocco, and other western African nations.

Administration and social conditions. *Government.* Upon independence in 1960, a constitution was granted that guaranteed parliamentary democracy through universal suffrage and provided for an elected National Assembly; the provisions of the constitution were not fully implemented, however. The military government that took power in 1968 suspended the constitution and forbade all political activity. A new constitution was approved in a national referendum on June 2, 1974. The president was elected to a six-year term and the 82 members of the National Assembly to three-year terms. This was changed by the Assembly in 1985, and there is no longer a limit to the length of time a president may hold office. The country is ruled by a single party, the Mali People's Democratic Union (UDPM).

The country is divided into the seven *régions* of Kayes, Koulikoro, Sikasso, Ségou, Mopti, Gao, and Timbuktu and the district of Bamako. Each of the *régions* is further divided into administrative units called *cercles* ("circles"), which are in turn subdivided into smaller units called *arrondissements.* Each *région* is administered by a governor, who coordinates the activities of the *cercles* and implements economic policy. The *cercles* provide nuclei for the major government services; their various headquarters provide focal points for the health service, the army, the police, local courts, and other government agencies. The *arrondissement* is the basic administrative unit, and its centre usually houses a school and a dispensary. It is composed of several villages, which are headed by chiefs and elected village councils.

Justice. At the head of the judicial system, the Supreme Court exercises both judicial and administrative powers; it is the court of first and last resort in matters concerning the government. The Court of Appeal, located in Bamako, tries all cases on appeal from ordinary tribunals. The Special Court of State Security holds trials for crimes against state property, especially concerning itself with charges of embezzlement. There are more than 50 tribunals and more than 70 magistrates. Justices of the peace have full powers to judge ordinary civil, commercial, and financial cases; they sit in the headquarters of the *cercles* and also travel to the major towns of the *arrondissements.*

Since 1960 new law codes have liberated women from traditional restraints, defined the rights and duties of citizens, and modified the penal procedure.

Education. French is the only language of instruction. Primary and secondary education are compulsory and free from six to 15 years of age and are combined in the nine-year curriculum of the *cycle fondemental* ("fundamental educational level"). The general secondary school, or *lycée,* provides the last three years of traditional secondary education. Higher education—geared directly to the needs of the government—is obtained in state colleges. These colleges include teacher-training colleges, a college of administration, an engineering institute, a polytechnic institute, and a school of medical assistants. Many of Mali's university students study abroad, especially in France; Senegal, and the Soviet Union.

Health and welfare. State hospitals at Bamako and Kati are supplemented by a network of medical centres, maternity centres, dispensaries, and a mobile service that visits patients in rural areas. Research centres in tropical ophthalmology and leprosy treat patients at Bamako. In spite of improvements in the statistics, rates of child and infant mortality remain among the highest in the world.

Cultural life. In spite of being one of the poorest nations in the world, Mali has long functioned as a crossroads between northern and western Africa and has developed a rich cultural tradition. Situated between the Arab world to the north and the black African nations to the south, it has for centuries been a cultural meeting place.

Music, dance, and traditional crafts

Music and dancing are the most common cultural activities; they form an especially rich heritage among the Malinke and Songhai peoples. The Bambara and the Voltaic groups excel in the creation of wood carvings of masks, statues, stools, and objects used in animist worship. The *tiewara,* or gazelle mask, of the Bambara is remarkable for its fineness of line.

Architecture is well developed in the Niger valley. The Sudanic style finds typical expression in the storied houses and mosques of Djenné and Timbuktu. Localized handicrafts include jewelry making by the Mandingo people, leatherworking around the Niger Bend, the weaving of geometric designs into cotton cloth, and the carving of statues for the tourist trade.

The Museum of the Institute of Research and Documentation at Bamako contains collections of art from most of the country's regions. The National Archives of Mali, the National Library, and the Institute of Human Sciences are also located in Bamako, as is the Municipal Library. The Centre of Arab Documentation is located in Timbuktu, and there is a French Centre of Documentation in Bamako. The Library of the Office of Niger in Ségou covers agriculture, irrigation, and general science.

The government promotes popular culture through the Committee of Youth and Sports. Youth associations organize sports, theatrical, musical, and dancing activities. Competitions are presented in Bamako during the biennial Youth Week. The Malian Ballet Troupe performs throughout the world. Artists are trained at the National Institute of Arts and at the Artisan Centre of Bamako.

Mali's one newspaper, *L'Essor,* is published by the UDPM and is far less effective in disseminating information than is the radio, not least because its circulation is limited to the literate and effectively to Bamako. There are two cultural journals of note—the first, *Sunjata* ("Lord Lion," leader in the ancient empire of Mali), is produced by the government, while the second, *Jamana* ("The Nation"), is independent.

The national radio station, Radiodiffusion Télévision Malienne, is run by the government and broadcasts for more than 5,000 hours per year. News bulletins, general information, and educational programs take up two-thirds of broadcast hours; the remainder are devoted to entertainment, cultural, and religious programs and to advertisements. There has been a dramatic rise in the number of radios in Mali, and it is these that permit communication with the more remote regions. Television, run by the UDPM, was introduced in 1983 and is limited to Bamako.

For statistical data on the land and people of Mali, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (K.M.B./Ed.)

HISTORY

Precolonial history. Rock paintings and inscriptions as well as Paleolithic and Neolithic remains have been found throughout Mali. In 1927, Neolithic human skeletal remains dating to 5000 BC were found in the Sahara at Asselar.

Rich gold deposits in the west and southwest constituted the principal resource in the economic life of early urban entrepôts and a succession of political states. An important trading centre, Djenné-Jeno, arose about 250 BC in the inland delta of the Niger River. It flourished until the 11th century AD, after which it was eclipsed by Djenné, a trading centre founded by Muslim Soninke about the 13th century AD. Terra-cotta statues, dating to as early as AD 800, have been found at Djenné-Jeno and at other sites in Mali.

The Mali empire

The export trade in gold and in slaves, ivory, civet, and gum arabic moved over trans-Saharan caravan routes from the Niger River valley to North Africa for almost a thousand years. This trade was controlled by the Soninke kingdom of Ghana (4th–11th century), which was established between the headwaters of the Niger and Sénégal rivers. Ghana was effectively destroyed by the Almoravid invasion of 1076, and its hegemony was ultimately assumed by the Malinke empire of Mali (12th–16th century), founded around the upper Niger. Under Mali the caravan routes moved east through Djenné and Timbuktu (founded about the 11th century AD). Mali's decline in the 15th century enabled the Songhai kingdom in the east to assert its independence. Under Songhai, Djenné and Timbuktu flourished as centres of both trade and Islāmic scholarship. In 1591 a Moroccan army of 4,000 men armed with muskets succeeded in crossing the Sahara and easily defeated the Songhai, who were armed only with bows and arrows and spears. With the destruction of Songhai hegemony, political chaos ensued, resulting in a disruption of trade. Eventually, new trade routes in gold and slaves were established, but these were toward the coast where Europeans were establishing trading posts. The Moroccans exiled or executed the Timbuktu scholars because they represented a political threat and dispersed most of their libraries of books and manuscripts. Moroccan military and political influence never extended beyond a short stretch of the Niger in the areas of Gao and Timbuktu. Eventually, political ties between Morocco and the descendants of the Moroccan invaders lapsed. In 1737 the Moroccans were defeated by the Tuareg, who seized control of the Niger Bend. To the west the Fulani (Peul) kingdom of Macina defeated the Moroccans at Diré in 1833. To the west of the Fulani a powerful animist kingdom had been established by the Bambara at Ségou beginning in the early 17th century.

The 19th century. Most of the 19th century was characterized by French colonial expansion from Senegal in the west and Islāmic jihads (religious wars) resulting in the establishment of theocratic states. Shehu Ahmadu Lobbo (Cheikou Amadou), a Fulani Muslim cleric, successfully overturned the ruling Fulani dynasty in Macina in 1810 and established a theocratic state whose capital was Hamdallahi. In the west, political events were dominated by al-Ḥājj ʿUmar Tal, a Tukulor Muslim cleric who led a series of jihads. ʿUmar conquered the Bambara kingdom of Ségou in 1861 and the Fulani empire of Macina in 1864. After ʿUmar was killed in a skirmish with the Fulani in 1864, his vast domains were divided among his sons and commanders. His eldest son, Amadou Tal, who had been installed at Ségou, unsuccessfully attempted to exert control over the whole Tukulor empire in a series of civil wars. He became head of the Ségou Tukulor empire, whose predominantly Bambara inhabitants led constant revolts against his rule.

The French, who established a fort at Médine in western Mali in 1855, viewed the Ségou Tukulor empire as the principal obstacle to their acquisition of the Niger River valley. Fearful of British designs on the same region, they engaged in a series of diplomatic overtures and military operations to push the boundaries of their control eastward. Between 1880 and 1881 the French succeeded in expanding their control from Médine 200 miles east to Kita, primarily through the diplomatic efforts of Captain Joseph-Simon Gallieni, who signed protectorate treaties with chiefs at Bafoulabé and Kita. In 1883 Gustave Borgnis-Desbordes launched a series of military campaigns against the Tukulor and the forces of Samory Touré, a Dyula Muslim leader who founded a state to the south. Desbordes captured Bamako that year, giving the French a presence on the Niger. Between 1890 and 1893, Colonel Louis Archinard launched a series of successful military operations that led to the final conquest of Ségou in 1893. Samory, who was driven into the Ivory Coast, was captured in 1898, the same year that the small Dyula kingdom of Kenedougou around Sikasso was conquered by the French under Colonel H.M. Audeod. Timbuktu was conquered in 1894 by the French officers Gaston Boiteaux, Eugène Bonnier, and Joseph-Jacques Joffre, and the southern Sahara was pacified by *méharistes* (camel corps) by 1899.

What is now Mali became a part of French West Africa. However, its borders were modified at different times and its name was changed as well. For most of its existence the territory was known as the French Sudan. The territory was headed by either a governor or lieutenant governor.

Independent Mali. Political parties were first formed in 1946, when a territorial assembly was established. The Sudanese Union-African Democratic Assembly (US-RDA) eventually became the dominant party under its charismatic Marxist leader, Modibo Keita. On Nov. 24, 1958, the territory became known as the Sudanese Republic and was an autonomous state within the French Community. On Jan. 10, 1959, Senegal and the Sudanese Republic joined to form the Mali federation under the presidency of Keita. Hopes that other Francophone states would join the union never materialized, and on Aug. 20, 1960, the federation broke up because of major policy differences between the Senegalese and Sudanese. On Sept. 22, 1960, a congress of the US-RDA proclaimed the independent country of the Republic of Mali.

Keita, the new country's first president, rapidly Africanized the civil service, distanced the country from France, established close diplomatic relations and economic ties with Communist bloc countries, and built a state-run economy. In 1962 Mali issued its own nonconvertible currency, although Keita entered into monetary negotiations with the French in 1967 to prop up a sagging economy. Keita, while claiming to be nonaligned, regularly supported the Communist bloc in international affairs. His radical socialist political and economic policies and a cultural revolution launched in 1967 led to widespread popular discontent, which created a favourable environment for a group of army officers to seize power. On Nov. 19, 1968, they launched a coup that overthrew Keita and his government. Led by Lieutenant Moussa Traoré, the officers formed a 14-member Military Committee of National Liberation that ruled Mali from 1969 to 1979, when a civilian government was elected. Personal rivalries and policy differences within the committee led to the removal of two officers in 1971. In 1978 Traoré accused four committee members who opposed a return to civilian rule of planning a coup. They were arrested and two, Lieutenant Colonel Kissima Doukara and Lieutenant Colonel Tiecoro Bagayoko, later died in prison.

The Traoré regime

In 1974, Malians overwhelmingly approved a new constitution. Under it the country returned to civilian rule in 1979, with a military-sponsored political party, the Mali People's Democratic Union (UDPM), in control of the government and President Traoré serving as head of state. Traoré was again elected president in 1979 and reelected in 1985. Eighty-two UDPM candidates have been elected to the National Assembly every three years. During the 1980s Traoré gave civilians access to the government through regular local and National Assembly elections and appointments to the Council of Ministers. Contested elections within the framework of the UDPM have provided the electorate with an opportunity for venting frustrations and disappointments by turning candidates out of office.

Cyclic droughts, a paucity of natural resources, inefficient parastatals (agencies serving the state but not officially under government control), and a bloated civil service have largely centred Mali's political focus on economic issues. During the early 1980s, student and teacher protests over stipends, salaries, and access to civil service employment challenged Traoré's rule. Traoré dealt effectively with these and with a number of coup attempts.

Traoré consistently followed a pragmatic foreign policy, maintaining close relations with both France and the communist bloc. During the 1980s he made concerted efforts to improve relations with other Western countries, including the United States. These were linked to his attempts to attract foreign investments, to diversify the economy, and to promote a private sector. Despite these attempts, Mali's political and cultural links to the communist bloc and the socialist economic structures put in place by Keita remain strong.

Mali has had two armed conflicts with Burkina Faso

over a border area, one in 1974–75 and another in 1985. The latter conflict took place in December 1985 and lasted five days; the territory in question was the Agacher Strip, a border region about 100 by 12 miles in size. The matter was referred to the International Court of Justice, which divided the territory in 1988 to the satisfaction of both parties.

For later developments in the history of Mali, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (P.J.I.)

Mauritania

The Islāmic Republic of Mauritania (al-Jumhūrīyah al-Islāmīyah al-Mūrītānīyah, or République Islamique de Mauritanie), a state in northwestern Africa, forms a geographic link between the North African Maghrib (a region that also includes Morocco, Algeria, and Tunisia) and the Senegal region of western Africa. Culturally it forms a transitional zone between the Arab-Berber region of North Africa and the region to the south of the tropic of Cancer known as the Sudan (a name derived from the Arabic *bilād as-sūdān,* "land of the blacks"). With an area of 398,000 square miles (1,030,700 square kilometres), Mauritania has the shape of an indented rectangle measuring about 930 miles (1,500 kilometres) from north to south and about 680 miles from east to west. It is bordered to the northwest by the Western Sahara (formerly Spanish Sahara), to the northeast by Algeria, to the east and southeast by Mali, and to the southwest by Senegal. Its Atlantic Ocean coastline, to the west, extends for 435 miles from the delta of the Sénégal River northward to the Cap Blanc Peninsula. The capital is Nouakchott.

Much of Mauritania forms part of the western Sahara, and a large proportion of the population is nomadic. The country's mineral wealth includes large reserves of iron ore, copper, and gypsum, which are now all being exploited. Mauritania, formerly French administered, became independent on Nov. 28, 1960. By the terms of the constitution, Islām is the official state religion, but the republic guarantees freedom of conscience and religious liberty to all. Arabic is the national language, and the official languages are Arabic and French.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief, drainage, and soils.* Both land relief and drainage are influenced by the aridity that characterizes the greater part of the country. The impression of immensity given by the landscape is reinforced by its flatness; the coastal plains are lower than 150 feet (45 metres), while the higher plains of the interior vary from 600 to 750 feet. The interior plains form a plateau of which the culminating heights, occurring at different levels, form many tablelands joined to one another by very long, gentle slopes of about 2°. The topography is relieved by vestiges of cliffs (generally cuestas); by sloping plains that terminate at one end of the slope with a steep cliff or faulted scarp, which may reach heights of 900 feet; or by inselbergs (steep-sided residual hills), of which the highest is Mount Ijill at 3,002 feet (915 metres), an enormous block of hematite.

Plains and plateaus

Structurally, Mauritania may be divided into three principal zones. The first of these, located in the north and northwest, consists of underlying Precambrian rock (about 2.7 billion years old), which emerges to form not only the backbone of northern Mauritania's Rigaibāt ridge region but also the Akjoujt rock series that forms a vast peneplain (a land surface worn down by erosion to a nearly flat plain) studded with inselbergs. The second zone is located partly in the extreme north but mostly in the centre and east. In the north it consists of primary sandstone, which covers the Tindouf Syncline (a fold in the rocks in which the strata dip inward from both sides toward the axis); in the centre is the vast synclinal basin of Taoudeni, bounded by the Adrar, Tagant, and 'Açâba plateaus. The basin is scarcely indented to the south by the Hodh Depression, with the Affollé Anticline (a fold in which the rock strata incline downward on both sides from a central axis) lying in its centre. The third zone is formed by the Senegalese-Mauritanian sedimentary basin, which includes coastal Mauritania and the lower Sénégal River valley of the southwest.

The Mauritanian landscape in general, as a result of the arid phases it underwent during the Quaternary period (*i.e.,* within the past 1.6 million years), presents three different aspects; these are represented by skeletal soils, regs (desert surfaces consisting of small, rounded, tightly packed pebbles), and dunes.

Skeletal soils are formed where outcrops of the underlying rock have been slightly weathered or where they have been covered with a patina or chalky crust. To these may be added the saline soils of the salt flats, formed from the caking of gypsum or of salt derived from the evaporation of former lakes. The regs form plains often of great extent, carpeted with pebbles and boulders. The dunes cover about 50 percent of the total area of the country. They are stretched out, often for several dozen miles, in long ridges known as *'alâb,* which are sometimes 300 feet high; they frequently overlap with one another, forming a network of domes and basins.

It is only to the south of the 10-inch isohyet (an imaginary line connecting points with equal rainfall) that the sands bear a brown type of soil. This soil is characteristic of the steppe (treeless plains) and contains 2 percent humus. It is also only in the extreme south of the country that the iron-bearing lateritic soils of the Sudanic zone begin; in the lowest places occur patches of hydromorphic soils (*i.e.,* soils altered by waterborne materials).

The drainage system is characterized by a lack of pattern. Normal drainage is limited to inland southwestern Mauritania, where tributaries of the Sénégal River, which forms the frontier between Mauritania and Senegal, flow southward and are subject to ephemeral flooding in summer. In the greater part of the country, however, the plateaus are cut into by wadis (dry riverbeds), where the rare floods that occur dissipate their waters into a few permanent drainage basins called *guelt* (singular *guelta*). In the wastes of the north and the east, rainfall is so rare and slight that there is practically no runoff.

Climate. The climate owes its aridity to the northeastern trade winds, which blow constantly in the north and throughout most of the year in the rest of the country; the drying effect produced by these winds is increased by the harmattan, or east wind. With the exception of the few winter rains that occur as a result of climatic disturbances originating in the mid-latitude regions, precipitation essentially results from the rain-bearing southwesterly winds, which progressively extend throughout the southern half of the country at the height of the summer. The duration of the rainy season, as well as the total annual amount of rainfall, diminishes progressively from south to north. Thus, Sélibaby in the extreme south receives about 25 inches (635 millimetres) between June and October; Kiffa, farther north, receives 14 inches between mid-June and mid-October; Tidjikdja receives seven inches between July and September; Atar receives seven inches between mid-July and September; and Nouadhibou receives between one and two inches, usually between September and November. Because of opposition between the wet southwesterlies and the harmattan, rains often take the form of stormy showers or squalls.

Aridity

The strength of the sun and the lack of haze in these latitudes result in high temperatures. In the summer months afternoon temperatures exceed 100° F (38° C) at most stations, and daily highs of 115° F (46° C) are not uncommon in the interior. The average temperature in the coldest month at most stations is in the region of 68° F (20° C), while the average temperature during the hottest month rises to about 75° F (24° C) at Nouakchott in September, to about 79° F (26° C) at Kiffa in May, to 81° F (27° C) at Atar in July, and to 84° F (29° C) at Néma in May.

Plant and animal life. Vegetation zones depend upon the degree of aridity, which increases from south to north. The Sudanic savanna, studded with baobab trees and palmyra palm trees, gradually gives way in the extreme south to a discontinuous belt of vegetation known as the Sahel (an Arabic word meaning riverbank, or shore, which

is also used to designate the southern Sahara borderlands). In the Sahel, trees are rare and the vegetation consists principally of acacias, euphorbia bushes (plants of the spurge family that have a milky juice and flowers with no petals or sepals), big tufts of morkba (*Panicum turgidum,* a type of millet), or fields of *cram-cram,* or Indian sandbur (*Cenchrus biflorus,* a prickly grass). Between the six-inch and four-inch isohyets the steppe rapidly disappears, giving way to desert. Vegetation is restricted to such places as the dry beds of wadis, beneath which water continues to flow, or to oases.

In the savanna, big antelope are hunted by lions, and, in the hilly Afollé region of southern Mauritania, herds of elephants are found. The steppe is frequented by gazelles, ostriches, warthogs, panthers, hyenas, and lynx; crocodiles are found in the *guelt.* Only addax antelope venture out into the waterless desert. Animal populations have been much reduced by hunting, obliging the authorities to introduce measures for conservation.

The Sahara and Sahel regions

Settlement patterns. The Sahara region to the north, where habitation is generally limited to oases, stands in contrast to the Sahelian steppelands to the south, where regular rainfall permits extensive stock raising and some agriculture. While transition occurs between one zone and another, a convenient line of demarcation is represented by the four-inch isohyet.

The heartland of Mauritania consists of the vast Adrar and Tagant plateaus, known as the Trab el-Hajra (Arabic: "Country of Stone"). There, at the foot of cliffs, are found several oases, among which some—such as Chinguetti, Ouadane, Tîchît, Tidjikdja, and Atar—were the sites of well-known cities in the Middle Ages. To the north and the west extend the vast desert peneplains. The exploitation of the Zouîrât mines and the development of Nouadhibou have transformed this once-abandoned region into the focus of Mauritania's economic life.

Coastal southwestern Mauritania (under colonial rule, the Trarza *cercle*) is covered with regularly aligned dunes and is important for stock raising; Arab-Berber (Moorish) culture is represented by important marabout families.

Inland, southwestern Mauritania is inhabited by Moors and Fulani (Peul), who engage in both livestock raising and agriculture. In the extreme south, large villages are surrounded by fields of millet, constituting the first sign of the Sudanese landscape.

In the southeast the vast Hodh Basin, with its dunes, sandstone plateaus, and immense regs, is a major livestock-raising region, the economy of which has many links with neighbouring Mali.

Mauritania is a country suited for nomadic life. Livestock supplies the nomads with milk and meat, while transport is provided by riding camels and pack camels and, in the south, by pack oxen and donkeys as well. The women dye sheep's wool, with which they then braid long brown bands that are sewn together to make tents; they also tan goats' skins to make *guerbas* (waterskins). Movement is governed by the search for water and pasturage. In the Sahara nomadic movements are irregular because of the extreme variability of rains; in the Sahel, however, regular seasonal movement occurs—to the south in the dry season and back to the north in the wet season. Sizes of nomadic encampments also vary from south to north. In the coastal southwest, agglomerations of up to 300 tents may be found, whereas in northern Mauritania only groups of a few tents are to be seen.

The decline of nomadism

For a variety of reasons—including changes in agricultural patterns as well as in political and commercial relations—the traditional nomadic way of life has been declining. Dams to conserve floodwaters have been built in the wadis, and palm tree culture has been considerably extended. The cumulative result of these developments has been that the nomads now tend to remain longer in the south near their millet fields and palm groves—becoming, in effect, seminomads.

Since the 1960s a movement toward settlement has been evident, largely because of growing dissatisfaction with the harsh conditions of nomadic life; this movement, however, has been restricted by the constant lack of water. Only in the extreme south—on the banks of the Sénégal River (in Mauritania the region called the Chemama)—is a normal, settled agricultural life possible. A series of droughts halted cultivation in the Sahel in the 1970s, and livestock raising was seriously constrained, affecting the livelihood of nomads and sedentary peoples alike.

The exploitation of the iron ore reserves of Mount Ijill, as well as the achievement of independence, have transformed the urban geography of Mauritania. The ancient cities that lived by caravan traffic—*e.g.,* Tîchît, Chinguetti, Ouadane, and Oualâta—and traded with Casablanca and, above all, with Dakar have grown idle beneath their palm trees. Only Tidjikdja and Atar have maintained a certain activity. Kaédi, on the Sénégal River, has expanded and is still growing. Three new towns have been built: Nouakchott, the capital; Fdérik (formerly Fort-Gouraud); and Nouadhibou (formerly Port-Étienne).

Nouakchott was founded in 1958. Formerly, Mauritania shared a capital, Saint-Louis, in Senegal, with the other members of the French West African Federation. Nouakchott is located near the sea and provides access to the Sahara and Sahel regions alike. In addition to being the seat of government and administration, Nouakchott is a growing commercial centre and has a printing works, a hospital, and numerous schools. Its water supply is provided by a plant that desalinizes seawater.

Fdérik is an administrative centre situated about 15 miles from the mining town of Zouîrât. The port of Nouadhibou, which was long stagnant, owes its more recent expansion both to expanding iron ore exports and to the establishment of a fishing complex there. A few miles to the south lies Cansado, a residential suburb.

The people. *Ethnic and linguistic groups.* The Moors constitute more than two-thirds of the population; about half of them are white, or *bīdān,* Moors of Arab and Berber descent, and about half are black Moors, of Sudanic origin. Moorish society historically was divided into a hierarchy of castes. At the head of the socioeconomic structure were the noble castes, composed of *'arab*s, or warriors, and *Murābit* (marabouts), or priests and scholars of the Qur'ān. The warriors were usually Arab, and the marabouts were usually Berber. The mass of the *bīdān* population were vassals who received protection from the warriors or marabouts in return for tribute. There were two artisan classes—the blacksmiths and the griots (who were at once musicians and genealogists). Servant classes were formed of black Moors and were subdivided into *'abid,* or slaves, and *hartani,* or freedmen. Among the ethnic and racial groups, blacks became the better educated and held most technical, professional, and diplomatic posts at the time of independence. Members of this "servant" caste, which developed as the bureaucratic class, became increasingly aware of their rights as citizens. Slavery was abolished by the French before independence and was officially abolished again on July 5, 1980, but subsequent reports claimed that the practice had continued.

The Moors

The Moors speak Ḥassānīyah, a dialect that draws most of its grammar from Arabic and uses a vocabulary of both Arabic and Berber words. Most of the members of the aristocratic castes also know literary Arabic.

The remaining population, generally referred to as *kewrin,* consists of Tukulor (Toucouleur), who live in the Sénégal River valley; Fulani, who are dispersed throughout the south; Soninke (Sarakole), who inhabit the extreme south; and Wolof (Oulof), who live in the vicinity of Rosso in coastal southwestern Mauritania. The Tukulor and the Fulani speak Fulfulde (Poular), and the other ethnic groups have retained their respective languages.

Of Mauritania's total population an estimated one-fourth are nomads, and about one-third live in and around urban centres. Because of the country's large desert area, the average density is the lowest in western Africa. Three-fourths of Mauritanians live in the Sénégal River valley. Life expectancies stand at 44 and 47 years respectively for men and women.

Most of the non-Africans in Mauritania are French nationals engaged in technical assistance, commerce, and mining; Spaniards represent the second largest foreign community.

Religious groups. About 99 percent of all Mauritanians

are Muslim. Most Moors belong to the Qādirīyah order. The Tukulor and some of the Tagant tribes belong to the Tijānīyah order. Many tariqas (mystical sects) flow from these orders.

The economy. In the Sahel region of Mauritania a traditional subsistence economy is maintained, composed of livestock raising, agriculture, crafts, and petty trading. In the Sahara region, however, a modern economy is developing, based on the exploitation of iron-ore and copper resources and of the ichthyologically-rich continental shelf; the modern economy receives much needed capital investment and technical assistance from abroad. More than three-quarters of the Mauritanian population still lives by traditional activities, among which livestock raising is the most important. In numbers, goats and sheep are the most important livestock, followed by cattle, camels, donkeys, and horses. Cattle are raised primarily in the southern region, whereas goats and sheep are dispersed as far north as the limits of the Sahara. Camels are raised mostly in the north and the centre, especially in the Adrar region. The growth of the Mauritanian economy slowed in the 1980s after a lengthy period of rapid expansion in the 1960s and '70s. Agriculture and fishing account for almost one-third of the gross national product, with the industrial sector, including mining, contributing about one-quarter, public administration about 15 percent, and the remaining sectors about 30 percent.

Livestock raising

The state imposes indirect taxes on imports, a turnover tax on exports and mining, a service tax, and taxes on cattle, vehicles, wages and salaries, and profits from industrial and commercial concerns.

Mauritania's budget, usually in deficit, was nominally balanced in the late 1980s. In the mid-1980s, principal and interest on a relatively large foreign indebtedness was rescheduled, but indebtedness remains a significant problem. Foreign aid, both bilateral (from France, Japan, China, the United States, and the Arab states) and by multilateral agencies (such as the African Development Bank, International Monetary Fund, International Bank for Reconstruction and Development, and the European Economic Community), is primarily targeted to assist in project development but is also used for budgetary and food support.

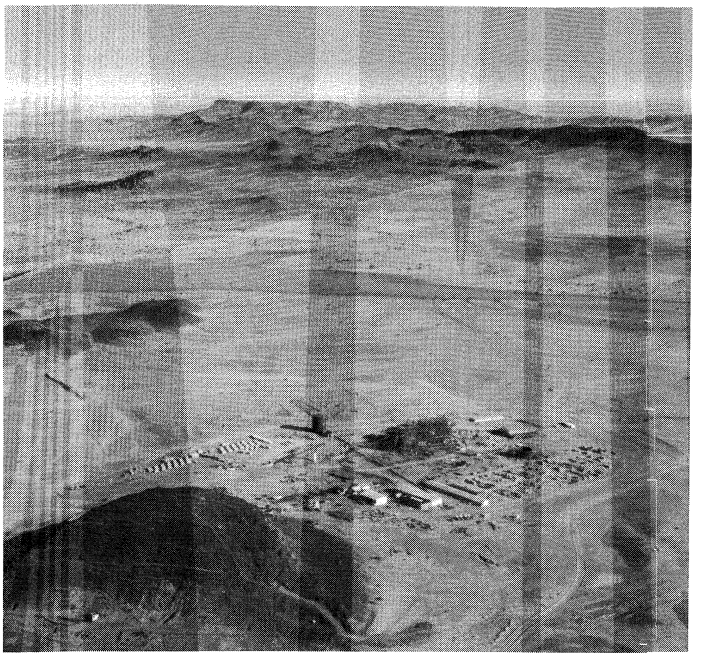
Agriculture and fishing. Agriculture is necessarily dependent upon rainfall. Where the rainfall exceeds 17 inches a year, millet (fonio) and dates are the principal crops, supplemented by sorghum, beans, yams, corn (maize), and cotton. Seasonal agriculture is practiced on the easily flooded riverbanks and in the wadis of the Sahelian zone, upstream from the dams. There, too, millet, sorghum, beans, rice, and watermelons are grown. Irrigated agriculture is practiced in areas supplied by water-control projects and at oases, where well water is available; corn, barley, and some millet and vegetables are grown. The output of gum arabic is less than it was in former years. Agricultural production in Mauritania has continued to decline because of drought. Crop production fell by approximately two-thirds in the period from 1970 to 1980. From the late 1970s Mauritania was unable to produce more than half of its total food requirements.

In agriculture the aim of successive Mauritanian governments has been to increase the amount of irrigated land in the Sénégal River valley and, above all, to increase the production of rice, of which Mauritania is still obliged to import large quantities, to plant fresh palm trees to replace those destroyed by the cochineal insect, to drill fresh wells, to improve the quality of dates, and to encourage the cultivation of vegetables.

Rich fishing grounds lie off Mauritania's Lévrier Bay. Mauritania stopped issuing fishing licenses in 1979, however, and in 1980 formed joint companies with Portugal, Iraq, South Korea, Romania, and the Soviet Union. In 1987, fisheries agreements were signed with the European Communities. At Nouadhibou fish are canned, frozen, or processed as fish flour. Several tons each year are dried and exported to other African countries.

Mining. A rail link connects the mining town of Zouîrât with the port of Nouadhibou, the only deepwater roadstead on the Saharan coastline, accommodating ships of

Iron ore and copper



Iron ore mine at Zouîrât, Mauritania, with the Adrar plateau in the background.
Shostal Associates/Superstock

up to 150,000 tons. Iron exploitation was organized and begun in 1963 by Miferma (of which 56 percent of the financing was by French groups and the remainder by British, West German, and Italian groups and by the Mauritanian government). The company was nationalized in 1974 and was renamed Cominor (Complexe Minier du Nord). The iron ore deposits of Mount Ijill have nearly been exhausted, and exploitation of reserves at Guelbs began in 1984. Iron exports fell from a peak of 12 million tons in 1974 to an annual average of nine million tons in the 1980s.

The copper deposits of Akjoujt are extensive, with a copper content of more than 2 percent. Exploitation was begun in 1969 by Somima (Société Minière de Mauritanie), of which 54 percent of the shares were held by British and U.S. interests, 25 percent by the Mauritanian government, and the remainder by French interests. Somima was nationalized in 1975, but operations were suspended in 1978. Reactivation of the mine remains a constant preoccupation, as does the extraction of gold from the copper ore. Processed copper had been exported along a highway to the wharf at Nouakchott. There are substantial gypsum deposits near Nouakchott; most of the annual production is exported to Senegal. Other mineral resources are minor, and salt output has declined. Reserves of ilmenite (the principal ore of titanium) have been located, and phosphate deposits have been identified near Bofal in the south. Oil prospecting has so far yielded no results.

Finance and trade. The national Banque Centrale de Mauritanie was established in 1973. In addition there are five commercial banks. There are no securities exchanges in Mauritania, and the government controls all insurance business.

Foreign trade is difficult to estimate because, while imports and exports of the modern sector are well known, there are no statistics for the traditional sector. Mauritania is nevertheless known to import from or by way of Senegal quantities of millet, tea, rice, sugar, cotton goods, and hardware, while it exports to Mali and to Senegal cattle, sheep, and goats. Iron ore together with fish and fish products constitute the major exports.

Transportation. Transport by pack animals—camels in the north, oxen and donkeys in the south—has retained considerable importance in a society in which a subsistence and barter economy prevails, although transport between cities and regions is increasingly by truck. Considerable hazards, however, confront road builders; among these are moving sand dunes, flash floods in the south,

and steep cliffs. Only the main road running from Rosso via Nouakchott, Akjoujt, Atar, Fdérik, and Bir Mogreïn to Tindouf, Algeria, is passable throughout the year. The road between Rosso and Nouakchott is surfaced, as is the section between Nouakchott and Akjoujt. The Trans-Mauritania highway, linking Nouakchott, Kaédi, Kiffa, 'Ayoûn el-'Atroûs, and Néma, was completed in 1982. The railroad from Zouîrât to Nouadhibou is used only for transporting iron ore. There are international airports at Nouakchott and Nouadhibou, and other cities are linked by regular air services.

The irregularity of the flow of the Sénégal River limits its use as a waterway; Kaédi can only be reached by ships drawing about seven feet at the high-water season, normally from August to October. The port facilities at Nouakchott can accommodate 320,000 tons of shipping a year. Nouadhibou, in addition to being an iron ore and fishing port, is also a commercial port.

Administration and social conditions. *Government.* The Mauritanian state had a presidential regime from 1960 until 1978, when a coup d'état installed a military government. A civilian government established in December 1980 was replaced the following April by a largely military administration.

The primary task of Mauritania's successive governments has been to transform a community of diverse tribes, hierarchical in social structure and very strongly differentiated, into a nation. Many of the local barriers to cooperation have been overcome, and traditional regional boundaries have been redrawn. There are 12 administrative regions, each directed by a governor, with the capital forming a separate district. The regions are subdivided into 49 departments.

Islāmic law and jurisprudence have been in force since February 1980. Qadis (judges of the Sharī'ah, or Islāmic law) in rural and town communities hear matrimonial cases. Nouakchott is divided into six regional sections where magistrates and judges hear cases in lower courts; there are also labour and military courts, the Court of State Security, and a Supreme Court, which deals with administrative as well as judicial matters.

Education. Less than one-third of Mauritania's adult population is literate. Schooling is compulsory to age 14, but only a small minority of children benefit from it. In addition to primary schools in the urban centres, there are secondary schools, a university with faculties of letters and human sciences and of law and economics, and a research institute for mining and industry in Nouakchott. The capital is also the site of a national library. There are traditional local libraries in some urban centres.

Health and welfare. Modern health facilities are scarce in Mauritania, with only one major hospital in Nouakchott and 25 other regional health centres, of which 15 are maternity clinics. Free medical services are available to the poor. Traditional medicine flourishes. In addition to the usual health problems that are associated with tropical areas, tuberculosis and intestinal and eye maladies are widespread. Disability payments for industrial accidents and occupational diseases, family allowances, and old-age pensions are administered by the National Social Insurance Fund.

Cultural life. Moorish society is proud of its Arab and Muslim heritage. Theology, poetry, and music flourish. Goldsmithing is a fine art. Kewri societies have a rich and varied folklore.

Mail, telephone, and telegraph services are combined in the main post offices. Fewer than 6,000 telephones are in use, and administrative contact is primarily through radiotelephone. International telephonic communications are run through Paris. The media are owned or controlled by the government. A national radio network broadcasts in the prevailing languages and in French. There are also two Earth satellite stations with telecasts in French and Arabic. A daily, *Ach-chaab;* a fortnightly, *Journal Officiel;* and a bimonthly, *Le Peuple,* are published in the two official languages. Movie theatres are found in the main urban centres and *cercles* (social, sporting clubs) provide recreational opportunities in Nouakchott, Nouadhibou, and Rosso.

For statistical data on the land and people of Mauritania, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (C.H.T./A.G.G.)

HISTORY

Numerous discoveries of Acheulian and Neolithic remains have been made in the north.

The Berber Almoravids

Mauritania was first peopled by Negroes and by the Sanhadja Berbers. It was the cradle of the Berber Almoravid movement, which in the 11th century laid waste the Ghana empire before conquering Morocco. The Almoravids imposed Islām upon all the neighbouring peoples. A caravan route at that time linked Mauritania with Morocco. Arab tribes infiltrated by this route and in the 15th century submerged the Berbers. The nomadic tribes formed several powerful confederations: Trarza and Brakna, which dominated the Sénégal River valley; Kunta in the east; and Rigaibāt (Regeibat) in the north.

In 1442, Portuguese vessels rounded Cape Blanc, and in 1448 the Portuguese founded the fort of Arguin, whence they derived gold, gum arabic, and slaves. Later the French and English frequented Portendick, and then the French settled at Saint-Louis at the mouth of the Sénégal River. In 1858 Colonel Louis Faidherbe ended Moorish domination over lower Sénégal. After 1898 an Islāmic specialist, Xavier Coppolani, succeeded, without a struggle, in rallying all the Moors of the south to France. He was assassinated in 1905, and his work was completed by Colonel Henri Gouraud, who occupied Tagant in 1907 and Adrar in 1909. The Rigaibāt were not finally pacified until 1955.

Mauritania was constituted a territory of French West Africa in 1920 and later became a colony; it was at first governed from Saint-Louis in Senegal. In 1946 Mauritania became an overseas territory and in 1957, after repulsing an attack by Moroccan irregulars on the north, elected a government under Moktar Ould Daddah, who established the new capital at Nouakchott. In 1958 Mauritania voted to become a member state of the French Community, and on Nov. 28, 1960, its full independence was declared. It became a member of the United Nations in October 1961. (H.J.D.)

Relations with Morocco

The small political elite was divided over whether the country should be oriented more toward Senegal and black French-speaking Africa or toward Morocco (which sought to absorb it). The winning faction, under Sidi el-Moktar N'Diaye and his successor, Ould Daddah, chose independence with close ties to France and full participation in the Organization of African Unity. Reversing Moroccan policy, King Hassan II recognized Mauritanian independence in 1969 as part of his plan to gain control of Spanish Sahara. Mauritania joined the Arab League in 1973.

As president since independence, Ould Daddah appeared securely established in spite of occasional strikes by miners and demonstrations by students, for his policies seemed attuned to a largely tribal population, 90 percent of whom were engaged in agriculture or pastoralism. The difficulties of suppressing guerrillas of the Polisario Front in Mauritania's portion of the Western Sahara (the former Spanish Sahara), which had been divided between Morocco and Mauritania in 1976, contributed to his downfall. On July 10, 1978, he was deposed and exiled in a military coup led by the chief of staff, Colonel Mustapha Ould Salek. (N.B./L.C.Br./A.G.G.)

Ould Salek resigned his position in June 1979 and was replaced by Lieutenant Colonel Mohamed Mahmoud Ould Louly. He was in turn replaced in January 1980 by the prime minister, Lieutenant Colonel Mohamed Khouna Ould Haidalla. In December 1984 Colonel Maaouya Ould Sidi Ahmed Taya took over the presidency and the office of prime minister from Haidalla in a bloodless coup.

The Mauritanian effort to disentangle itself from the Western Sahara worsened relations with Morocco. After signing a peace treaty with the Polisario Front in August 1979, Mauritania renewed diplomatic ties with Morocco in 1985 and sought a solution to the dispute in the Western Sahara.

For later developments in the history of Mauritania, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (Ed.)

Niger

The Republic of Niger (République du Niger), a landlocked western African country, takes its name from the Niger River, which flows through the southwestern part of its territory. The name Niger derives in turn from the phrase *gher n-gheren,* meaning "river among rivers," in the Tamashek language. The republic has an area of 458,075 square miles (1,186,408 square kilometres). It is bounded on the northwest by Algeria, on the northeast by Libya, on the east by Chad, on the south by Nigeria and Benin, and on the west by Burkina Faso and Mali. The capital is Niamey.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Niger extends for about 750 miles (1,200 kilometres) from north to south and about 930 miles from east to west. It tends to monotony in its features, is intersected by numerous depressions, and is dominated by arid highlands in the north. Rainfall increases as one proceeds southward so that the country divides naturally into three distinct zones—a desert zone in the north; an intermediate zone, where nomadic pastoralists raise cattle, in the centre; and a cultivated zone in the south. It is in this southern zone that the greater part of the population, both nomadic and settled, is concentrated.

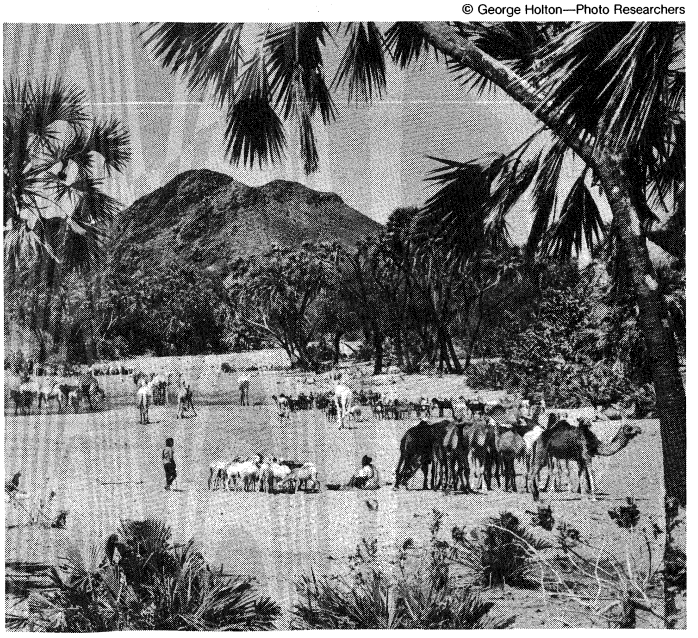
Three geographic zones

The highlands of the north are cut by valleys (*kori*) of the Aïr Massif, which is an extension of the Ahaggar (Hoggar) Mountains of Algeria, and consists of a range running north to south in the centre of Niger, with individual mountain masses forming separate "islands": from north to south these are Tazerzaït, where Mount Gréboun reaches an altitude of 6,379 feet (1,944 metres); Tamgak; Takolokouzet; Angornakouer; Bagzane; and Tarouadji. To the northeast is a series of high plateaus, which form a bridge between the Ahaggar Mountains of Algeria and the Tibesti Mountains of Chad. From west to east these are the plateaus of Djado, Mangueni, and Tchigaï.

The sandy regions of the Nigerian Sahara extend to either side of the Aïr. To the west the Talak region includes the Tamesna area in the north (where fossil valleys are filled with moving sand dunes) and the Azaoua area in the south. East of the Aïr is the Ténéré region, covered partly by an expanse of sand called an erg, partly by a stony plain called a reg.

The plateaus of the south, which form a belt about 900 miles long, may be divided into three regions. To the west is the Djerma Ganda region. Its large valleys are filled with sand, while *dallol* (fossilized valleys of rivers that formed tributaries of the Niger in ancient times) descend from the Aïr and the Iforas Massif of neighbouring Mali. The central region consists of the rocky Adar Doutchi and Majia areas; it is the region of the *gulbi* (dried-up valleys of former tributaries of the Sokoto River) and the Tegama—a tableland of sandstone, ending, toward the Aïr, at the Tiguidit scarp. To the east the underlying rock reappears in the Damagarim, Mounio, and Koutous regions, to the north of which is the region of Damergou, consisting of clays. In the Manga region, in the east, traces of ancient watercourses appear on the sandy plain.

© George Holton—Photo Researchers



Tuareg with their herds at an oasis in the Aïr Massif, northern Niger.

Drainage and soils. It is convenient to make a distinction between the ancient hydrographic system, which allowed agriculturalists, fishermen, and pastoralists to live in the Aïr region about 5,000 or 6,000 years ago, and the present simple system, which forms the basis of the marked difference between the northern and southern parts of the country. The present system includes to the west the Niger River basin and to the east the basin of Lake Chad; between the two occur vestiges of the older system, such as the *dallol* and the *gulbi.*

To the west the Niger River crosses about 350 miles of Niger's territory. Because of the change in river flow, which occurs because of the dispersal of its waters in its interior delta region in Mali, it is only in January and February that it flows past Niamey in flood. At other times the river is fed by certain temporary watercourses that flow in from the right bank. These are the Gorouol, the Dargol, the Sirba, the Goroubi, the Djamangou, the Tapoa, and the Mékrou; the last two flow through the "W" National Park (so called because the Niger flows through the area in the form of a W). On the left bank, proceeding eastward, appear the *dallol,* the vestiges of the older watercourses. Generally running from north to south, they constitute zones of dampness, although a few still contain waters that flow toward the Niger. The best known are the Bosso, the Foga, and the Maouri wadis. Other vestiges consist of the *kori,* which run down from the Aïr and from former tributaries that had their sources in the Iforas Massif, and which flowed to a confluence at what is now the Ti-m-merhsoï Wadi. No waters flow through the *kori* now, but water is still to be found beneath their sands. Other remnants of the old system are formed by the *gulbi,* through which water still flows annually, occasionally causing damage.

Ancient watercourses

To the east is situated the basin of Lake Chad, a large, shallow lake, which at its highest contemporary level has an area of approximately 9,650 square miles; of this, Niger possesses about 1,100 square miles. Its extent is considerably reduced during the dry season. The Komadougou Yobé River, which flows into Lake Chad from the west, forms part of the frontier between Niger and Nigeria. Its water level, which begins to rise in August, from January to May consists only of some stagnant pools.

In addition to the drainage system described, it may be noted that rainwater collects in several basins, so that some permanent lakes or pools also exist; these are found at Keïta and Adouna in the Adar Doutchi region, at Madaroumfa in the Maradi *gulbi,* and at Guidimouni to the east of Zinder. The water table in some areas can also be tapped to produce artesian wells.

The soils fall into three natural regions. In the Saharan region in the north the soil is infertile, except in a few oases where water is found. In the region known as the Sahel, which forms a transitional zone between the Sahara and the region to the south, the soils are thin and white, being covered with salty deposits resulting from intense evaporation that forms an infertile surface crust. The third region (in the south) is cultivated. In this area the soils are associated with extensive dunes or uplands or with basins or depressions. Some of the soils in the latter, such as those in the Niger basin and in the *gulbi,* are rich. Black soils occur in the Kolo basin. Throughout the region, however, and above all on the plateaus, less fertile lateritic (leached iron-bearing) soils occur.

Climate. Niger extends southward from the tropic of Cancer, and the northern two-thirds of its territory lies in dry tropical desert. In the southern part of the country the climate is of the type known as Sahelian, which is characterized by a single, short rainy season. In January

and February the continental equivalent of the northeast trade winds, the harmattan, blows southwestward from the Sahara toward the equator. Typically dust-laden, dry, and desiccating, the harmattan hinders normal living conditions on the southern fringe of the desert. From April to May the southern trade winds blowing from the Atlantic reach the equator and are diverted toward the Sahara where they meet with the harmattan—an encounter that results in violent line squalls and that signals the beginning of the rainy season. The rains last from one to four months, according to the latitude; August is the rainy month everywhere except in the far north, where the rainfall is unpredictable.

Temperature and rainfall

Niger lies in one of the hottest regions of the world. Temperatures rise from February to May and drop during the "winter" rainy season, rising again somewhat before falling to their annual minimum averages in December or January. During May (the hottest month), afternoon temperatures are high everywhere, ranging from a low of about 108° F (42° C) at Nguigmi on Lake Chad to 113° F (45° C) at Bilma and Agadez, both in the northern desert. In January, afternoon temperatures average more than 90° F (mid-30s C) at most stations but at night may drop to freezing level in the desert. The daily range is greater in the desert north than in the south and is also more extreme during the dry season.

Rainfall varies according to location as well as season. The 10-inch isohyet (line on a map connecting points having equal rainfall) follows a line from near Tahoua to Gouré, in effect marking the northern limit of nomadic pastoral life, for the rainfall permits a sparse vegetation to grow. To the extreme south the 30-inch isohyet marks the southern limit of this zone, after which the southern agricultural zone begins. In the course of the same rainy season a most irregular spatial pattern of rainfall may occur, while from one year to another the total amount of rainfall may also vary; in addition, the rainy season itself may arrive early or late, thus jeopardizing crops.

Plant and animal life. The vegetation of the desert zone clusters around the oases; it includes the date palm and cultivated corn (maize). Animal life, which must be able to endure hunger and thirst, includes the dromedary.

In the Sahel zone, where the doum palm and the *cram-cram* (*Cenchrus biflorus,* a prickly grass) appear, the vegetation has a short life cycle and is principally used for grazing. Animal life includes the ostrich and the gazelle.

In the cultivated zone the vegetation includes acacia trees, doum palms, and palmyra palms, as well as baobabs. Wildlife, which has partially disappeared, includes antelope, elephants, and warthogs; giraffes are found in the Zarmaganda and Damergou regions, and hippopotamuses and crocodiles on the banks of the Niger. The extreme southwest is a savanna region where baobabs, kapok trees, and tamarind trees occur. Animal life is preserved in the "W" National Park, where antelope, lions, buffalo, hippopotamuses, and elephants may be seen.

Settlement patterns. The southern part of Niger's territory is situated in the vast region of Africa known as the Sudan, in which, in former times, large political states arose, such as Ghana, Mali, and Songhai, as well as the Hausa states, the empire of Sokoto, and Bornu. The northern part of Niger remains the domain of the Tuareg. The country comprises a multitude of traditional regions, the names of which remain despite the establishment of contemporary administrative divisions. All these regions have a fluctuating political, economic, and geographic significance: the Hausa regions, for example, have been cut in two and divided between Niger and Nigeria. Most regions, moreover, have been and remain zones where contact takes place between different peoples—between the Hausa and the Tuareg in the Adar Doutchi region, for example; between the Tuareg and the Kanuri in the Damergou region; and between Hausa and Zerma in the Aréoua area.

About one-fifth of the population live in towns. The rural population comprises nomads and sedentary peoples. There are some 10,000 villages, of which approximately half have only a few hundred inhabitants; there are practically no villages in the desert zone. Fulani (Peul) herdsmen, who breed horned cattle and oxen, and the Tuareg, who raise goats, sheep, and dromedaries, tend to travel over the northern region during the winter. They meet together to permit the cattle to lick the salty soil of the In Gall region during August and September but move southward during the dry season. Both Fulani and Tuareg live in tribal groups, in temporary or portable shelters, and gain their subsistence from their livestock. The Fulani subsist above all on milk in various forms; the Tuareg live on meat and dates.

Nomadic peoples

Sedentary peoples, such as the Hausa, the Songhai-Zerma, and the Kanuri, who inhabit the Niger and Chad basins, live largely by agriculture. They raise millet, rice, corn, peanuts (groundnuts), and cotton. They also work as blacksmiths and shoemakers, while on the banks of Lake Chad and the Niger the Buduma and Sorko peoples are fishermen. Sedentary peoples live in dwellings that vary from those made of straw to those made of banco (hardened mud), although the Wogo people live in tents of delicate matting.

There is a tendency among the nomads to settle down, and the already sedentary peoples are expanding the lands under cultivation toward the north. Rural life, above all in its sedentary form, tends to slow its pace during the long dry season; it is at this time of year that migration to the towns or other countries occurs.

It was approximately in the 15th century that a few towns, such as Agadez or Zinder, were first established as halting places, or depots, on the trans-Saharan caravan routes. As commercial routes gradually developed on the coasts, however, these northern towns lost their former economic importance, while other centres, such as Birni Nkonni, and Tessaoua, declined in the course of the 19th century as a result of the colonial era.

Population centres

There are four principal towns in Niger. Niamey, the political capital, has experienced rapid growth. It has a cosmopolitan character and a transient population. Its characteristic life varies between the European and African rural styles, including various intermediate steps, of which the life-style of the *évolués* (educated Africans) is the most distinctive. Zinder, for which the African name is Damagaram, is an older town than Niamey; a Hausa town, it was the capital of Niger until 1926 and has a number of skilled craftsmen, especially leatherworkers and dyers. The town has experienced some industrial growth and has close links with Nigeria. Maradi has developed rapidly. The town is situated in the heart of the peanut-growing region near the Nigerian frontier. Many European companies have established branches there; the town is particularly renowned for its red goats, the skins of which are exported to Europe and the Americas. Tahoua has grown up on the edge of the desert. There it forms a large livestock market, where nomad pastoralists and sedentary farmers meet. All of the towns remain little more than modest administrative and commercial centres, but because of the discovery of uranium ore Agadez has experienced a spectacular growth.

The people. *Linguistic groups.* The largest linguistic group is formed by the Hausa, whose language, also spoken in Nigeria, is one of the most important in western Africa. A large percentage of the inhabitants of Niger understand Hausa, which possesses an abundant literature that has been printed in Latin characters in Nigeria. Songhai is the second most important language; it is also spoken in Mali, in northern Burkina Faso, and in northern Benin. In Niger itself it is divided into various dialects, such as Songhai proper, Zerma, and Dendi. The language of the Fulani is Fulfulde; in Niger it has two dialects, eastern and western, the demarcation line between them running through the Boboye district. Tamashek is the language of the Tuareg, who often call themselves the Kel Tamagheq, or Tamashek speakers. The language is also spoken in Algeria and Mali and possesses its own writing, called *tifinagh,* which is in widespread use. Kanuri is spoken not only in Niger but also in Cameroon and Nigeria; the tongue is called Beriberi by the Hausa. While these five languages are the principal ones spoken in Niger, there is also an important Teda linguistic group in the Tibesti region. In addition, many of the peoples of Niger speak Arabic, and a still larger number read and write in

that language; Agadez possesses one of the oldest Arabic schools in Africa. The use of the Arabic alphabet resulted in Fulfulde and Hausa becoming written languages; the script is called *ajami;* a search for more old manuscripts in *ajami* is being conducted.

By using Hausa and Songhai, one may make oneself understood from one end of the country to the other. French, however, remains the official language, as well as the language of instruction, although it remains understood only by a small minority. English is taught as the principal foreign language in secondary schools.

Ethnic groups. Ethnic groups correspond to the five linguistic groups already mentioned. The Hausa are the largest group, constituting more than half of the present population, though the majority of the Hausa people live in Nigeria. The Hausa occupy the centre of southern Niger as far as Dogondoutchi. The Songhai-Zerma are found in the southwest; the Songhai proper live along the Niger, where they are assimilating the Kurtey and Wogo peoples. The majority of the Songhai people as a whole, however, live in Mali. The Zerma (Djerma) live on the left bank of the Niger, remaining in close contact with the Mauri and Arewa peoples. The Fulani, who are dispersed throughout the country, are mostly nomadic; they are also found dispersed throughout western Africa. The Tuareg, also nomadic, are divided into three subgroups—the Iullemmiden of the Azaouak region in the west, the Asben (Kel Aïr) in the Aïr region, and the Itesen (Kel Geres) to the south and east of Aïr. The Tuareg people are also found in Algeria and in Mali. The Kanuri, who live to the east of Zinder, are divided into a number of subgroups—the Manga, the Dogara (Dagara), the Mober, the Buduma, and the Kanembu; they are also found living in Chad, Cameroon, and Nigeria. Apart from the nomadic Teda of the Tibesti region, who constitute an important minority, the remainder of the population consists of Arabs, black Africans from other countries, and Europeans, of whom the greater part are French.

Religious groups. More than 95 percent of the population adhere to the Sunnite branch of Islām. Although the Annaawaa group of Hausa have always refused to accept Islām, as have a group of Fulani, the Wodaabe—who distinguish themselves from other Fulani for this reason—Islām remains the religion of the majority of both Hausa and Fulani. Christianity (Roman Catholicism and Protestantism) remains a religion of the towns, particularly of Niamey. There are several Christian missions in the Songhai and Arewa areas. Christianity is primarily a European religion, although it is also practiced by some black Africans from other countries. The traditional animist religions of the black Africans continue to manifest themselves in strength.

The economy. The economic system is based upon planning but accords an important role to private enterprise. The three main policy objectives are the maintenance of national unity, the elevation of the living standards of the population, and the attainment of economic independence. The private sector of the economy consists partly of a multitude of small enterprises and partly of enterprises belonging to large French or international companies. The government, through the agency of the Development Bank of the Republic of Niger, which is funded partly by aid from abroad, has promoted the establishment of many companies, including real estate, road transport, air transport, and agricultural processing enterprises.

The private sector of the economy

Niger is encouraging economic links between African countries. Apart from its membership in the Organization of African Unity, Niger is a member—together with Côte d'Ivoire, Benin, Burkina Faso, and Togo—of the Conseil de l'Entente, a regional cooperative group, as well as of the Organisation Commune Africaine et Mauricienne, another group of French-speaking African states.

Resources. Salt is traditionally exploited in the Kaouar and Aïr regions, as well as in the *dallol,* and in the Manga district. Natron (hydrated sodium carbonate) is extracted locally. Cassiterite (an ore of tin) is mined at open workings in Aïr. Small quantities of gold are obtained by panning in the Sirba River. Limestone and an important deposit of gypsum have been located at Malbaza and in the Ader Doutchi and Majia region. Niger's known reserves of uranium rank among the most important in the world. Apart from tungsten in the Aïr region, traces of copper, lignite (a brownish black coal), molybdenum, zinc, phosphates, and titanium have been found and are the subject of further prospection. A reserve of iron ore, with an iron content of about 50 percent, has been located in the Say region; and petroleum deposits have been discovered in the Lake Chad area.

The exploitation of plant resources has long been practiced but on a small scale. The doum palm and the palmyra palm provide wood for construction, while the palms of the Manga oasis produce dates. Small amounts of kapok (a silky down from the kapok tree, used for insulation, life jackets, and so forth) and of gum from the acacia gum tree are exported. Skins of ostriches, crocodiles, and snakes are used for making handicrafts that are exported to Europe. Fish from the Niger River and Lake Chad are exported southward to the coastal countries.

Agriculture. Agriculture and agricultural products constitute the largest sector of Niger's economy in terms of the number of persons employed and the percentage of gross national product (GNP). Millet and sorghum, the main food crops, are grown in the south, as are cassava and sugarcane. Rice is grown in the Niger River valley. Peanuts are the most important cash crop; other important crops include cotton and pulses.

Livestock is an important sector of the agricultural economy and is a major export. Cattle, sheep, and goats are raised for meat, milk, and hides.

Niger's ability to remain self-sufficient in food and livestock production is closely linked to rainfall, and periods of drought have resulted in shortfalls requiring imports and food aid. To increase production and avoid cereal shortfalls, the government has invested in irrigation projects and an "off-season growing program" of small-scale production and irrigation operations.

Industry. Niger is one of the world's leading producers of uranium, which is mined at Arlit, Akouta, and Tassa. Extraction at the Arlit site is undertaken by the French-controlled Société des Mines de l'Aïr (SOMAIR). The second major mining concern, the Compagnie Minière d'Akouta (COMINAK), is owned partly by the government of Niger and partly by foreign interests. The Tassa mine opened in 1986 and is operated by SOMAIR. The uranium industry was seriously affected by the fall in uranium prices in the early 1980s. Development of additional sites is dependent upon an increase in world uranium prices.

Some manufacturing industries have been established, mostly at Niamey. They produce chemicals, food products, textiles, farm equipment, and metal furniture. There are many small craft industries in the principal towns.

Imported petroleum, supplemented by locally mined coal, is used to generate about half of Niger's electricity, and the remaining amount is imported from Nigeria. The Office of Solar Energy has produced solar batteries, which are used in the country's telecommunications network, and peanut shells have been experimentally used to supplement hydrocarbon fuels since 1968. Wood is the traditional domestic fuel.

Transportation. While the economically active zone of Niger runs from east to west across the southern part of the country, the principal lines of communication run southward toward the coast. The two ports used by Niger—Cotonou in Benin and Lagos in Nigeria—are each more than 600 miles away, and Niger possesses no railroad. Traditional systems of transport and communication are still largely relied upon. These include camel caravans in the northern Sahel region, canoes on Lake Chad and the Niger, and individual travel on horseback or on foot. Only a small tonnage of goods is transported.

Trucks maintain transport communications between Maradi and Zinder in Niger and Kano in Nigeria, and between Niamey and Parakou in Benin. A road completed in 1981 connects the uranium-producing centres of Arlit and Akouta to Nigerian transport links. The principal west–east road axis enters the country from Gao in Mali, runs on the banks of the Niger as far as Niamey, and then continues eastward to Nguigmi on Lake Chad. From this

Roads

central route, roads branch off southward. Toward the north, routes running via Tahoua and Tânout converge near Agadez, linking Niger to Algeria via Tamanrasset.

Air Niger is responsible for domestic air services linking the country's airports, including those of Tahoua, Maradi, Zinder, Agadez, Diffa, and Arlit. Niamey has an international airport.

Administration and social conditions. *Government.* The constitution promulgated in November 1960, which established a presidential type of regime, was suspended after the military coup of April 1974. Executive power is held by a Supreme Military Council composed of army officers. The president of the Supreme Military Council appoints a Council of Ministers. A charter for civilian rule was drafted in 1986.

Following the 1974 coup the Supreme Court was replaced by the Court of State Security. At lower levels are the Court of Appeal, district courts, and justices of the peace. There are separate labour and assize courts.

Niger is divided into eight *départements*—Agadez, Diffa, Dosso, Maradi, Niamey, Tahoua, Tillabéry, and Zinder—each of which is administered by a prefect. Each *département* is divided into several *arrondissements,* each of which is administered by a subprefect.

Education. Education in Niger is free, but only a small proportion of children attend school. Primary and secondary schools and teacher-training colleges are the responsibility of the Ministry of National Education. Other ministries are responsible for technical education. Niger has one of the lowest adult literacy rates in western Africa, and literacy programs are conducted in the five principal African languages. Niamey has a university, and the Islāmic University of Niger at Say opened in 1987.

Health and welfare. The general state of health in the country is poor, and health care facilities are inadequate. The rural population has only limited access to any medical treatment. The infant mortality rate, at about 150 per 1,000 live births, is one of the highest in western Africa. Health services concentrate on the eradication of certain diseases in rural areas, as well as on health education. Campaigns have been successfully waged against sleeping sickness and meningitis, while vaccinations against smallpox and measles are also administered. Other diseases, however, such as tuberculosis, malaria, and leprosy, remain endemic. Antituberculosis centres are located at Niamey, Zinder, and Tahoua. Lack of finances and shortage of trained personnel remain the principal obstacles to the improvement of health conditions.

Cultural life. Niger forms part of the vast Sahelian cultural region of western Africa. Although the influence of Islām is predominant, pre-Islāmic cultural traditions are also strong and omnipresent. Since independence, greater interest has been shown in the country's cultural heritage, particularly with respect to traditional architecture, handicrafts, dances, and music. With the assistance of the United Nations Educational, Scientific and Cultural Organization, a regional centre for the collection of oral traditions has been established at Niamey. An institution prominent in cultural life is the National Museum at Niamey.

For statistical data on the land and people of Niger, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (Di.L./Ed.)

HISTORY

One of the central themes of the history of Niger is the interaction between the Tuareg (and also Tubu) nomads of the vast Saharan north and the agriculturalists of the sedentary south; the interaction, that is, between opposed yet complementary ways of life and civilizations. Among the agriculturalists the main ethnic groups are the Songhai-Zerma in the west, the Hausa in the centre, and the Kanuri in the east. The Hausa have always been the most numerous. They constitute nearly half of the total population of Niger, whereas the Songhai-Zerma constitute only about a quarter.

The kingdom of Takedda

In the 14th century (possibly also earlier and later) the Tuareg-controlled kingdom of Takedda, west of the Aïr Massif, played a prominent role in long-distance trade, notably due to the importance of its copper mines. Copper was then used as a currency throughout western Africa. Archaeological evidence attests to the existence of communities of agriculturalists, probably Songhai-speaking, in this region, which is now desert, at the time of the kingdom of Takedda. Takedda was succeeded at an unknown date by the (probably) much more loosely organized sultanate of Agadez, centred on Aïr.

For many centuries the southeastern third of present-day Niger constituted one of the most important provinces of the Kanuri empire of Bornu. The might of Bornu was based on the control of a number of salt-producing sites and of long-distance trade, notably along the string of oases between Lake Chad and the Fezzan via Kawar.

The great drought of about 1735–56—the prelude to the present dry cycle, which set in about 1880—had an adverse effect upon the natural environment. This may explain why both the communities of agriculturalists west of Aïr and the oases between Lake Chad and Kawar disappeared. It may perhaps also explain in part why the Tuareg were able to extend their control over a fair portion of the sedentary south.

By the time of the colonial conquest, the disparate regions the French molded into an entity known as Niger may be best described as an assemblage of peripheral borderlands. As borderlands, however, these regions had played a significant role as zones of refuge—the west after 1591 and the Moroccan conquest of the Songhai empire and the Hausa region much later, after the 1804 Fulani jihad in central Hausaland (*i.e.,* present-day northern Nigeria). In both cases the refugees were people who had lost in the military as well as the religious struggles of their respective homelands. Thus both regions became bastions of "traditionalism" in the face of partly alien conquerors attempting to impose Islām.

French conquest

The French conquest began in earnest only in 1899. It nearly met with disaster due to the local population's determined resistance against the notorious expedition in 1899 led by the French officers Captain Paul Voulet and Captain Charles Chanoine. It was only in 1922, after the severe drought and famine of 1913–15 and the Tuareg uprising of 1916–17, that the French felt safe enough to establish a regular administration under civilian control. By then the power of the Tuareg had been decisively and definitively broken. Niger had thus become a basically sedentary country.

As elsewhere, the peace in French West Africa (*pax gallica*) meant among other things the rapid spread of Islām, a steep demographic increase, and, although exclusively among the Hausa, the extension of cash crop cultivation. The Songhai-Zerma, on the other hand, responded to the French tax demands by engaging themselves as seasonal labourers in the coastal regions.

Through the reforms of 1946, France's African subjects in theory were granted full citizenship. Thus Niger, along with the other colonies (renamed "overseas territories") of black Africa, was represented in the French Parliament. Consultative-legislative assemblies were also set up locally. These reforms secured the ascent of a tiny new elite, the so-called *évolués—i.e.,* those who had been trained in French schools. Many were descendants of former slaves, and most were Songhai-Zerma. Indeed, the people of the west had proved to be far more open to European influence than, for instance, the Hausa.

At least until 1954–55 the French administration (headed for 12 years by Governor Jean Toby) remained firmly in control of the political situation. The first local executive was established in 1957. Its head, the left-wing trade unionist Djibo Bakary, advocated a no vote in the referendum of 1958. However, 72 percent of the votes cast were in favour of a continued link with France.

Independence

Nevertheless, under Bakary's successor, his cousin and fellow Songhai-Zerma Hamani Diori, independence was proclaimed on Aug. 3, 1960. Diori, who set up a single-party dictatorship, was toppled in a coup in 1974. There followed a military dictatorship headed first by Seyni Kountché (until his death in 1987) and then by Ali Seibou.

The peanut boom in the 1950s, the relatively favourable climatic conditions between the late 1930s and the 1970s, and the uranium boom in the 1970s raised some hopes for

the future. These hopes have been largely thwarted, however, notably by a succession of droughts and unfavourable international developments. Thus Niger remains an underdeveloped, drought-ridden, vulnerable, and dependent country, which is increasingly faced with a problem of overpopulation.

For later developments in the history of Niger, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (F.F.)

Senegal

Situated at the western extremity of Africa's tropical zone, the Republic of Senegal (République du Sénégal) has an area of 75,955 square miles (196,722 square kilometres). It is bounded to the north and northeast by the Sénégal River, which separates it from Mauritania; to the east by Mali; to the south by Guinea-Bissau and Guinea; and to the west by the Atlantic Ocean. The Gambia constitutes a finger of territory 20 miles (32 kilometres) wide and 200 miles long that thrusts from the coast eastward into Senegal along the Gambia River.

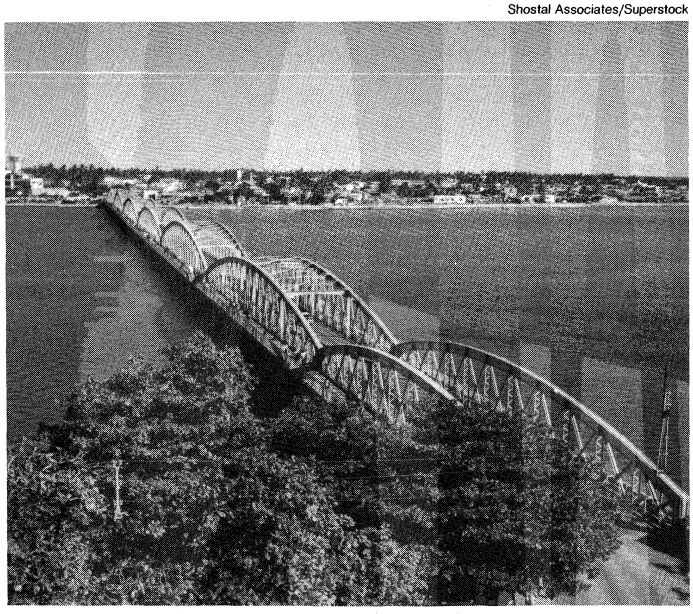
Senegal—which gained its independence in 1960, first as part of the short-lived Mali Federation and then as a sovereign state in its own right—is among the principal producers of peanuts (groundnuts); its light soils and its climate are well suited to this crop. Food crops such as millet and sorghum also are important. Although the economy is planned, the moderate controls to which it is subjected are applied in a flexible rather than an authoritarian manner. Private investors, whether foreign or Senegalese nationals, are encouraged to establish new enterprises; an investment code grants tax exemptions and permits the withdrawal of profits. Nationalization has in general been avoided, and the denationalization of some state companies is proceeding. Economic policy favours private initiative even in the production and marketing of the peanut, which is so essential to Senegal. The economic life of Senegal is characterized by its membership in the Franc Zone, as a result of which the country benefits from French financial support. French is the official language.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Senegal is a flat country, lying in the depression known as the Senegal-Mauritanian Basin. Altitudes of more than about 330 feet (100 metres) are found only on the Cape Verde Peninsula and in the southeast of the country. The country as a whole falls into three structural divisions. These are: first, the Cape Verde headland, which forms the western extremity and consists of a grouping of small plateaus made of hard rock of volcanic origin; second, the southeastern and the eastern parts of the country that consist of the fringes of ancient massifs (mountain masses) contiguous with those buttressing the massif of Fouta Djallon on the Guinea frontier, with the highest point reaching an altitude of 1,640 feet; third, an immense but shallow basin lying between Cape Verde to the west and the edges of the massif to the east.

Structural divisions

Shostal Associates/Superstock



Faidherbe Bridge, spanning the mouth of the Sénégal River, at Saint-Louis.

Washed by the Canary Current, the Atlantic coast of Senegal is sandy and surf-beaten. Like the rest of the country, it is low except for the Cape Verde Peninsula, which represents the westernmost point of the African continent and which shelters Dakar, one of the finest ports in Africa. To the south of the peninsula, the surf on the coast is less heavy. To the south of the Saloum River mouth, the coast consists of rias (drowned valleys) and is increasingly fringed with mangroves.

Drainage. The country is drained by the Sénégal, Saloum, Gambia (Gambie), and Casamance rivers, all of which are subjected to a climatic regime characterized by a dry season and a rainy season. Of these rivers, the Sénégal, which rises in the Fouta Djallon highlands of Guinea and was long the main route providing access to the interior, is the most important. After traversing the old massifs the river rapidly drops downward before reaching Senegalese territory. At Dagana it forms the so-called False Delta, which supplies Guiers Lake on the left bank. At the head of the delta is the town of Richard-Toll (the Garden of Richard), named for a 19th-century French nursery gardener. The slope of the land is so gentle on this stretch of the river that at low water salty seawater flows about 125 miles upstream. The island on which the town of Saint-Louis stands, at the mouth of the river, is situated about 300 yards from the sea in the False Delta, whose true mouth lies 10 miles to the south.

The False Delta of the Sénégal

Soils. Despite its apparent uniformity, Senegal contains a great diversity of soils. These fall generally into two types—the valley soils and those found elsewhere.

The soils of the Sénégal and Saloum river valleys, in their middle courses, are alluvial and consist of sandy loams or clays. Near the river mouths the soils are salty and favourable for grazing. Similar conditions are associated with the Gambia and Casamance rivers, except that near their mouths the banks are muddy, while their upper courses have sandy clay soils.

Many different types of soils are found in the various regions. In the northwest the soils are ochre-coloured and light, consisting of sands combined with iron oxide. These soils, called "Dior soils," constitute the wealth of Senegal; the dunes they form are highly favourable to peanut cultivation, while the soils between the dunes are suitable for other food crops, such as sorghum. In the southwest the plateau soils are sandy clays, frequently laterized (leached into red residual iron-bearing soils). In the centre and the south the country is covered by a layer of laterite hidden under a thin covering of sand. These soils afford only sparse grazing during the rainy season. In Casamance heavily leached clay soils with a high iron oxide content predominate. Whether they are deep, as in western Casamance, or shallow, as in the southeast of the region, they are suitable for cultivation.

Climate. Senegal's climate is conditioned by two major factors: first, the tropical latitude of the country and, second, the seasonal migration of the intertropical convergence zone (ITCZ)—the line, or front, of low pressure at which hot, dry continental air meets moist oceanic air and produces heavy rainfall. The prevailing winds, characterized also by their area of origin, fall into two categories—those that are dry, originating in the continental interior, and the moist maritime winds that bring the rains.

The dry winds consist of the northeast trade winds, known in the winter and spring when they are strongest as the harmattan; they bring no rains at all, apart from a very light precipitation which, to the Wolof people of Senegal, is known as the "Heug." The rain-bearing winds blow primarily from the west and northwest. Beginning in June with the northward passage of the ITCZ these winds usher in the summer monsoon. As the ITCZ returns southward beginning in September, the rainy season draws to a close. The slow north–south migration of the ITCZ results in

a longer, heavier rainy season in the southern part of the country.

From the combination of these various factors, three principal climates may be distinguished, each of them associated with a characteristic type of vegetation.

The coastal (Canarian) climate occurs along a coastal strip about 10 miles wide running from Saint-Louis to Dakar. Its winters are cool, with minimum temperatures reaching about 63° F (17° C) in January and maximum temperatures in May not exceeding 81° F (27° C). The rains begin in June, reach their height in August, and cease in October. The average rainfall is about 20 inches (500 millimetres).

The Sahelian climate occurs in a zone bounded to the north by the Sénégal River and to the south by a line running from Thiès (a town on Cape Verde Peninsula) to Kayes in the neighbouring country of Mali. The month of January is cool, especially in the mornings before sunrise, when the temperature drops to about 57° F (14° C); afternoon temperatures, however, may rise to higher than 95° F (35° C). In May, minimum temperatures do not fall below about 72° F (22° C), while maximums often rise above 104° F (40° C). The dry season is quite distinct and lasts from November to May. Certain places, such as Podor and Matam, are particularly noted for their dryness and heat. Between July and October the rainfall averages about 14 inches, moderating the temperature somewhat. Maximum temperatures at this season reach about 95° F (35° C).

The Sudanic climate occurs in the remainder of Senegal. Regional nuances are in evidence. Thus, from north to south three climatic subdivisions may be recognized, each of which is characterized by the amount of average annual rainfall. First, in the Kaolack–Tambacounda subdivision, annual rainfall averages between 29 and 39 inches, occurring on about 60 days between June and October. Cultivation without irrigation is possible in this region. Second, in the Gambian region, the rainfall frequently amounts to 50 inches, resulting in the growth of a continuous belt of light forest and patches of herbaceous undergrowth. Third, in the Casamance region, rainfall everywhere exceeds 50 inches, falling on 90 days of the year. The forest is dense and green and is continuous, without undergrowth. Oil palms, mangroves, and rice fields are characteristic of this climate zone. The Sudanic climate in general is very hot, humid, and uncomfortable. The town of Kaolack, for example, has average afternoon temperatures near 100° F (38° C) throughout the year and is rendered yet more oppressive by the salt wind.

Rainfall in the Sudanic area

Animal life. While it is true that large mammals have disappeared from the western part of the country, owing to human settlement, such animals as elephants, antelopes, lions, panthers, cheetahs, and jackals may still be encountered in the interior. Herds of warthogs abound in the marshes, especially those of the False Delta. Hares are ubiquitous, and monkeys of all types congregate in noisy bands, above all in the upper Gambia and upper Casamance river valleys. Among the great numbers of birds the quelea, or "millet eater," which is destructive of crops, may particularly be noted, as well as the partridge and the guinea fowl. Reptiles are numerous and include pythons and cobras and other venomous snakes. Crocodiles, hippopotamuses, and turtles are found in the rivers. The rivers and the coastal waters are rich in fish and crustaceans.

Settlement patterns. While such physical factors as geology, soil, climate, and vegetation have resulted in regional differentiation, man has also been a determining factor in the delimitation of different regions, each marked by a traditional type of human settlement. Thus from north to south five principal traditional regions may be distinguished.

The Ferlo region is the central region of Senegal; very extensive, it is distinguished by its semidesert aspect and by the poverty of its soils. Vegetation appears only in the south, the north consisting of the Sahelian type of savanna parkland (an intermediate zone between the Sahara and the savanna proper); it affords light grazing for the flocks of nomadic Fulani (Peul) pastoralists.

Fouta-Toro is based upon the Sénégal River, extending approximately from Bakel in the east to Dagana in the north and consisting of a strip of territory that is relatively densely inhabited. Cultivated lands, which are watered by the river and its tributaries in the dry season, are of importance thanks to highly developed agricultural and pastoral use of the soils and vegetation. Most of the region is inhabited by Tukulor people. Fouta-Toro is bounded to the west by the False Delta, also known as the Oualo, which is peopled by the Wolof, who cultivate millet and carry on stock raising, employing Fulani shepherds.

The Dianbour, Cayor, Djolof, and Baol region is a diverse area situated between the Ferlo region and the Atlantic and extending from the Oualo in the north to the Cape Verde Peninsula in the south. The soils are sandy, the winters cool, and peanuts are the primary crop. The population is as diverse as the region itself and includes Wolof in the north, Serer in the Thiès region, and Lebu on Cape Verde.

The Sudan region is bounded by Cape Verde to the northwest, the Ferlo to the north, and the lower Casamance Valley to the southwest. It is composed of the following subregions—the "Little Coast," Sine-Saloum, Rip, Yacine, Niani, Boundou, Fouladou, and the valleys of the Gambia and upper Casamance rivers. In general, the region benefits from ample rainfall, which becomes abundant toward the south. The clayey soil is suitable for agriculture, despite the lateritic crust that appears intermittently. The region, as a result, is relatively densely populated. The estuaries are muddy and salty, with marshy saline depressions known as *tannes* occurring occasionally. The region as a whole is inhabited by a very diverse population composed of all the ethnic groups living in Senegal; the majority are, however, Malinke (Mandingo).

The lower Casamance region is small but strongly characterized. It is covered by dense vegetation of the Guinean type. Mangroves, oil palms, and raffia palms predominate. The rainy climate favours the cultivation of rice, which has long been a specialty in this part of the country.

Rural settlement

The great majority of the Senegalese population lives in the countryside. There are numerous villages, each with an average population of a few hundred persons. Usually each village has a shaded public gathering place, a mosque, and a water source, whether a well, a spring, or a small stream. The village is administered by a chief who is either traditionally nominated or appointed by the government. Religious life is directed either by a person literate in Arabic, called a marabout, or by a leading sorcerer. Various types of village may be distinguished, according to ethnic characteristics.

Whether it is situated in the western Ferlo or the Cayor regions, the Wolof village is small, being inhabited by about a hundred farmers. The houses are built of locally obtained materials. Each village may easily be moved from place to place, as the topography provides no natural obstacles to this. Harvests are kept in straw granaries, located far from the compounds for fear of fires. In the eastern Saloum region, the Wolof village is surrounded by three concentric zones of vegetation. The first of these—the inner zone—consists of fields and vegetable gardens and is known as the Tol-keur (literally "kitchen garden"). The second circle consists of land that has been exhausted, except for peanut cultivation, and it is known as the Diatte. The third, the farthest from the village, is the Gor, in which cereal crops are cultivated.

The typical village of the Sudanic region of Casamance consists of a Malinke agglomeration; it is a heritage from the epoch when the Sudanic peoples conquered the region. Each village has between 200 and 300 inhabitants living in enclosed compounds and crowded together in geometrically aligned rectangular huts. Agriculture and stock raising are the principal activities. The chief of each village is generally a marabout, conservative in his ways.

The Serer village differs from the Wolof and Malinke village because of its family compounds, called M'Binds, being loosely dispersed; each M'Bind is autonomous. On the islands at the mouth of the Saloum River, the houses of the Serer Nyiominka people are solidly built and trim. The granary is located in the compound.

Diola villages are substantial rural agglomerations with populations of up to 5,000 people or more. One of the characteristics of this type of village is that it is usually built on the edge of a plateau, or on ground that overlooks the rice fields with which Diola life is associated. As in the Serer villages, the compounds are not grouped in any distinguishable hierarchy. The houses are the best built and the most permanent among the different types of village dwellings to be found. On occasion they constitute veritable fortifications, for example, in the Thionck-Essyl and Oussouye regions; the villages of the Essyl region are often equipped with a rainwater catchment. The Diola and Serer villages have no chiefs with authority or prestige comparable to those of the Wolof or Malinke villages.

Major towns

The town of Saint-Louis was founded in 1633, and Dakar in 1857; other towns were, however, more recently founded. All the towns are of colonial origin. Dakar, as was formerly Saint-Louis, is the political and administrative capital. The other towns usually owe their foundation to the peanut trade, for which they were collection points that later developed into urban centres. These towns were often stops along the railroad lines, as at Thiès, Tivaouane, Mékhé, and Louga (between Dakar and Saint-Louis), or Khombole, Bambey, Diourbel, Gossas, Kaffrine, and Koungheul (between Thiès and Kayes, Mali). Certain ports also became towns: among these are Kaolack, Foundiougne, and Fatick (on the Sine-Saloum rivers), and Ziguinchor, Sédhiou, and Kolda (on the Casamance River). With the exception of Dakar, Saint-Louis, Rufisque, Thiès, Kaolack, and Ziguinchor, these towns have not as yet emerged, as it were, from their rural cocoons. Furthermore, all of the towns—including Saint-Louis, Rufisque, and Gorée, all of which had great importance in the past—are today dependent upon Dakar, in which about one-fifth of the Senegalese live.

Apart from the division of the countryside into traditional regions, one may observe that the best lands in Senegal are concentrated in the west and in the river valleys. The remainder of the land becomes increasingly poor and less settled as one continues toward the north or the east.

The people. *Linguistic composition.* The scientific study of languages in Senegal has not progressed far enough for even a rough type of classification to be attempted. Specialists nevertheless recognize certain imprecise groupings. These are: (a) the Atlantic (West Atlantic) group, including Wolof, Lebu, Serer, Tenda, and Diola; (b) Fulfulde, the Fulani (Peul) language, which also shares some of the characteristics of the Atlantic group; Fulfulde possesses numerous linguistic particularities, and has a complex grammar; and (c) the Mande group, including Bambara, Dyula, Malinke, and Soninke (Sarakole).

Ethnic composition. There are seven major ethnic and religious groups, and a number of other less significant groups. The major groups are located in the Sahel and savanna regions which formerly supported the ancient empires of the western Sudan, such as Ghana, Mali, and Songhai. Until recently the societies composing this grouping were strictly hierarchical in organization, consisting of the princely caste, the nobility, the freemen, the lower castes, and finally the slaves.

The Wolof represent about one-third of the total population. Their language is the most widely used in the republic. The Wolof predominate in the sandy western region. In the Cayor district they are initiates of the Tijānī Muslim brotherhood; the other brotherhood, that of the Murīdīyah, is very influential, and its expansion toward the southern part of the country is concurrent with that of peanut cultivation. Members of the Murīdīyah brotherhood, strong adherents of Islām, are primarily agriculturists.

The Serer are densely settled in the western part of the southern Ferlo region. They are experienced farmers, practicing both cultivation and cattle raising. Originally animist by religion, they are now becoming increasingly either Muslim or Roman Catholic.

Also known as the Peuls, Foulah, Fulbe, and Fellata, the Fulani are distributed throughout Senegal; they are particularly found in the Ferlo, the Upper Casamance, and Oualo regions, where their settlements are substantial. Characteristically nomadic pastoralists, many of them have become settled agriculturists, above all in the Fouta-Toro region and on the Senegal–Guinea border. They are Muslim.

The Tukulor (or Toucouleur) are often hard to distinguish from either the Wolof or the Fulani, with both of whom they have often intermarried. The name Tukulor is a distortion of the name of the ancient realm of Tekrur. The Tukulor live primarily in the middle course of the Sénégal River Valley. They are also found in dispersed groups living on the Gambia and Saloum rivers. The Tukulor were the first Senegalese people to become Muslim, having accepted Islām probably in the 11th century; many are literate in Arabic. Primarily farmers, they are increasingly migrating to the towns, particularly to Dakar and Saint-Louis.

The Diola occupy the lower Casamance Valley and the southwest of the Gambia Valley. They are skilled farmers, specializing in rice growing, but turning to the cultivation of peanuts and millet as the distance from the sea increases. In the Fogni district they are Muslim, but the majority remain animist. A few have accepted Christianity.

The Malinke came originally from the Niger River Valley and have spread out into various regions of Senegal, especially into the Gambia, Upper Casamance, and Saloum river valleys. Farmers and energetic traders, they are Muslim.

The Soninke are a minority group of Berber descent. They represent an extension into Senegal of the Malinke families of Mali. They are in the process of abandoning an unfruitful agricultural terrain in order to migrate toward the towns, where they often become small traders. The Soninke are Muslim.

The numerically less significant Senegalese comprise such peoples as the Mauri, who live especially in the north of the country where they are stock raisers or traders; the Lebu of Cape Verde, who are fishermen and often wealthy landowners; and the Basari, an ancient people who are found in the rocky highlands of Fouta Djallon.

The economy. Economic life is characterized by two factors. The first is the division of the country into two regions—the western region, which is wealthy and dynamic, and the remainder, and larger part, of the country, which remains poor and economically stagnant, depending upon a subsistence economy. The second factor is the existence of a single crop economy, which leads to partial unemployment, an insufficient income, and a dependence on an unpredictable climate and the international market.

Before independence, the economy was virtually entirely in the hands of the private sector. Since the economy depended primarily on the peanut trade, the large French companies that marketed the peanuts also controlled the importation of European manufactured goods. After independence, however, the Senegalese government created a state agency responsible for virtually all aspects of the peanut trade; in consequence, while the private sector remained important in the economy as a whole, it received its principal impulse from the state. An investment code is composed of various guarantees and long-term tax concessions, as a result of which capital investment has been attracted from many quarters.

The public sector is of primary importance in a country in which, for historic reasons, a middle class in the Western sense has never existed. The intervention of the state, moreover, is not a recent phenomenon; already in existence during the colonial era, it has been given a new form since independence by the creation of the National Organization of the Rural Sector. Apart from buying and selling peanuts, rice, and millet, the organization also sells fertilizer, seed, tools, and equipment and is thus the primary instrument used by the state in giving form to its policy of "African socialism."

The National Organization of the Rural Sector

In the 1980s the government began to move away from continued state participation in much of the Senegalese economy and to encourage the reintroduction of private initiatives. As part of this general trend the peanut industry has increasingly involved private organizations made up of the producers themselves. The state intervenes less and less.

As already mentioned, the government of Senegal encourages investment by granting tax benefits. Most governmental revenue is obtained from indirect taxes, which take the form of local taxes on alcoholic beverages, gasoline, tobacco, firearms, automobiles, and commerce. Direct taxation consists of land taxes, professional licenses, and personal taxes, such as taxes on profits and income taxes.

Since the late 1970s the demographic explosion, uncontrolled movements to the city, and the declining prices for primary materials have created a downturn in economic activity. Only substantial foreign aid has prevented a decline in the standard of living. At this stage the future of Senegal is unclear. The diversification of cultures participating in the reorganization of opportunities in West Africa will certainly give a better economic prospect for the Senegalese, however.

Resources. The economy is essentially agricultural and is primarily based upon the peanut crop. The known mineral resources are only of minor importance and consist primarily of phosphates of lime, located at Taïba near Tivaouane, about 60 miles northeast of Dakar, and aluminum phosphates at Palo Dial, near Thiès. Significant mineral reserves include petroleum deposits discovered off the Casamance coast and high-grade iron-ore reserves located in the upper Falémé Valley. The saltworks of Kaolack have a considerable potential production. Production of gum arabic, which is obtained from acacia trees, is only of secondary significance; and other forest products also have limited commercial value. The herbaceous vegetation nevertheless permits a relatively important amount of stock raising. By improving the grazing land available, Senegal has the potential to increase the numbers of its cattle herds to a considerable extent.

The waters off Senegal, particularly those at some distance from the shore, are rich in economically significant schools of fish, although the coastal waters are also known for their large variety of fishes; in this respect Senegal is better endowed than most other tropical countries on the Atlantic seaboard.

Agriculture and fishing. Agriculture occupies at least three-quarters of the economically active population and provides the basis for industry as well. Although a certain balance between the raising of livestock and peanut cultivation is maintained, it is the peanut production which earns the foreign exchange that the country needs.

Importance of the peanut crop

The main economic problem affecting national agriculture is the country's excessive dependence upon the peanut crop. This situation has prevailed in spite of efforts at diversification made over the past 30 years to develop the production of cotton, rice, corn (maize), sugar, and tomatoes. Each year the sale of the peanut crop results in much economic activity throughout Senegal; the resulting wealth has favoured the development of the smaller towns. The traffic in peanuts has also resulted in the establishment of the river ports of Kaolack, Foundiougne, Fatick, Sédhiou, and Kolda. Much of the activity at Dakar itself is due to the peanut trade.

Apart from peanuts, a number of food crops are also grown. Extensive acreage is devoted to sorghum and pennisetum (a genus of Old World grasses). Rice is cultivated both in naturally wet areas and by irrigation. Its large-scale cultivation is restricted to the lower Casamance Valley and the lower Sénégal River Valley below Richard-Toll. In addition corn, cassava (manioc), beans, and sweet potatoes are grown in significant quantities.

The climate and the savanna type of vegetation encourage the raising of livestock—including cattle, goats, sheep, horses, donkeys, camels, and pigs—which is carried on in almost all regions but is especially characteristic of the north. Stock raising is not a major source of income for the farmer, however; the meat is consumed locally, and only the hides and skins are exported.

While many fish are obtained from the rivers, the greater part of the catch is obtained from the sea. Fishing products have come to head all exports (producing almost double the income from peanut products in the late 1980s). This is the result of many years of efforts to build up the fishing industry.

Industry. Industrial production in Senegal is more developed than in most western African countries. Both processing and handicraft industries are established. Most of the processing industry is located in the Cape Verde region, where there are peanut-oil processing plants. Plants are also in operation in Dakar, Rufisque, Kaolack, Diourbel, and Ziguinchor. In addition, there are fish canneries, a shoe factory, and a cement manufacturing plant. Both the shoe factory and the cement plant are located at Rufisque. Other industrial establishments, all of which are located in Dakar, include flour mills, a textile plant, a sugar refinery, a tobacco factory, and a brewery, in addition to a naval shipyard, chemical plants, and an automobile assembly plant.

A number of craftsmen are engaged in traditional handicrafts; the more skilled among them are established at Dakar and Saint-Louis.

Of secondary significance in the national economy, mineral production consists of lime phosphates and aluminum phosphates.

Electric energy is produced by two joint public-private companies, in which public ownership dominates. They are the Senegal Electrical Society, owner of six steam-generating stations (two in Dakar and one each in Kaolack, Saint-Louis, Ziguinchor, and Tambacounda), and the National Society of Electric Energy Distribution (SENELEC).

Before the 1980s all energy produced in Senegal was of thermal origin. Cheaper hydroelectric energy began to be produced with the construction of the three-nation hydroelectric projects on the Sénégal River, with dams at Diama in Senegal (completed in 1985) and Manantali in Mali (completed in 1988). The forecast for electric production from the Manantali project alone is about 800 million kilowatt-hours per year.

Role of Senegal as a West African financial centre

Finance and trade. In finance, Senegal has benefited from the fact that in colonial times it was the principal territory of the administrative grouping known as French West Africa. Goods entering the country from the Franc Zone are liable to a simple fiscal tax; those from the Common Market countries enjoy a preferential tariff. There are a considerable number of banks offering financial services. Currency is issued by the Central Bank of West African States, which is the agency of the West African Monetary Union to which Senegal belongs. Other financial institutions include the Building Society of Cape Verde; the National Development Bank of Senegal; the Senegalese Banking Union; and private banks.

The value of imports is usually greater than the value of exports. The principal imports are petroleum products, rice, sugar, machinery, and vehicles. The principal exports are peanuts and peanut products, phosphates, fresh fish, and canned fish.

Peanuts remain the leading agricultural product of Senegal, and this makes the country, in good years, the leading producer of peanut oil in French-speaking black Africa. Nevertheless, the world market for this product is decreasing. The economic disengagement policy of the government has led to the creation of cooperatives selling directly to local oil producers. Peanuts in the shell are no longer exported as they were at the time of the colonial trade.

Transportation. The development of a transport network has taken place primarily in the western part of the country, within the area bounded by Saint-Louis, Kaolack, and Dakar. About half of Senegal's extensive road network is passable year-round.

The rail system includes a line from Saint-Louis to Dakar, with a branch line running from Louga inland to Linguère, and the line from Dakar to the Niger River, at Koulikoro. Locomotives are run entirely on diesel fuel. The network is in need of extensive renovation, but the financial situation of the Senegalese Railroad Administration (RCFS) has only permitted the renovation of equipment such as locomotives. Phosphates represent more than 90 percent of the freight carried by rail.

Senegal's three ports are Kaolack, Ziguinchor, and Dakar. Only Dakar is an international port, the other ports handling only local traffic. The second port of western Africa after Abidjan in Côte d'Ivoire, Dakar accommodates ships up to 100,000 tons in nearly 50 berths along six miles of quay. The quays provide refrigerated facilities that harbour

1,000 fishing boats each year. At the end of 1988 a terminal for handling containers was put into service, which will considerably reduce costs and reinforce the important position of the port.

The international airport of Dakar-Yoff at Dakar is served by a number of airlines. It possesses three runways that can accommodate all aircraft currently in service. The airport at Saint-Louis has two runways for local traffic. Domestic services are provided by Air Afrique. There are also airports at Thiès, Ziguinchor, Kaolack, Rosso, Podor, Matam, Tambacounda, Kédougou, Sementi, and Kolda.

Since the end of the 19th century the rivers, of which the Sénégal River has always been the most important, have lost much of their importance. The Sénégal is navigable year-round from Saint-Louis to Podor by boats drawing about three feet of water. Other reaches are only navigable in the rainy season; Kayes in Mali, for example, can be reached only at that time. Navigation of the Sénégal has been aided by the completion of the Diama and Manantali dams. Activity on the Saloum River is due to the presence of the peanut port of Kaolack, while traffic on the Casamance is due to the port of Ziguinchor.

Administration and social conditions. *Government.* The first constitution of Senegal was promulgated in 1963 and has been revised many times since. It proclaims its attachment to fundamental human rights, respect for political, trade-union, and religious freedoms, and also for individual and collective property rights. The Senegalese state is a democratic and secular state, with French as its official language.

The most important feature of the amended constitution of 1983, which was modeled after the 1963 constitution, was the reestablishment of a strongly centralized presidential regime. The amended constitution abolished the office of prime minister and provided for the president of the republic to rule directly. The president is elected for five years. Ministers are appointed by the president, who no longer has the power to dissolve the National Assembly. The assembly, likewise, has lost its power to introduce a motion of censure of the president. Members of the National Assembly are elected to five-year terms at the time of the presidential election. Judicial, executive, and legislative powers are separated.

The office of the president

Senegal is divided into 10 *régions,* which in turn are divided into *départements* and *arrondissements.* Each *région* is administered by a governor whose role is coordinative; he is assisted by two deputy governors, one dealing with administration, and the other with development. A regional assembly composed of general councillors deals with local taxation. In each *département* the prefect represents the republic, as well as the ministers. There are also autonomous urban communes. Dakar is governed by an elected municipal council.

The Senegalese played a pioneering role in the development of a modern political system in the territories of French West Africa. Unlike most single-party African States, in Senegal the existence of a multiplicity of political parties is a constitutional provision. The political-party system is solidly entrenched, and the concept of a single-party system is generally held to be repugnant. This is the reason President Abdou Diouf in 1981 established the pluralism that still prevails today, with more than 15 parties representing socialist, liberal, and Marxist positions. But after the elections in February 1988 a tendency for some parties to regroup began to develop.

At the beginning, political life was of concern only to a limited élite consisting of the intellectuals, the traditional chiefs, and above all the inhabitants of four communes—Saint-Louis, Dakar, Rufisque, and Gorée—who had been French citizens since 1916. After World War II universal suffrage was introduced by stages, with the electorate increasing from 890,000 voters in 1958 to 1,932,265 in 1988. Senegalese citizens today participate in the elections of the president, of the deputies to the National Assembly, of regional councillors, and of municipal councillors.

The electorate

In addition to participation in political-party and trade-union activities, other institutions also permit participation in the political process. These include societies for mutual assistance, which are organized both on a regional and on a village basis; youth associations; and religious groupings, which are most influential. The Muslims, particularly Sunnites, sensible of their popular strength, have called for the establishment of an Islāmic state. The government takes no official notice of such claims.

Justice is administered in the *départements* by justices of the peace and in the *régions* by courts of first instance. Criminal cases are judged by assize courts held at Saint-Louis, Kaolack, Ziguinchor, and Dakar. Dakar is the seat of the Court of Appeal.

Education. In addition to continuing the educational expansion begun during the colonial period, Senegal has made particular efforts to increase school enrollment in rural areas. Among the secondary schools, the Faidherbe Lycée at Saint-Louis and the Van Vollenhoven Lycée at Dakar are the most renowned and the oldest. Technical education is expanding; technical training is provided by institutions in Dakar, Saint-Louis, Diourbel, Kaolack, and Louga.

University education was begun in Dakar in 1957. Following disturbances in 1968, Senegal concluded an agreement with France on higher education by the terms of which the University of Dakar became oriented to its African context. The staff of the university is now about 70 percent Africanized. The Africanization of courses has meant that university degrees, with the exception of those gained in medicine, are no longer equivalent to those obtained in France. About 20 percent of the students are foreign, most from the French-speaking countries of Guinea, Mali, and Burkina Faso. The university was renamed the Cheikh Anta Diop University of Dakar, in honour of the late Senegalese historian and politician, in 1987. The College of Sciences and Veterinary Medicine for French-speaking Africa is also located in Dakar, and a polytechnic college opened at Thiès in 1973.

The Africanization of university courses

Health and welfare. Although still insufficient, Senegal nevertheless has a considerable range of medical facilities. They include hospitals, clinics, maternity homes, and various services specializing in certain diseases, such as tuberculosis, poliomyelitis, syphilis, and leprosy. The Senegalese Red Cross is also active, while environmental studies are conducted by the French Overseas Office of Scientific Research and Technology in cooperation with the World Health Organization.

Thanks to the efforts of doctors during the colonial era, Senegal is rid of the plagues that raged in the early 20th century. If yellow fever, tuberculosis, and similar diseases are still to be encountered, they at least no longer assume epidemic proportions. In rural areas, however, sanitary conditions leave much to be desired, owing to the lack of training and facilities. It must be admitted, however, that this situation is due, above all, to the fact that facilities are primarily concentrated in Dakar, which accounts for much of the total budgetary expenditure for health. The death rate for children between the ages of one and four is five times higher in rural Senegal than it is in Dakar.

Health facilities in Dakar

In rural areas, dwellings, which are usually well constructed, are roofed with straw, while walls are made of either earth or straw. In more prosperous villages, roofs are sometimes of corrugated iron, and walls are sometimes of cement brick. Houses in the towns are of cement with roofs either of tile or of corrugated iron; many families are usually crowded together. The drift from the countryside and the expansion of town populations has frequently resulted in the proliferation of shantytowns.

The standard of living in the countryside is low. Manufactured products are more expensive there than in the cities, which, despite often startling social inequalities, may appear to be a kind of paradise in comparison with living conditions in the countryside.

As is emphasized by the foregoing, the most marked socioeconomic cleavage in Senegal today is between those living in towns and those living in the country. This division results from the modernizing process that has taken hold in the western fringes of the country.

Country life, in its traditional communal village form, is still followed by four-fifths of Senegalese. The family is extended and lives, under the authority of its most elderly member, in a compound consisting of a group of thatched

dwellings, engaging in agriculture for a livelihood. The influence of the marabouts—interpreters of the Qur'ān—remains uncontested. Liquid funds are virtually nonexistent, a subsistence economy being followed.

The way of life of the city dwellers stands in contrast to this, since it typifies the modernizing process at work in Senegal. To speak of city life, moreover, is virtually tantamount to speaking of Dakar, the capital, since three-quarters of town dwellers live in this one city. The primary factor distinguishing the rural farmer from the town dweller is monetary income, since the inhabitants of the towns earn from 10 to 12 times more than those living elsewhere. Better lodging, better food, better education, and better health conditions are available in the towns. Socially, a person who lives in the city, particularly in Dakar, follows a way of life that represents a break with the traditional values of communal village life in its pure form. Senegal today, in sum, presents two distinct aspects to the contemporary observer. On the one hand stands Dakar with its inhabitants in communication with international society and a universal civilization; on the other hand stands the remainder of the country, following a provincial pattern of life that has lost its primary impulse and that is growing increasingly stagnant.

Cultural life. Both the rhythm of life in Senegal and the Senegalese mentality have evolved over a long period of time in a setting that was unacquainted with technology in the Western sense of the word. The attitudes of Senegalese in their relations with nature are consequently different from those of Europeans in general. Fear, magic, and collectivism are dominant in traditional Senegalese life. Writing is absent or constitutes at best the prerogative of no more than the few. The cultural heritage is preserved in oral tradition, of which the guardians have been the most experienced, that is to say the oldest, men. Society thus forms a hierarchy, at the summit of which stand the oldest people.

The social hierarchy

The traditional Senegalese cultural heritage remains much alive. Rites and initiations are actively practiced in rural areas—for example, by the Basari of Kédougou. Among Muslims, youths must be circumcised before being accorded the responsibilities of adulthood.

Art, sculpture, music, and dance remain typically Senegalese in expression. Sculpture is characterized by abstraction and by the ideogram; a sculptured gazelle, for example, may be represented solely by its horns and its neck, while an elephant may be represented only by the immense fan formed by its ears and its trunk. The Senegalese artist thus neglects the material aspect in order to give free rein to ideas and to feelings. Similarly, in the absence of written music, the imagination of the musician is released. Without falling into the realm of fantasy, the *griot* (a West African troubadour and historian) recites poems or tells stories of warrior deeds, drawing upon his own sources of inspiration. Both dance and music owe a great deal to improvisation, which, combined with rhythm, produces an intense effect upon the entire community.

Senegalese literature is incarnated by the former president Léopold Sédar Senghor. The quality and the importance of his work resulted in his election in 1983 as the first black member of the French Academy. He is the poet associated with Negritude, a concept that he defined as consisting, on the one hand, of an attitude of defense of the traditional values of black Africa and, on the other, of tension toward the modernization of these same values. From this concept Senghor drew his political philosophy concerning not only Senegal but the whole of black Africa. Besides Senghor one may also cite the names of Birago Diop, who revived local legends, as well as of such writers as Ousmane Socé, David Diop, Alioune Diop, Cheikh Anta Diop, Cheikh Amidou Kâne, Abdoulaye Sadji, Abdoulaye Ly, Sembène Ousmane, and Bakary Traoré, all of whom are known for works which combine intelligence with the savour of Senegalese life.

Since the World Festival of Negro Arts was organized at Dakar in 1966, a number of existing institutions have been reoriented toward African traditions, while new institutions also were created. Among the new institutions are the Dynamique Museum, the Daniel Sorano Theatre, and the Tapestry Factory of Thiès. The existing institutions that underwent some transformations are the Fundamental Institute of Black Africa, the Houses of Youth and of Culture, and the craft village of Soumbedioune in Dakar, which has become a centre for Senegalese sculpture and goldsmithing.

Senegal was the first of the former French West African territories to have a press. The *Moniteur du Sénégal* was founded in 1854 and was succeeded by the *Journal Officiel de la République du Sénégal.* Important Dakar publications are the *Soliel, Afrique Nouvelle,* and *Le Moniteur Africain.* Radio Sénégal broadcasts in French and English and in several African languages. Television is widely available.

For statistical data on the land and people of Senegal, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (Ca.C.)

HISTORY

Senegal has been inhabited since ancient times. Paleolithic and Neolithic axes and arrows have been found near Dakar, and stone circles, as well as copper and iron objects, have been found in the Sine-Saloum region.

The Tukulor

The Tukulor occupied the lower Sénégal Valley in the 11th century. The name Senegal appears to be derived from that of the Zenaga Berbers of Mauritania. Toward 1040 Zenaga Berbers established a Muslim monastery, perhaps on an island of the river; this became the base of the Almoravids, who converted the Tukulor, defeated the Soninke (Sarakole) empire of Ghana, conquered Morocco, and crossed into Spain. Between 1150 and 1350 the legendary leader Njajan Njay founded the Jolof kingdom, which in the 16th century split into four competing Wolof states: Jolof, Walo, Kajor (Cayor), and Bawol (Baol); while the Serer of Sine and Saloum also established independent kingdoms. Islāmic influence spread throughout the region in variable strength; it gained a new impetus from the later 17th century, and after 1776 Tukulor Muslims established a theocratic confederacy in Fouta-Toro.

Portuguese navigators reached Cape Verde about 1444; they established trading factories at the mouth of the Sénégal, on the island of Gorée, at Rufisque, and along the coast to the south. In the 17th century their power was superseded by that of the Dutch and then the French.

A French factory at the mouth of the Sénégal River was rebuilt in 1659 at N'Dar, an island in the river that became the town of Saint-Louis, and in 1677 France took over Gorée from the Dutch. As bases for French trading companies that bought slaves, gold, and gum in the Sénégal Valley and along the coast, these two towns became homes of Francophone communities of free Christian Africans and Eurafricans. (H.J.D./J.D.H.)

After two periods of British occupation, Saint-Louis and Gorée were returned to France in 1816. When attempts to grow cotton near Saint-Louis proved unprofitable, the colonial economy came to depend on trade for gum in the Sénégal Valley, where an upriver station was founded at Bakel. In the 1830s two coastal factories, at Carabane and Sédhiou, were acquired in Casamance. In 1848 the ailing colonial economy was further disrupted when the Second Republic outlawed slavery on French soil.

In 1854, at the request of local merchants, Napoleon III appointed as governor Commandant Louis-Léon-César Faidherbe, who began to establish French military hegemony. He soon came into conflict with al-Ḥājj ʿUmar, a Tukulor from Fouta-Toro who, having become regional head of the Tijānīyah fraternity, was establishing an economic and military power base in the upper Niger Valley; but a military stalemate after 1857 led to a truce of coexistence. When Faidherbe retired in 1865, French power was paramount over most of the territory of modern Senegal; and growing exports of peanuts, through the new colonial port of Dakar, were providing some economic resources.

In 1879 the French government approved a large program of railway construction. One line was designed to facilitate penetration of ʿUmar's empire; another linked Saint-Louis with Dakar through the main peanut area in Kajor, where commercialization and indebtedness were already disturbing Faidherbe's system of collaboration. In 1886 the deposed *damel* (king), Lat Jor, died in battle

against the French; Islāmic legitimacy among the Wolof now passed to his kinsman Amadu Bamba Mbake; he became spiritual leader of a new fraternity, the Murīdīyah, whose devotees were exhorted to discharge their religious obligations by diligent cultivation of peanuts. Meanwhile, France was consolidating direct control over the rest of Senegal and other African colonies. In 1895 Jean-Baptiste Chaudié became first governor-general of French West Africa, and in 1902 its capital moved to Dakar.

Before this new autocratic empire had established its rigid administrative control over such traditional chiefs as it still tolerated, the Third Republic had recognized the inhabitants of Saint-Louis, Gorée, Dakar, and Rufisque, regardless of race, as French citizens. In 1914 the African electors succeeded in sending to the National Assembly in Paris as their deputy a former colonial official, Blaise Diagne. In return for assistance in recruiting African soldiers in World War I (some 200,000 in all from French West Africa), Diagne obtained confirmation of the French citizenship of this urban minority, even if they chose to retain their status under Muslim law. These privileges were lost between 1940 and 1942, when French West Africa passed under control of the wartime Vichy government, but were restored under the Fourth Republic.

Two socialist deputies elected in 1946, Lamine Guèye and Léopold Sédar Senghor, at first concentrated on restoring these rights of full French citizenship and extending them to the whole Senegalese population. But political life was increasingly influenced by nationalist movements elsewhere in Africa and Asia, as well as by perceptions of strong internal tensions, notably those revealed by a sustained strike of railway workers in 1947–48. Senghor, a poet and philosopher who sought some synthesis between an authentic African identity and French civilization, built a strong political position upon partnership with the leaders of the Murīdīyah and other socially conservative Muslim orders, but he was increasingly driven toward claiming political independence. In 1958 the Senegalese electorate accepted his advice to vote in favour of membership in Charles de Gaulle's proposed French Community; but two years later Senegal claimed and received independence (initially within the short-lived Mali Federation).

As president, Senghor maintained collaboration internally with the *grands marabouts* and externally with France, which continued to provide economic, technical, and military support. But the economy remained vulnerable both to fluctuations in world prices for peanuts and phosphates and to the Sahelian droughts, and the government found it increasingly difficult to satisfy the expectations of the working class or of the rapidly growing student body. Though Senegal remained more tolerant and pluralist than many African states, there were encroachments on political freedoms. In 1976, however, Senghor authorized the formation of two opposition parties; and under Abdou Diouf, to whom he transmitted presidential power in January 1981, these freedoms were tentatively extended.

Under Diouf, the Socialist Party (PS) maintained Senghor's alliance with the Muslim hierarchies. When the PS secured more than 80 percent of the votes in the 1983 elections there were complaints of unfair practice, and the eight deputies returned by the Senegalese Democratic Party (PDS) of Abdoulaye Wade initially refused to take their seats. Nevertheless the framework of parliamentary democracy survived the continuing economic stringency of the 1980s. In 1988 Diouf's presidential majority dropped to 73 percent, and the PDS won 17 of the 120 parliamentary seats. Charges of inequity and fraud, and considerable violence, were followed by the declaration of a state of emergency. Wade was imprisoned; but he was subsequently pardoned.

Diouf found it increasingly difficult to meet prescriptions for economic adjustment while trying to contain social pressures caused by falling export values, rising costs of living, and mounting unemployment. In 1981, when Senegalese troops entered The Gambia to defeat a coup, a Senegambian confederation had been proclaimed; but this never became a reality. During 1989 a long-standing border dispute with Guinea-Bissau aggravated the problem of regional discontent in Casamance, and disputes with Mauritania escalated, resulting in movements of refugees across the border in both directions.

For later developments in the history of Senegal, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (J.D.H.)

COUNTRIES OF THE GUINEA COAST

Benin

The People's Republic of Benin (République Populaire du Bénin)—until 1975 the Republic of Dahomey—has an area of 43,450 square miles (112,600 square kilometres). It consists of a narrow wedge of territory extending for about 420 miles (675 kilometres) southward from the Niger River, which forms part of its northern frontier, to the Gulf of Guinea in the Atlantic Ocean, on which it has a 75-mile seaboard. Benin is bounded to the northwest by Burkina Faso, to the northeast by Niger, to the east by Nigeria, and to the west by Togo. The official capital is Porto-Novo. Cotonou is the largest city, the de facto capital, and the chief port.

A former French colony, Dahomey became an autonomous republic within the French Community in 1959 and gained full independence in 1960. Thereafter, Dahomey underwent more than a decade of political instability, caused partly by internal regional rivalries. The majority of the people inhabiting the southern part of the country are related Fon peoples (including the Adja), who migrated into the region in the 13th century and later founded the Porto-Novo kingdom and the Dahomey (or *Dã-ho-mé,* "on the belly of Dã") kingdom. The Dahomey kingdom—also called the Abomey kingdom after its capital city of Abomey—conquered smaller kingdoms in the south in the 18th century. Dahomey's warriors were defeated by French colonizers in 1893–94. The kingdom then lent its name to the whole country, which became a division of the federation of French West Africa (Afrique Occidentale Française) in 1904. Descendants of Portuguese settlers and returning freed slaves (called Brésiliens, or Brazilians) and the French colonialists were instrumental in spreading Christianity and Western education in the south; by the 1950s, Dahomey was known as the "Latin Quarter" of French West Africa.

A grouping of ancient kingdoms

Large segments of the population in the northern part of the country are descendants of peoples who migrated from the interior of the continent. These peoples were predominantly Muslim, and until independence they had relatively little contact with their neighbours in the south. This cleavage between the north and the south and that in the south between Abomey and Porto-Novo traditionally have fueled conflict among political leaders. Another factor contributing to political instability has been the failing economy, largely dependent upon the export of palm oil and palm kernels and upon external aid, primarily from France. (D.Ro.)

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Benin consists of five natural regions. The coastal region is low, flat, and sandy, backed by tidal marshes and lagoons. It is composed of, in effect, a long sandbar on which grow clumps of coconut palms; the lagoons are narrower in the western part of the country, where many have become marshes because of silting, and wider in the east, and some are interconnected. In the west the Grand-Popo Lagoon extends into neighbouring Togo, while in the east the Porto-Novo Lagoon provides a natural waterway to the port of Lagos, Nigeria, although its use is discouraged by the political boundary. Only at Grand-Popo and at Cotonou do the lagoons have outlets to the sea.

Behind the coastal region extends the *barre* country—

The *barre* country

the word being a French adaptation of the Portuguese word *barro* ("clay"). A fertile plateau, the *barre* region contains the Lama Marsh, a vast swampy area stretching from Abomey to Allada. The landscape is generally flat, although occasional hills occur, rising to about 1,300 feet (400 metres).

The Benin plateaus, four in number, are to be found in the environs of Abomey, Kétou, Aplahoué (or Parahoué), and Zagnanado. The plateaus consist of clays on a crystalline base. The Abomey, Aplahoué, and Zagnanado plateaus are from 300 to 750 feet high, and the Kétou plateau is up to 500 feet in height.

The Atakora Mountains, in the northwest of the country, form a continuation of the Togo Mountains to the south. Running southwest to northeast and reaching an altitude of 2,103 feet (641 metres) at their highest point, they consist of a highly metamorphosed quartzite interior.

The Niger plains, in the northeast of Benin, slope down to the Niger River Valley. They consist of clayey sandstones.

Drainage. Apart from the Niger River, which, with its tributaries the Mékrou, Alibori, and Sota, drains the northeastern part of the country, the three principal rivers in Benin are the Mono, the Couffo, and the Ouémé. The Mono, which rises in Togo, forms the frontier between Togo and Benin near the coast. The Couffo, near which stands Abomey, flows southward from the Benin plateaus to drain into the coastal lagoons at Ahémé. The Ouémé rises in the Atakora Mountains and flows southward for 280 miles; near its mouth it divides into two branches, one draining to the east into Porto-Novo Lagoon and the other to the west into Nokoué Lake. The Atakora Mountains form a divide between the Volta and Niger basins.

Two climatic zones

Climate. Two climatic zones may be distinguished—a southern and a northern. The southern zone has an equatorial type of climate with four seasons—two wet and two dry. The principal rainy season occurs between mid-March and mid-July; the shorter dry season lasts to mid-September; the shorter rainy season lasts to mid-November; and the principal dry season lasts until the rains begin again in March. The amount of rain increases toward the east. Grand-Popo receives only about 32 inches (800 millimetres) a year, whereas Cotonou and Porto-Novo both receive approximately 50 inches. Temperatures are fairly constant, varying between about 72° and 93° F (22° and 34° C), and the relative humidity is often uncomfortably high.

In the northern climatic zone, there are only two seasons, one dry and one rainy. The rainy season lasts from May to September, with most of the rainfall occurring in August. Rainfall amounts to about 53 inches a year in the Atakora Mountains and in central Benin; further north it diminishes to about 38 inches. In the dry season the harmattan, a hot, dry wind, blows from the northeast from December to March. Temperatures average about 80° F (27° C), but the temperature range varies considerably from day to night. In March, the hottest month, diurnal temperatures may rise to 110° F (43° C).

Plant and animal life. The original rain forest, which covered most of the southern part of the country, has now largely been cleared, except near the rivers. In its place, many oil palms and rônier palms have been planted and food crops are cultivated. North of Abomey the vegetation is an intermixture of forest and savanna (grassy parkland), giving way further north to savanna. Apart from the oil and rônier palms, trees include coconut palms, kapok, mahogany, and ebony.

In the extreme north is the "W" National Park (1,938 square miles), which extends into Burkina Faso and Niger. Its varied animal life includes elephants, leopards, lions, antelope, monkeys, wild pigs, crocodiles, and buffalo. There are many species of snakes, including pythons and puff adders. Birds include guinea fowl, wild duck, and partridge, as well as many tropical species. The Pendjari National Park (1,062 square miles) borders on Burkina Faso.

Settlement patterns. The southern third of the country is by far the most densely populated region. The cultivation of subsistence crops, such as corn (maize), cassava, and yams, is intensive on the outskirts of the towns. The *barre* region and the Benin plateaus are planted with oil palms, which form the cash crop, as well as with subsistence crops. To the north, the aspect of the countryside changes as savanna vegetation increases and the population diminishes; some areas are uninhabited, except by Fulani nomads. Villages, instead of being encountered frequently as in the south, become scattered.

Benin's towns

Although some of the towns have a modern aspect, vestiges of former centuries are still to be seen, such as the old Portuguese fort at Ouidah, dating from the time of the slave trade. Cotonou is primarily a colonial town and is ethnically heterogeneous, whereas Porto-Novo, an old African-founded town, combines both African and European features. Abomey, the ancient capital, now somewhat in decline, is an important market and a centre for crafts. Parakou is an important market town. (S.S.A./D.Ro.)

The people. About two-thirds of the people live in the southern third of Benin. Many of these are clustered about the port of Cotonou, which is the focus of the commercial and political life of the country.

Ethnic and linguistic groups. While French is the official language, the most widely spoken languages are Fon and Ge (Mina), both dialects of Ewe; Bariba; Yoruba; and Dendi.

The Fon are predominant in the south and are related to other southern groups, including the Adja and the Gun (Goun). The Aizo have been largely assimilated into the Fon and Adja. The Yoruba, related to the Yoruba of Nigeria, are called the Nago (Nagot) in Benin and include the small Holli and Ketu groups.

The Bariba, a Voltaic-speaking group, are the major ethnolinguistic group in the north. Others in the north are the Somba, who live in the Atakora Mountains and in adjoining parts of Togo; the Pila Pila; the Dendi, who are associated with the Niger Valley; and the nomadic Fulani (Peul).

A few thousand foreigners, mostly French and Lebanese, also live in Benin.

Religious groups. Although Christian missions have been active in the coastal region since the 16th century, only one-fifth of the total population is Christian; of the Christians, about four-fifths are Roman Catholic. Islām has adherents in the north and southeast; about one-sixth of the total population is Muslim. Most of the population adheres to traditional religions. In the south, animist religions, which include fetishes (objects regarded with awe as the embodiment of a powerful spirit) for which Benin is renowned, retain their traditional strength.

Traditional religions

Demographic trends. Benin's rate of population growth is high for western Africa, resulting primarily from a birth rate that is higher than the regional average and a death rate that is lower. Moreover, nearly one-half of the population is less than 15 years of age, assuring the country's continued high growth rate. Life expectancy for males is about 47 years and for females about 51 years. Only about one-fifth of the population is urban, most of whom live in Cotonou, the only city with a population of more than 500,000. (D.Ro.)

The economy. Until 1974 Dahomey had a market economy in which France, as former colonial power, had a central role. One of the main objectives of the 1974 ideological shift of the regime of President Mathieu Kérékou was to remedy economic conditions, which had been deteriorating since independence. This shift (known as the Beninese Revolution) meant the complete restructuring of the society along Marxist-Leninist principles, including nationalization of most sectors of the economy (with compensation paid, ironically, by the French), complete or partial state control of enterprises, and attempts at disengaging from French economic dependence while turning to the Soviet Union and other socialist countries. This switch proved a complete failure. The restructured economy did not result in a balanced budget (public debt grew steadily), did not eliminate corruption, and did not bring about growth in investments. A new policy of "liberalism" initiated in 1982 has begun denationalization of state-owned enterprises, encouragement of private internal and foreign investment, and encouragement of the growth of French financial assistance.

Resources. Only a few stretches of forest remain in Benin, mostly in the south and central areas; these contain mahogany, iroko, teak, samba, and other tropical hardwoods. The rivers and lagoons are rich in fish. Mineral deposits include iron-ore both in the Atakora Mountains and northeast of Kandi, limestone deposits at Onigbolo, chrome ore and a little gold in the northwest near Natitingou, marble at Dadjo, an important deposit of pottery clay at Sakété, and ilmenite (a mineral source of titanium) near the coast. Offshore oil was discovered in 1968 in the Sémé field near Cotonou and has been exploited since 1982.

Agricultural production

Agriculture and fishing. About 70 percent of the working population depends on agriculture. Traditional farming produces yams, cassava, corn, and sorghum. Palm oil and palm kernels constitute the main cash crop; other cash crops are cotton, cocoa, peanuts (groundnuts), coffee, and tobacco. Livestock include cattle, sheep and goats, pigs, horses, and poultry. Substantial quantities of fish are caught annually in the lagoons and rivers, while coastal fishing produces a smaller, but growing, amount. Most of the fish is exported to Nigeria or Togo. Shrimp and deep-sea fishing are developing, using modern vessels.

Industry. Manufacturing plants and secondary industries include several palm-oil-processing plants, in Ahozon, Avrankou, Bohicon, Cotonou, Gbada, and Pobé; cement plants at Onigbolo and Pobé; several cotton ginning facilities in the north; a textile mill at Parakou; a sugar refining complex at Savé; a soft-drink plant; a brewery; and two shrimp-processing plants.

Electricity is generated thermally by plants located at Bohicon, Parakou, Cotonou, and Porto-Novo. About half of Benin's demand for electricity is met by importing power from Ghana's Volta River Project at Akosombo. In 1988 operations commenced at the hydroelectric installation of the Mono River Dam, a joint venture between Benin and Togo on their common southern boundary.

Finance. The Banque Centrale des Etats de l'Afrique de l'Ouest (BCEAO), established in Paris in 1958 and transferred to Dakar, Senegal, in 1974, is the central bank of most but not all French-speaking western African states. It has a branch in Benin. In addition, there are the Banque Béninoise de Développement, a state-owned bank that provides financing of development projects; the Banque Commerciale du Bénin, the only commercial bank also handling governmental transactions; and the Caisse Nationale de Crédit Agricole, a state-owned organ for providing agricultural credit. The Caisse Centrale de Coopération Economique is an official French agency through which French aid is administered. The common currency of most former French colonies in West and Equatorial Africa, including Benin, is the CFA (Communauté Financière Africaine) franc, which is fully guaranteed by and pegged to the French franc.

Trade. The main exports are petroleum, cotton, palm oil, and cocoa. Among the principal imports are miscellaneous manufactured goods, tobacco, and machinery. In the late 1980s, the value of imports was reduced from 10 times to four times the value of exports. More than 90 percent of exports and some 40 percent of imports are exchanged with the countries of the European Economic Community. Alongside these official trade relations, however, flourishes a significant informal trade across the long international boundaries with Togo, Burkina Faso, Niger, and, especially, Nigeria.

Transportation. There are two principal roads, one running along the coast from the Togo border in the west via Cotonou and Porto-Novo to the Nigerian border in the east, and the other running north from Cotonou to Parakou and on to the Niger frontier. Road transport is usually by truck.

The railroad consists of three lines. One line runs from Cotonou to Séboroué; a second line runs from Cotonou via Porto-Novo to Pobé; and the third line runs north from Cotonou via Pahou to Parakou.

The lagoons, running parallel to the coast, are used by small craft. The Ouémé River is navigable for about 125 miles, the Couffo for about 80 miles, and the Mono for about 60 miles. The country's only port, at Cotonou, was expanded in the mid-1980s to handle three million metric tons of cargo each year. The airport at Cotonou links Benin with other western African countries and with France, the United Kingdom, the United States, and the Soviet Union. The government maintains a small domestic air service. (S.S.A./D.Ro.)

Administration and social conditions. *Government.* The Kérékou regime, which came to power in 1972, has achieved a shaky but unprecedented stability. The Marxist rhetoric introduced in 1974 culminated in repressive military rule in the late 1970s, but this had largely ceased by the early 1980s. Nevertheless, the People's Revolutionary Party of Benin (PRPB) is still the only legal political party. Its leadership nominates candidates for elections in seven socio-professional categories, one of which is the military. From among the candidates, 196 commissioners to the National Revolutionary Assembly are elected by citizens over 18 years of age in universal, direct, and secret balloting. Elections are held every five years. The assembly elects the president, who is the head of state.

Local administration

Following the administrative reorganizations of 1974 and 1978 the country was divided into six *provinces,* each structured to bring all levels of the population into the political process. In each *province* there is a prefect and an advisory Provincial Revolutionary Council composed of the heads of the army and the gendarmerie in the *province* and five delegates from each *district* into which the *provinces* are divided. The six prefects in turn are members of the government, which is composed of both military and civilians. There are 84 *districts,* each with a District Revolutionary Council (CRD) that advises the *district* chief appointed by the government. The executive secretaries of all the CRDs are also members of the National Revolutionary Assembly. The *districts* in turn are composed of *communes.* Each *commune* has an elected Local Revolutionary Council, which advises the chief of the *commune.*

At the head of the judicial system is the High Court. There are also a council of the magistracy, a court of appeal at Cotonou, a tribunal of the first instance, and local tribunals of traditional African or Muslim law.

Education. During the 1975–76 school year major reforms were introduced to raise school attendance, to change the curriculum in primary and secondary schools to conform to Marxist-Leninist ideology, and to fit the academic calendar to the agricultural cycle. Local languages were to substitute for French as the language of instruction. Apart from raising attendance substantially, the reforms largely failed, as standards were lowered and had to be readjusted. One problem was that many teachers either left the teaching profession for administrative positions or took positions in other French-speaking African countries where salaries, and often standards, were higher.

The National University of Benin in Abomey-Calavi, near Cotonou, was established in 1970 with French financial and pedagogical assistance. In 1976, following educational reforms, three polytechnic schools opened in Pobé, Bohicon, and Natitingou.

Health and welfare. Health services are limited by budgetary restrictions and by the paucity of physicians and other medical personnel. The country's largest hospitals are located at Cotonou, Porto-Novo, Abomey, Ouidah, Natitingou, and Parakou. In addition, there are medical centres, dispensaries, maternity centres, and leprosariums. Campaigns have been waged against meningitis, leprosy, malaria, and sleeping sickness.

Diversity of cultural traditions

Cultural life. Each ethnic group in Benin has a centuries-old cultural tradition. This diversity is clustered in two distinct regions: the largely animist and Muslim north and the largely animist and Christian south. In both regions traditions are not merely folkloric remnants shown in festivities and museums but are encountered at any hour of the day practically everywhere. Cultural traditions are seen in the variety of colourful dress, languages, jewelry, and the open-air markets. The former French colonial cultural impact is preserved in the wide use of the French language among educated town dwellers.

Daily life. The extended family is the basic social unit, and in much of the country it is interlinked with other families sharing the same ancestral line. In villages and in

traditional towns such as Porto-Novo, related, often polygamous, families live in compounds in the same neighbourhood, where together they practice religious rites, celebrate festivities, and share or coordinate cooking, marketing, and care for the young and old. Traditional ceremonies with music and dance held by neighbourhood clubs or "secret societies" are commonplace. Celebrations of Christian and Muslim holy days and festivals are imbued with traditional flavour.

The arts. Artistic traditions in Benin are very old and are represented in practically every village. Plastic art is the most prominent, as carved wooden masks representing images and spirits of the departed are made and used in traditional ceremonies. Other artistic items are bronze statuettes, pottery, appliquéd tapestries recounting the history of kings of precolonial Dahomey, and fire engraving on wooden bowls, which often have religious meaning. Probably the best-known art objects are the Yoruba wooden masks called *guelede* from the region of Porto-Novo.

Cultural institutions. An artisan village is attached to the Historical Museum of Abomey (formerly the Royal Palace). There is an excellent ethnographic museum in Porto-Novo, a historical museum in Ouidah, and the Open-Air Museum of Ethnography and Natural Sciences in Parakou. The National Library is in Porto-Novo. Art galleries are the Cultural and Artistic Centre and the French Cultural Centre, both in Cotonou, and the CAZAM in Porto-Novo.

Recreation. The national sport played by several teams is football (soccer). There is a modern sport stadium in Cotonou. In the far north the Porga National Park, where wildlife is protected, attracts many tourists.

Press and broadcasting. Radio programs are broadcast from Cotonou in French, English, and a number of local languages. There is also a limited television service. A daily newspaper, *Ehuzu,* is published in Cotonou and is controlled by the government; there are also two other dailies and several weekly or biweekly publications. Newspapers published in Senegal and Côte d'Ivoire (in French) and newspapers and magazines from elsewhere may be found in bookstores and newsstands.

For statistical data on the land and people of Benin, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (D.Ro.)

HISTORY

As a political unit, Benin was created by the French colonial conquest at the end of the 19th century. In the precolonial period, the territory comprised a multiplicity of independent states, differing in language and culture. The south was occupied mainly by Ewe-speaking peoples, who traced their traditional origins to the town of Tado (in modern Togo). During the 16th and 17th centuries, the most powerful state in this area was the kingdom of Allada (Ardra), but in the 18th and 19th centuries its place was taken by Dahomey. In the north, the largest group was the Bariba, the most important state being the kingdom of Nikki, which formed part of a confederacy including other Bariba states located in what is today Nigeria.

The slave trade. The Portuguese first explored the coast of Benin in 1472 but did not begin trading there until 1553. During the 17th century the Dutch, English, French, and other Europeans also entered the trade. The principal export before the mid-19th century was always slaves. The volume of slave exports was at first small, but it increased rapidly in the second half of the 17th century, when this area became known to Europeans as the "Slave Coast," and remained high until the 1840s. The principal centre for the trade was the coastal kingdom of Ouidah (Whydah), which was originally a tributary of Allada but which had become an independent state by the 1680s. The slaves exported were predominantly war captives and were drawn from the entire area of modern Benin, including northern peoples such as the Bariba as well as communities near the coast. The Atlantic slave trade had a substantial and deleterious impact in Benin, causing the depopulation of certain areas as well as a general militarization of society. The prominence of slaves from this area in the transatlantic trade is reflected in the survival of elements of its culture in black communities of the New World, especially in the "voodoo" religion of Haiti, which incorporates many spirit cults and deities of the Ewe-speaking peoples.

Beginning of European trade

The kingdom of Dahomey. Dahomey was the state of the Fon people; its capital was at the town of Abomey, north of Allada. It was originally a dependency of Allada, but during the 17th century a ruler called Wegbaja declared himself king and made Dahomey an independent state. Under King Agaja (reigned 1708?–40) Dahomey overran the coastal area, conquering Allada in 1724 and the commercial centre of Ouidah in 1727, thus establishing itself as the dominant power in the area. A section of the royal family of Allada, however, founded the new kingdom of Porto-Novo, on the coast to the east, which successfully resisted Dahomian authority and competed with Ouidah for control of the Atlantic trade. Dahomey itself was attacked and defeated by the kingdom of Oyo, to the northeast (in modern Nigeria), to which it was obliged to pay tribute from 1730 onward. Dahomey attained the height of its power under kings Gezo (1818–58) and Glélé (1858–89). Gezo liberated Dahomey from its subjection to Oyo by defeating the latter in 1823. Dahomian attempts at expansion eastward, however, brought it up against the powerful state of Abeokuta (also in Nigeria). Dahomian attacks upon Abeokuta in 1851 and 1864 were decisively defeated.

Dahomey was a despotic and militaristic kingdom. Its power was based upon a highly trained standing army, which included a female contingent (called the "Amazons" by Europeans) drawn from the king's wives. The king's authority was buttressed by an elaborate cult of the deceased kings of the dynasty, who were honoured by the offering of human sacrifices at yearly public ceremonies (the "annual customs"). Its rulers succeeded in uniting the disparate communities which they absorbed into a new national identity, so that the conquered subjects of Dahomey came to regard themselves as Fon. During the 18th and early 19th centuries, Dahomey was a major supplier of slaves for the transatlantic trade, but by the mid-19th century the volume of the slave trade was in decline. In 1852 King Gezo was forced by a British naval blockade to accept a treaty abolishing the slave trade, although this was evaded in practice. From the 1840s onward Gezo promoted the export of palm oil, produced by slave labour on royal plantations, as a substitute for the declining slave trade.

The French conquest and colonial rule. During the 17th century several of the European nations engaged in the Atlantic slave trade maintained trading factories in the Dahomey area, and during the 18th century the English, French, and Portuguese all possessed fortified posts in Ouidah. The French first established a factory in Allada in 1670 but moved from there to Ouidah in 1671. Although this factory was abandoned in the 1690s, the French built a fort (known as Fort Saint Louis) in Ouidah in 1704. The European forts in Ouidah were, however, all abandoned around the end of the 18th century, the French establishment being withdrawn in 1797.

In 1842 the French fort at Ouidah was reoccupied as a base for the new trade in palm oil, and in 1851 the French government negotiated a commercial treaty with King Gezo of Dahomey. Subsequently fears of preemption by British colonial expansion led to the extension of formal French rule in the area. A protectorate was briefly established over the kingdom of Porto-Novo in 1863–65 and was definitively reestablished in 1882. Treaties purporting to secure cession of the port of Cotonou, between Ouidah and Porto-Novo, were also negotiated with the Dahomian authorities in 1868 and 1878, though Cotonou was not actually occupied until 1890. King Behanzin, who had succeeded to the Dahomian throne in 1889, resisted the French claim to Cotonou, provoking the French invasion and conquest of Dahomey in 1892–94. Behanzin was then deposed and exiled, and Dahomey became a French protectorate.

French ambitions to extend their control into the interior north of Dahomey were threatened by the rival expansionism of the British, who were established in what was to become their colony of Nigeria to the east, and

in 1894 both the British and French negotiated treaties of protection with the kingdom of Nikki. The Anglo-French convention of 1898, however, settled the boundary between the French and British spheres, conceding Nikki to the former. The boundary with the German colony of Togo to the west was settled by the Franco-German conventions of 1885 and 1899. The present frontiers of Benin were definitively established in 1909, when the boundaries with the neighbouring French colonies of Upper Volta (modern Burkina Faso) and Niger were delimited. The colony was at first called Benin (from the Gulf of Benin, not the precolonial kingdom of Benin, which is in Nigeria to the east), but in 1894 it was renamed Dahomey, after the recently incorporated kingdom. From 1904 Dahomey formed part of the federation of French West Africa, subordinate to the governor-general in Senegal.

Settlement of boundaries

Decolonization and independence. With the liberalization of the French colonial system after World War II, in 1946 Dahomey became an overseas territory of France. It was created an autonomous republic within the French Community in 1959 and achieved complete independence on Aug. 1, 1960. During the period of decolonization, the nationalist movement in Dahomey became fragmented, with the emergence of three regionally based political parties—led by Sourou-Migan Apithy (president in 1964–65), Justin Ahomadégbé (president in 1972), and Hubert Maga (president in 1960–63 and 1970–72), drawing their principal support respectively from Porto-Novo, Abomey, and the north. After independence in 1960, these political problems were exacerbated by economic difficulties, reflected in student and trade union unrest. The ensuing political instability resulted in no fewer than six successful military coups d'état between 1963 and 1972 and periods of army rule in 1965–68 and 1969–70. In a last military coup, on Oct. 26, 1972, power was seized by Major (later General) Mathieu Kérékou. From 1974 Kérékou pursued a Marxist-Leninist policy, based on nationalizations and state planning of the economy, and on Nov. 30, 1975, he renamed the country the People's Republic of Benin.

For later developments in the history of Benin, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (R.La.)

Cameroon

The Republic of Cameroon (République du Cameroun) is a triangle-shaped state wedged between western and central Africa. It covers an area of 179,714 square miles (465,458 square kilometres) and is bordered by Nigeria to the northwest, Chad to the northeast, the Central African Republic to the east, Congo to the southeast, Gabon and Equatorial Guinea to the south, and the Atlantic Ocean to the southwest. The capital of Cameroon is Yaoundé.

The country's name is derived from Rio dos Camarões (River of Prawns)—the name given to the Wouri River estuary by Portuguese explorers of the 15th and 16th centuries. Camarões was also used to designate the river's neighbouring mountains. The French form is Cameroun. Until the late 19th century, English usage confined the term the Cameroons to the mountains, and the estuary was called the Cameroons River or, locally, the Bay. In 1884 the Germans extended the word Kamerun to their entire protectorate, which largely corresponded to the present state.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Cameroon can be divided into the four geographic regions of the south, the west, the central, and the north. The southern region extends from the Sanaga River to the southern border and from the coast eastward to the Central African Republic and Congo. It consists of coastal plains that are about 25 miles (40 kilometres) wide and a densely forested plateau with an average elevation of a little more than 2,000 feet (600 metres).

The western region extends north and west from the Sanaga River and continues north along the Nigerian border as far as the Bénoué (Benue) River. The relief is mostly mountainous, the result of a volcanic rift that extends northward from the island of Bioko (Fernando Po). Near the coast, the active volcanic Mount Cameroon rises to the highest elevation in western Africa—13,435 feet (4,095 metres).

Upland areas

The central region extends east from the western highlands and from the Sanaga River north to the Bénoué River. The land rises progressively to the north and includes the Adamawa (Adamaoua) Plateau, with elevations that are everywhere more than 3,000 feet and average about 4,500 feet.

The northern region extends from the Bénoué River north to Lake Chad. Composed largely of vast savanna-covered plains, the region contains scattered inselbergs, or mounds of erosion-resistant rock rising above the plains. The Mandara Mountains in the west, with elevations of more than 4,700 feet, are the highest features in the north.

Drainage. The rivers of Cameroon form four large drainage systems. In the south, the Sanaga, Wouri, Nyong, and Ntem rivers drain into the Atlantic Ocean. The Bénoué River and its tributary, the Kébi, flow into the Niger River basin of Nigeria. The Logone and Chari rivers—which form part of the eastern border with Chad—drain into Lake Chad, whereas the Ngoko River in the extreme southeast joins the Sangha River and flows into the Congo (Zaire) River Basin.

Climate. Lying wholly within the tropics, the country is hot throughout the year; mean annual temperatures range between 70° and 82° F (21° and 28° C). They are lowest in areas of high elevation, and the seasonal range of temperatures almost everywhere is small, except in the dry north, where the diurnal range is also greatest.

Rainfall decreases from south to north. Along the coast, the rainy season lasts from April to November, and the relatively dry season lasts from December to March; a transition period from March to April is marked by violent winds. The mean annual rainfall of more than some 100 inches (3,800 millimetres) occurs in about 150 days.

In the central plateau region, rainfall decreases to 60 inches. There are four seasons—a light rainy season from May to June, a short dry season from July to October, a heavy rainy season from October to November, and a long dry season from December to May. The north, however, has only a dry season from October to May and a rainy season from June to September, with a mean annual rainfall averaging about 30 inches, but dropping as low as 15 inches near Lake Chad. The wettest part of the country lies in the western highlands. Debundscha Point on Mount Cameroon has a mean annual rainfall of more than 400 inches (10,000 millimetres)—an average attained by only two other locations in the world—most of which falls from May to October.

Plant and animal life. The hot and humid south supports dense rain forests, in which mahogany, ebony, obeche, dibetu, and sapelli may grow to more than 200 feet. There are large numbers of orchids and ferns.

Plants of the tropical rain forest

Mangroves grow along the coasts and the mouths of rivers. The rain forest gives way to the semi-deciduous forest of the central region, where a number of tree species shed their leaves during the dry season. North of the semi-deciduous forest, the vegetation is composed of wooded savanna with scattered trees 10 to 60 feet high. The density of trees decreases toward the Chad Basin, where they are sparse and mainly of *Acacia* species.

Between 4,000 and 8,000 feet, the tropical rain forest differs from that of the lowlands; the trees are smaller, are of different species, and are festooned with mosses, lichens, and other epiphytes. Above the forest zone are drier woodland, tall grassland, or patches of mountain bamboo. Above about 7,800 feet in the interior and above about 10,000 feet on Mount Cameroon, short grasses predominate.

The country's dense forests are inhabited by screaming red and green monkeys, chimpanzees, and mandrills, as well as rodents, bats, and numerous birds—from tiny sunbirds to giant hawks and eagles. A few elephants survive in the forest and in the grassy woodlands where baboons and several types of antelope are the most common animals. The Waza National Park in the north, which was originally created for the protection of giraffes and antelope, abounds in both forest and savanna animals.

Settlement patterns. In general there is a cultural division between the north and the south. The northern savanna plateau is inhabited by Sudanic and Arab pastoralists who migrate seasonally in search of grazing land, whereas the forested and hilly south is peopled by Bantu agriculturists in permanent villages. Social and economic divisions in general follow ethnic lines. In the north Muslims dominate, whereas the more urban south is dominated by those who have acquired Christianity and a Western education. Because of the multiplicity of tribes, precise divisions are difficult to draw.

Population density is greatest in the western highlands, the southern forest, and along parts of the coast; it is lowest in the southeast interior. Douala is the largest city and the country's main port. Yaoundé is an important transportation and communications centre. Garoua is a port on the Bénoué River.

The people. The country has been described as a "racial crossroads" because of its more than 100 different ethnic groups. The Bantu settled in the Cameroons from Equatorial Africa. The first group that invaded the country included the Maka, Ndjem, and Duala. They were followed at the beginning of the 19th century by the Fang (Pangwe) and Beti peoples.

The principal linguistic and ethnic groups

The Sudanic-speaking peoples in the north include the Sao, who live on the Adamawa Plateau; the Fulani; and the Kanuri. The Fulani came from the Niger Basin in two waves—in the 11th and 19th centuries; they were Muslims who converted and subjugated the peoples of the Logone valley and the plains areas to the south along the Kébi and Faro rivers. The third ethnic group consists mainly of small tribes, except for the Bamileke, who live between the lower slopes of the Adamawa Plateau and Mount Cameroon. Other western Bantu-speaking tribes include the Tikar, who live in the Bamenda region, the Bali, the Mum, and the Keaka.

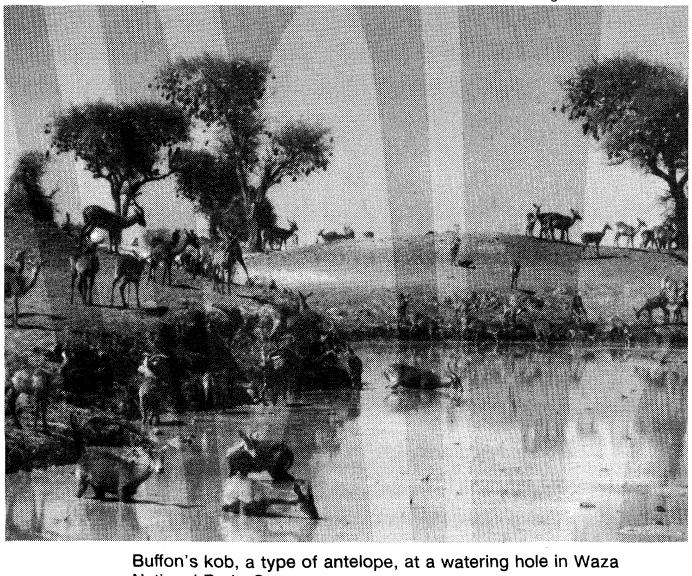
The Pygmies, locally known as the Baguielli and Babinga, are found within the southern forests. They have been hunters and food-gatherers for thousands of years and live in small hunting bands.

European missions and colonization led to the introduction of European languages. During the colonial era German was the official language; it was later replaced by English and French, which have retained their official status.

About one-quarter of the population continues to adhere to traditional religious beliefs. About a third of the population is Roman Catholic, one-fifth is Protestant, and one-fifth is Muslim.

The economy. Cameroon has an agricultural economy supplemented by petroleum and mineral extraction. The economy—which is dominated by French financial institutions and banking houses—has maintained steady growth since independence. Cameroon has suffered only moderate trade deficits, since imports and exports are usually in reasonably good balance. The government dispensed with external budgetary support after 1969.

© Michael Fogden—Animals Animals



Buffon's kob, a type of antelope, at a watering hole in Waza National Park, Cameroon.

The government sets the guidelines for development in five-year plans and attracts private capital for certain sectors of the economy. It has stressed balanced development of economic sectors and has avoided hasty industrialization. Cameroon is an associate member of the European Economic Community (EEC, or Common Market) under the Lomé conventions and a member of the Central African Customs and Economic Union (Union Douanière et Économique de l'Afrique Centrale; UDEAC).

Resources. The largest mineral reserves are those of bauxite at Minim-Martap and Ngaoundere on the Adamawa Plateau; the deposits amount to more than one billion tons. There is some gold in eastern Cameroon, and cassiterite occurs at Mayo Darlé in the northwest. Surveys have found iron ore near Kribi, on the coast south of Douala. Production from offshore oil deposits in the Rio-del-Rey estuary near the border with Nigeria increased markedly in the 1980s.

More than half of Cameroon is forested, but only about three-fourths of that area is amenable to exploitation. Tropical hardwood timber is the primary forest resource. The rivers and offshore waters are a rich source of fish, and the grasslands support valuable herds of cattle. The wild animal life provides a firm basis for tourism.

Agriculture, forestry, and fishing. Agriculture accounts for about half of the country's export earnings but employs nearly three-fourths of the economically active population. The main subsistence crops include plantains, yams, and cassava in the south and peanuts (groundnuts), millet, and cassava in the north. Most cultivation is carried out on small farms by traditional simple methods, but Cameroon is generally self-sufficient in food production. Cash crops are grown on large plantations owned by French interests in the east and by agricultural cooperatives and a government agency—the Cameroon Development Corporation—in the west.

Cocoa and coffee production

Cameroon is one of the world's leading cocoa-producing countries. The cacao plant, which yields the cocoa bean, is grown on small farms throughout the southwest, in parts of the coastal areas, and in western Cameroon. Much attention was given to the improvement of cocoa production under the second five-year plan; under "Operation 100,000 Tons," agricultural extension workers helped farmers to tend their cacao plants and to fight pests and diseases.

Among the French-speaking African countries, Cameroon is the second largest coffee exporter, after Côte d'Ivoire. Eastern Cameroon produces about 90 percent of the coffee crop. About three-fourths of the coffee produced is robusta, the rest arabica.

Bananas are an important export product by value, but production and exports declined in the 1960s because of the lack of new markets and the loss of imperial preference in the United Kingdom in 1967. Bananas are grown on plantations in the southwest and on small farms.

Cotton was introduced in 1952; it is grown largely in the northern grasslands by private farmers. Yields and annual production figures have expanded steadily since the 1950s.

Rubber is produced on government and domestic company plantations at Dizangué in the southern area of eastern Cameroon and in western Cameroon. Rubber is produced mainly for export, but a small quantity of rubber produced at Dizangué is used domestically.

Other cash crops include rice, sugarcane, tropical fruits, tobacco, and palm oil. Except for pigs, which are prohibited in areas where Islām is predominant, livestock (mainly cattle, sheep, and goats) is concentrated on the northern Adamawa Plateau. Some livestock, meat, and hides are exported to neighbouring countries.

Forestry

Forestry has been well developed in the most accessible areas along the Douala–Yaoundé railway and along the main roads. The logs are transported by water, road, and rail. Logs, cut timber, and palm kernels are produced for export, whereas wood for construction purposes, firewood, charcoal, kola nuts, and palm oil and kernels are produced for domestic use.

Marine fishing is carried out by individual fishermen using traditional methods as well as by large companies. The fishermen go to sea in dugout canoes and use nylon nets, whereas the companies employ trawlers based in Douala. Fishing in the rivers often yields an even larger catch.

Industry. Mining has become an important sector of the economy. Oil production rose sharply in the 1970s and by the 1980s accounted for more than one-third of export earnings. Cameroon's substantial bauxite reserves remain largely undeveloped, however. The remote locations of the deposits and the country's lack of necessary infrastructure have prohibited their exploitation. The aluminum smelter at Edéa is supplied by imported bauxite, mostly from Guinea. Limestone and pozzolana are quarried for the cement industry, and smaller quantities of gold and cassiterite are mined.

The manufacturing industry of Cameroon employs a relatively small percentage of the working population. The industrial sector consists of a major heavy industrial complex, the Edéa aluminum smelter, a petroleum refinery at Victoria, and hundreds of small and medium-sized enterprises engaged in processing, various manufacturing procedures, and the assembly of radios.

Apart from a few establishments, most industrial enterprises have been set up since the 1960s. The largest development has been in cotton textiles, centred around factories in Douala. Food-processing plants include a sugar refinery at Mbandjok; palm-oil plants in Bota, Lobe, and Edéa; a chocolate factory at Yaoundé; a cassava starch mill at Bertoua; and a fish canning plant and flour mills at Douala.

The main source of hydroelectric power is the Sanaga River; the power stations at Edéa on the Sanaga Falls and at Songo-Loulou are the country's largest. The Akam-Wouri River in the west also provides hydroelectric power as it drops from the central plateau to the coastal region. In the north, the hydroelectric station at Lagdo on the Bénoué River was built with aid from China.

Finance. Together with 13 other countries in western Africa, Cameroon is a member of the Franc Zone. All members share a common currency, the CFA (Communauté Financière Africaine, or African Financial Community) franc, which is linked to the French franc at a fixed rate of exchange. France is represented on the board of directors of the Central Bank in Yaoundé; its notes and coins are legal tender in each member state of the UDEAC. Each member state has its own monetary committee, on which France is represented, and a National Credit Council. The latter body advises the respective governments on credit policy and banking organization and methods. There are a number of foreign commercial banks in Cameroon, including French, British, and U.S. institutions. Most insurance companies are French-owned. There is no stock exchange.

Monetary relations with France

Trade. The main exports are agricultural products, crude petroleum, and lumber. Coffee, cocoa, and cocoa by-products together account for almost half of the total value of exports. Aluminum, cotton, and bananas are other important exports. Capital equipment and semifinished products for the growing industrial sector have become leading imports. Other imports include pharmaceuticals, beverages, and synthetic fabrics.

Most trade is carried on with the EEC countries. France is the largest individual trading partner; it supplies almost half of Cameroon's imports and takes about one-fifth of the country's exports.

Transportation. The difficult terrain and heavy rainfall in the south have been contributory factors to the absence of an adequate transportation network linking the different regions of the country. The construction of the railway line between Mbanga and Kumba in 1965 enabled the transportation of products from eastern Cameroon to the port of Douala in western Cameroon, and subsequent expansion improved contact with the north. Transportation is more developed in some regions than in others; the best roads are in the coastal region, whereas the roads in eastern Cameroon are few and are often in bad condition.

Problems of transport development

Most of the roads can be traveled only during the dry season. Douala is linked to the large population centres by roads that run through the central region to Yaoundé, Yoke Yokon, and north to the Adamawa Plateau. Douala is also linked by road through Bertoua and Batouri to Bangui in the Central African Republic and to Gabon via Edéa and Kribi.

Railways include the Trans-Cameroon line, which runs from Doula to Yaoundé, and, since 1974, on to Ngaoundéré in the north. The two short railway lines of metre gauge in the southwest, comprising the West Line, were constructed by the Germans before World War I; they remained separated until 1955 when the Wouri Bridge made it possible for them to use a common terminus and to gain access to workshops at Bassa, just outside of Douala. A line to the western highlands runs from Douala to Nkongsamba and serves the coffee-and-banana region of western Cameroon; a line from Mbanga to Roumbou was completed in the late 1960s. A branch line off the track between Douala and Yaoundé runs to Mbalmayo on the Nyong River, which serves the country's major cocoa-producing area.

Port facilities at Douala

The main port is Douala, on the estuary of the Wouri River. One of the best-equipped ports in western Africa, it has docks for cargo ships, including a wood-loading dock and a tanker dock with adjacent facilities for the unloading and storing of minerals. The fishing port above the wood-loading dock is equipped with refrigeration facilities.

The minor ports include Kribi at the mouth of the Kienké River, which is used primarily for the shipment of logs and cocoa from the interior; the ocean port of Victoria in western Cameroon, which has become Cameroon's petroleum exporting port; and the port of Tiko, on a creek leading to the Wouri estuary, which handles bananas, wood, and rubber.

Douala is the main international airport, and Yaoundé and Garoua also handle international flights. There are domestic airports at Maroua and Ngaoundéré, as well as numerous airfields. Cameroon Airlines, owned by the government and Air France, operates all scheduled domestic flights and provides service to other African and to European countries.

Administration and social conditions. *Government.* By the constitution of 1961, the states of West Cameroon and East Cameroon were linked together into a federation. The constitution of 1972, subsequently revised, replaced the federation with a centralized government, the United Republic of Cameroon. Executive powers are conferred on the president, who is the head of the government and chief of the armed forces. The president is elected for a period of five years by direct and secret universal suffrage granted to all citizens over 21 years of age. A constitutional reform in 1984 abolished the office of prime minister and changed the name of the country to the Republic of Cameroon.

The federation of the two states

Legislative power is held by a unicameral National Assembly. It has 180 members, directly elected for five-year terms. The National Assembly shares with the president the initiative for proposing legislation, which it adopts on a simple-majority basis. When a particular bill is read twice, however, a majority of all numbers is required.

The republic is divided into 10 provinces, each administered by a governor. Each province is further divided into departments.

Evolution of the one-party state

In the 1950s there were no fewer than 84 political parties in East Cameroon, virtually all of which existed only at the tribal or regional level. Government action, terrorism, the imprisonment of opposition leaders, and mergers between parties in the 1950s and early '60s resulted in the creation of the Cameroon National Union (UNC). The country is now a de facto one-party state. In 1985 the UNC was renamed the Cameroon People's Democratic Movement (RDPC).

A Higher Judicial Council, which along with the president is responsible for guaranteeing the independence of the judiciary, advises the president on the nomination of magistrates and judges and acts as a disciplinary body. The legal system of Cameroon consists of the Supreme Court, two courts of appeal, and high courts as well as circuit courts. The Supreme Court decides whether a bill is receivable by the National Assembly in the event of a dispute between the president and the legislature. It also

passes judgment on appeals concerning administrative actions of the government and decisions of the Court of Appeal. The Court of Impeachment passes judgment on the president in case of high treason and on other government ministers in the event of a plot against the government.

Education. Educational services have expanded. About three-fourths of all children of primary-school age are enrolled either in government or in Christian mission schools. This attendance rate is not constant throughout the country, however, because the availability of school facilities varies regionally; virtually all children in the south may attend school, but adequate facilities exist for only about one-fifth of the children in the north.

There are general-education secondary schools, vocational schools, and teacher-training schools. Manual labour is compulsory in secondary and technical schools as a means of encouraging graduates to take up farming instead of seeking white-collar jobs in the cities. The University of Yaoundé was established in 1962 and in the early 1980s added four regional campuses.

Health and welfare. Malaria is prevalent everywhere except in the mountainous regions, where respiratory and pulmonary diseases and dysentery are common. There are incidences of leprosy and schistosomiasis, as well as syphilis, sleeping sickness, and rheumatism. The infant mortality rate, at about 100 per 1,000 live births, remains high by world standards, but is, nonetheless, comparatively low for western Africa.

The government emphasized the improvement of the nation's health facilities in the first and second five-year development plans and increased the number of hospitals, dispensaries, and elementary health centres about sevenfold. Hospitals in major cities were modernized, and in the late 1980s the country had one of the lowest population to hospital bed ratios in western Africa. A Health Sciences University Centre was established at the university in 1969 to train physicians and other medical personnel.

There is no government system of social security covering the whole population. Most assistance is obtained through the traditional kinship system. The National Social Insurance Fund, financed by employee and employer contributions, provides limited pension benefits for wage employees. There are, however, indemnities for occupational diseases or accidents, and the Public Health Service provides free services to the poor.

Cultural life. Each major ethnic group of the country has developed its own culture. The vigorous rhythms played on the drums by the people of the southern forest region contrast with the flute music of northern Cameroonians. In the Adamawa area, the Muslim Fulani produce elaborately worked leather goods and ornate calabashes (gourds used as containers), and the Kirdi and the Matakani of the western mountains produce distinctive types of pottery. The powerful masks of the Bali, which represent elephants' heads, are used in ceremonies for the dead, and the statuettes of the Bamileke are carved in human and animal figures. The Tikar people are famous for beautifully decorated brass pipes, the Ngoutou people for two-faced masks, and the Bamum for smiling masks.

Cultural institutions and organizations

L'Institut Français d'Afrique Noire (French Institute of Black Africa) maintains a library in Douala that specializes in the sociology, ethnology, and history of Africa. Of the several museums, the Diamare and Maroua Museum has anthropological collections relating to the Sudanese peoples, and the Cameroon Museum of Douala exhibits objects of prehistory and natural history.

Cultural organizations include the Cameroun Cultural Association, the Cameroun Cultural Society, and the Federal Linguistic and Cultural Centre. There are also numerous women's associations, youth organizations, and sporting associations.

For statistical data on the land and people of Cameroon, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (G.Be./Ed.)

HISTORY

Settlement by Europeans. The Portuguese navigator Fernão do Pó was probably the first European to sight the shore and estuary of the Wouri River in the 1470s. Dutch settlers followed the Portuguese into the region and were themselves succeeded by the English, who came with antislavery cruisers at the beginning of the 19th century. Liverpool and Bristol companies anchored pontoons to the banks of the Wouri to serve as shops, and in 1832 John Beecroft established a business at Bimbia. In 1849, when the Spanish occupation of the island of Fernando Po seemed likely, Beecroft was appointed consul and representative of the king to the Bights of Benin and Biafra and actively engaged in negotiations with local chiefs.

Activities of Alfred Saker

In 1845 the Baptist missionary Alfred Saker settled in Cameroon and on April 29, 1852, signed with Beecroft, King Akwa, and Charles Dido a treaty that regulated trade in the area and foreshadowed the abolition of slavery, the prevention of human sacrifice, the granting of religious liberty, and the protection of missionaries. In 1858, when the strictures of Roman Catholic Spain began to affect the Protestant Jamaicans in Fernando Po, Saker installed them at Ambas Bay (which he called Victoria) and founded many more settlements, notably at Deido and Bonabéri. He also succeeded in transcribing the Duala language.

In 1868 the Woermann Company of Hamburg built a warehouse on the estuary, an example soon followed by several other firms, and from 1874 (when the Duala were seeking British protection and the Malimba had already concluded a treaty with France) began to demand the appointment of a German consul. Despite the support of British businessmen, requests for British protection made by Duala chiefs in 1881 were forestalled by the German treaty of July 12, 1884, between King Bell and Gustav Nachtigal. An English annexation mission was similarly belated. German occupation marked by the appointment of Max Buchner as consul had little effect on Britain and France but locally took a far more serious turn. On December 15 the chiefs of Jos and Hickory, seeing themselves partially dispossessed, attacked Bell's village and took down the German flag. Three days later Admiral Knorr arrived in the *Bismarck* and took so stern a reprisal that Woermann's representative was murdered. Baron von Soden, who was appointed governor of Cameroon in mid-1885, set up the first administrative divisions and organized explorations. The French renounced their rights as far as the Campo River with the Berlin Agreement of Dec. 24, 1885, while on April 21, 1885, Britain agreed to the transfer of Victoria to Germany in exchange for St. Lucia in South Africa and Forcados in the Niger Delta.

The expeditions of such early explorers as Hugh Clapperton; Dixon Denham, who was probably the first to cross the interior in 1823; Heinrich Barth; Adolf Overweg, who visited Adamawa in 1851; and the Reverend George Grenfell, with those of their German and French successors, helped Germany to define new frontiers with Britain on Oct. 15, 1893, and with France on Feb. 4, 1894. Furthermore, in 1911, in return for assenting to France's protectorate over Morocco, Germany absorbed some 105,000 square miles of former French Equatorial Africa to constitute the territory of Neu Kamerun. When the territorial borders were defined in 1913, however, the governor had no means of enforcing his decisions. The recruitment of soldiers in Dahomey and Nigeria had serious effects, and the ensuing mutiny engendered the Kaiserliche Schutztruppe, a police force that dominated the country, although its western regions were not entirely subdued until the second half of the century.

Initially, chiefdoms were reinforced, a system of indirect administration was set up, and concessions were granted to commercial companies. The Gesellschaft Sud Kamerun governed the south and the Gesellschaft Nord-West Kamerun controlled the territories of the Bamileke and the Bamum. The number of officials increased, however, and direct administration to a large extent replaced indirect administration, which was maintained in the Muslim regions of the north.

The 20th century. In World War I joint action by French and British troops forced the Germans to retreat to Spanish Guinea. Early in March 1916 a temporary partition, which gave France about nine-tenths of the territory, was agreed upon, and the section adjoining Nigeria came under British control. On July 10, 1919, the Cameroons

League of Nations Mandates (1922)

was formally divided by Great Britain and France. In 1922 the League of Nations conferred mandates for the administration of the two zones, subject to the supervision of the League's Permanent Mandates Commission, and on July 20, 1922, the mandates were approved by the Council of the League of Nations.

Although the administration of the British Cameroons was integrated with that of Nigeria, French Cameroun constituted a separate entity and the former German Neu Kamerun was returned to French Equatorial Africa.

British Camerouns. The British trust territory consisted of a strip of land bisected by the Bénoué River along the eastern border of Nigeria. The capital of Southern Cameroons was Bouéa while Northern Cameroons comprised the Tigon–Ndoro–Kentu region, the Adamawa districts, and Dikwa in the extreme north.

Although the area allotted to Great Britain in 1919 was small, it nevertheless contained most of the German plantations (bananas, rubber, palms, and cocoa). These were reclaimed by their owners in 1924 but were sequestered at the outbreak of World War II. After World War II the Nigerian government absorbed them into the Cameroons Development Corporation, which employed several thousand workers and played an important part in the prosperity of Southern Cameroons.

French Cameroun. The administrative system of French Cameroun was based on that of other territories in French Equatorial Africa. Residing in the capital, Yaoundé, a commissioner (later high commissioner) with the rank of governor, assisted by a secretary-general, initiated and coordinated all governmental activities. The country was divided into 18 administrative regions which were themselves made up of further subdivisions. Its elected territorial assembly exercised wide powers. Personnel remained in the country for some time, thus assuring administrative continuity. Sleeping sickness was eradicated. Education along French lines was vigorously pursued. During World War II the territory joined the Free French Movement in August 1940, and Douala was for a time the headquarters of General Charles de Gaulle.

On Dec. 13, 1946, the first General Assembly of the United Nations approved Cameroun's status as a "trusteeship" within the French Constitution of 1946, which instituted Arcam (Assemblée Représentative du Cameroun). This body approved taxes and controlled almost all aspects of the territory's life. As the electoral body increased, from 16,088 in 1946 to 685,059 in 1956, this assembly, composed of 40, then 50 more members, became the voice of the people and the representative of diverse ethnic groups. In 1952 it was transformed into Atcam (Assemblée Territoriale du Cameroun) and in 1957 was reformed into Alcam (Assemblée Législative du Cameroun).

Independence. From 1941 the municipal regime applied to two urban communities, then to five others in 1947 and 1950, and in 1952 was extended to all urban and rural areas of Southern Cameroun in an initiative unique in Africa. Cameroun's constitution, approved by a decree of the Legislative Assembly on April 16, 1957, provided for the transfer of power to a legislative assembly, with finance, defense, and the diplomatic service the remaining responsibilities of the tutelary power. The government presided over by André Marie Mbida wielded executive power. In March 1959 the UN adopted the French proposal to end its trusteeship, and on Jan. 1, 1960, the territory became the Republic of Cameroun.

Republic of Cameroun. Political life did not really begin to take shape until the end of World War II when first Jeucafra (Youth of French Cameroon), then Unicafra (Union of French Cameroon), later Racam (Cameroonian Rally), were formed. In April 1948 the Cameroun section of the African Democratic Rally (RDA) took the title of the Union of the Populations of Cameroon (UPC). In 1955 an extremist splinter group of the UPC led by Ruben Um Nyobé whipped up unrest in Douala but failed to prevent the northern peoples from regrouping into the Cameroonian Democratic Bloc (BDC) of Ahmadou Ahidjo. In 1958 Ahidjo founded the majority party Cameroonian Union and in February 1958 assumed governmental control. After independence had been achieved on Jan. 1, 1960, the constitution was approved by 60 percent of the electorate on Feb. 21, 1960, and on April 10 the National Assembly was elected with Ahidjo's party in majority.

The Federal Republic. In the British Cameroons the question of reunification was put to a plebiscite. On Feb. 11, 1961, the northern regions voted in favour of joining the Federation of Nigeria, and the Southern Cameroons decided to join the Cameroun Republic, which then, on Oct. 1, 1961, became the Federal Republic of Cameroon.

(R.Co./Ed.)

In 1966 the Cameroonian Union merged with the Kamerun National Democratic Party and four opposition parties to form the Cameroon National Union (UNC), and Ahidjo assumed leadership of this new party.

In 1972 a referendum approved a new constitution and the formation of the United Republic of Cameroon. The federal system was abandoned for a centralized form of government. Maritime border disputes with Nigeria over the Rio-del-Rey offshore oil fields occurred in 1981 but were settled by the Organization for African Unity. In 1982 Ahidjo resigned, naming Prime Minister Paul Biya as president. A military coup threatened Biya's government in April 1984, but most of the army remained loyal and the government survived. Unofficial death counts from the attempted coup ranged from 500 to 6,000, and the government responded with a series of arrests and executions. In the following year, however, Biya consolidated his power and reconstituted the UNC as the Cameroon People's Democratic Movement (RDPC). He also announced that exiled members of the banned UPC were free to return home, but the UPC remained banned. In the same year, the official name of the country was changed to the Republic of Cameroon, its original name.

Biya's presidency

A major natural disaster struck Cameroon in August 1986, when clouds of toxic gas of natural origin emanated from Lake Nyos, near Wum. More than 1,700 deaths were caused by this incident, and more than 3,000 people were either injured or evacuated.

In 1988 Cameroon held its first multicandidate elections since independence, although the RDPC remained the sole legal party. In this election, President Biya officially won 98.75 percent of the vote, but many incumbents lost their seats in the National Assembly, which was expanded from 120 to 180 members.

For later developments in the history of Cameroon, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (Ed.)

Côte d'Ivoire

The Republic of Côte d'Ivoire (République de Côte d'Ivoire; also called the Ivory Coast) is a republic on the coast of western Africa. With a coastline more than 300 miles (480 kilometres) long, it forms an almost square block of territory with an area of 123,847 square miles (320,763 square kilometres). It is bounded to the north by Mali and Burkina Faso, to the east by Ghana, to the south by the Gulf of Guinea, to the southwest by Liberia, and to the northwest by Guinea. The de facto capital is Abidjan; the administrative capital designate is Yamoussoukro.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* The ground rises constantly as it recedes from the coast, and the northern half of the country consists of high savanna lying mostly 1,000 feet (300 metres) above sea level. Most of the western border with Liberia and Guinea is shaped by mountain ranges, whose highest point is Mount Nimba (5,748 feet [1,752 metres]), which is situated where the borders of the three countries meet.

Natural regions

One of four natural regions, the coastal fringe consists of a strip of land, no more than 40 miles wide, studded with lagoons on its eastern half. Access from the sea is made difficult by the surf and by a long submarine sandbar. Behind the coastal fringe lies the equatorial forest zone that, until a century ago, formed a continuous area more than 125 miles wide. It has now been reduced to an area roughly triangular in shape, with the apex lying a little to the north of Abidjan and with its base lying along the

Liberian border. The cultivated forest zone, which lies to the east of this triangle, consists of forest land that has been partly cleared for plantations, especially along the Ghana border and in the area around Bouaké. The fourth region, the northern savanna, consists of a sparsely populated plateau, offering open ground favourable for stock breeding. About 4,500 square miles in this region have been set aside to form the Komoé (Comoé) National Park.

Drainage. Apart from the Cavalla River, which forms most of the border with Liberia, major rivers from west to east are the Sassandra, the Bandama, and the Komoé, all of which drain southward into the Gulf of Guinea. Because all are broken by numerous falls and rapids, their value for transportation is minimal.

Soils. The forest soils of the south tend to lose their fertility because of excessive leaching and to turn into laterites, but the poorly drained, swampy soils, also found mainly in the south, more readily maintain their fertility and their rich yellowish silico-argillaceous character. In the savanna areas, crustlike "shields," formed as a result of rapid evaporation, alternate with rich black silico-argillaceous soils.

Climate. Equatorial and southern savanna types of climate prevail. North of about 8° N latitude, the southern savanna type of climate occurs, characterized by the parching wind known as the harmattan, which blows from the northeast from December until February. The dry season lasts from about November to March. There is a single rainy season, and the annual total rainfall amounts to approximately 45 inches (1,100 millimetres) in the northeast and centre to 60 inches in the northwest. The region is drier than the rest of the country and, because of the altitude, somewhat cooler. South of 8° N latitude, two rainy seasons occur, and three climatic subdivisions may be discerned. In the coastal fringe, rain falls mostly from May through July and to a lesser extent in October and November, averaging an annual total of about 77 inches at Abidjan; considerable variations are, however, experienced at different places along the coast. Average monthly temperature variation is small, and diurnal temperatures range from about 70° F (21° C) to the low 90s F (mid-30s C). In the forest zones and in the southern part of the savanna region, the rainy seasons are less pronounced. Diurnal temperatures vary between about 60° and 102° F (16° and 39° C), and the relative humidity is often high. On the mountains further west there is no dry season, and rainfall amounts to about 80 inches.

Plant and animal life. The tropical rain forest in the south contains valuable timber species, including African mahogany and iroko (or African teak). An important afforestation centre is the Banco National Park near Abidjan.

The animal life of the forest zone differs little from that of adjoining Ghana, although the larger ungulates (hoofed mammals) are lacking, with the exception of the bongo (a reddish brown antelope) and the forest buffalo. There are also about six kinds of dwarf antelope, ranging from the royal antelope to the yellow-backed duiker; the giant forest hog is widespread (although nowhere common), and the red river hog is locally plentiful. Manatees (herbivorous water animals) probably survive in some rivers. To the north, the savanna woodlands have some 10 species of antelope, as well as lions and occasional herds of elephants. In addition to Komoé National Park in the northeast, which is well stocked with wildlife, the Taï National Park, near the Liberian frontier, is notable for its pygmy hippopotamuses.

© Georg Gerster—Comstock, Inc.

Gallery forest along the Bandama River, near Korhogo, Côte d'Ivoire.

Settlement patterns. In the southeastern quarter of the country, most people live in compact villages and towns. The houses are rectangular dwellings of reeds, poles, or dried clay, traditionally covered with thatched roofs, though often today with corrugated iron sheets. The town centres grow quite lively when, every four days, the markets are held. There, the women are the sellers of the produce, which consists mostly of yams—the most basic national staple—corn (maize), the starch root cassava (manioc), peanuts (groundnuts), oil-palm nuts, and other vegetables. Fishermen on the lagoons ply their trade separately, maintaining their own markets. The entire area is divided into petty states with kings and an elaborate hierarchy of ministers and palace officials bearing elegant titles.

Rural life and regional characteristics

Among the Kru and other peoples of the southwestern forest zone, houses may be either rectangular or round, varying according to place rather than to tribe. Dwellings everywhere are clustered around a central open area, though markets are held only in a very few privileged towns. In other places, the central area is an evening meeting place as well as a spot in which "village democracy" is practiced by councils of elders. Women perform the bulk of daily work, both at home and in the fields, where they grow such crops as rice, cassava, bananas, yams, and corn. Rice is a comparatively recent introduction that has not yet been accepted by all. The men engage in hunting, gathering kola nuts and oil-palm nuts, and—on the coast—fishing.

The Malinke people of the northwestern corner of the country, as members of the Mande group, are inheritors of a culture made famous in the 13th to the 16th century by the Mali empire. Long before then, the Mande had effected a regional agricultural revolution, discovering the use of millet, which remains their staple food, and later introducing such other cereals as sorghum and corn. They have cultivated cotton for centuries. Cattle are kept by everyone, but for purposes of prestige and use on ceremonial occasions rather than for economic reasons; little milk is drunk. The men raise livestock and cultivate crops; they may also travel extensively for trade. The Malinke build round huts of mud and sun-dried brick surmounted by a conical thatched roof. Defense is a traditional concern, as is evident from the fences built around dwellings clustered in compounds and from the palisades surrounding large villages or towns. The people recognize a dual authority: on the one hand, the village chief, on the other hand, the chief representative of the linear descendants of the first settlers, a group forming a traditional nobility. Some people are born into certain trades, such as that of musician or *griot* (a historian minstrel).

The rest of the savanna is part of the domain of the Voltaic peoples, many of whom live in neighbouring Burkina Faso. Among them, the Senufo live immediately east of the Mande and have adopted many of their customs. Life in comparatively large villages and specialization on a hereditary basis may have helped them to reach their high level of artistic creation in wood carving and in weaving. All other Voltaic communities are split into dispersed homesteads. In the northeastern corner, some houses are of a peculiar type, found as far away as northern Benin. These are rectangular mud or brick structures crowned with crenellated parapets built around a flat roof, so that each house has some resemblance to a fairy-tale castle. Millet and sorghum are the staple food of all Voltaics;

the men do most of the work. All the people keep cattle, but only those who give them for keeping to outside herdsmen of Fulani origin ever taste milk. The Voltaics are great traders, distinguishing between local-market trading, which is conducted by women, and outside trading, which is conducted by the Dyula (Dioula), a subtribe of the Mande. Each community is run by the head of the main lineage group, who seeks above all to mediate in all disputes so the earth may never be defiled by blood spilling. In addition, the Senufo have chiefs who govern small districts.

One of the many trading ports built by Europeans along the African coast, Abidjan nevertheless has distinctive features of its own. A most striking one is its location on a lagoon, rather than on the sea. Its core is divided by a branch of the lagoon into Abidjan-Plateau, the first European settlement, to the north, and Treichville, the first large African settlement, to the south. The maintenance of physical communications between these quarters is one of the main problems facing the city.

Urban life

Plateau was recommended for settlement as early as 1898, but no Europeans actually lived there until 1903. Treichville, located behind the fishing village of Anoumabo, owes its importance to the boom in colonial trade that followed World War I. It remained a very small town until 1934, when the seat of colonial government was moved to Abidjan from Bingerville. Treichville, left to African homebuilders, assumed an appearance that still keeps alive the flavour of those comparatively leisurely times. After 1950, when the opening of the 1.7-mile Vridi Canal to the sea marked a new era of economic expansion, urban growth was exceptionally fast. Treichville, connected with Plateau by bridge, reached the saturation point with a gain of 150,000 inhabitants within a decade. Urban growth after 1960 was in all directions. Comprehensive planning was rendered impossible due to the many confining branches of the lagoon waters.

The first planned extension consisted of building a colonial army camp north of Plateau. There, and beyond, Adjamé and Attiécoubé offered an abundance of so-called moderate-rent dwellings, decent at first but fast deteriorating and inconsistent in design with all African traditions of family life. Across the small bay east of Abidjan, Cocody grew up in isolation as an area of expensive housing (including the presidential tower mansion) with a tourist centre with two hotel complexes offering facilities such as gambling and even ice-skating.

Petit-Bassam Island, where Treichville lies, also contains the settlements of Marcory and Koumassi, beyond which Port-Bouët grew up on the seashore, eight miles southeast of Plateau. Squatters helped in developing Yopougon-Attié and Abobo across the bay to the west, and in 1986 Greater Abidjan was organized into these 10 municipalities (each one with an elected council and mayor).

The people. *Ethnic groups.* There are more than 60 tribes, traditionally independent from each other, though larger groups among them may be recognized on the basis of cultural unity. Each one of these groups has tribal affiliations with larger groups living outside the borders of the republic. Thus the Baule, as well as other peoples living east of the Bandama River, are affiliated with the Akan group of Ghana. The lagoon fishermen farther south also have tribal brothers belonging to the same Akan group. The forest people west of the Bandama belong to the same group as the Kru boatmen of Liberia. In the interior, the Kru group is subdivided into tribes tiny in number but scattered over large areas of the forest and kept together by secret societies.

The savanna peoples may be divided into two main groups. The Mande group, which is particularly strong in Mali, is represented by the Malinke farmers and by the Dyula peddlers. The Voltaic group comprises the Senufo as well as the Lobi and Bobo subgroups, who live widely scattered over the northeastern region and across into neighbouring states.

Linguistic composition. All tribal languages belong to one of three subgroups of the Niger-Congo family: Kwa in the south, Mande in the northwest, and Voltaic in the northeast. A trade language, known as Dyula-Taboussi and akin to the Mande Bambara, is spoken throughout the country by Muslim traders, and *français de Moussa* is popular in Abidjan, but the official language is French.

Religions. Tribal cults continue to predominate among rural communities, but Islām claims the adherence of one-quarter of the population, mostly in the northwest and in Abidjan. The growth of Christianity suffers from the quick rise and fall of prophetic sects, but almost one-third of the population is Christian, mostly Roman Catholic or Methodist.

Demographic trends. Côte d'Ivoire has one of the highest population growth rates in sub-Saharan Africa and in the world. Its high rate of natural increase together with the huge influx of immigrants from the impoverished countries to the north, which its comparatively strong economy attracts, have been the main reasons for its rapid growth. Immigrants constitute about one-fourth of the total population, and, since a large percentage are men, the country has an unnaturally high male-to-female sex ratio. Nearly one-half of the population lives in urban areas, the largest such figure in western Africa, and among the urbanites there is a large French community as well as numbers of Lebanese and Syrians.

The economy. Côte d'Ivoire has a good financial reputation, which it maintained in the 1980s when the government agreed to reschedule its debt over a period from 1993 to 2002, including sums that had benefited from earlier agreements. Ivoirian policy is fundamentally liberal, and investments are welcomed through tax exemptions and legal protection against nationalization. Increased privatization became government policy in the mid-1980s, mainly owing to the fact that the government had participated in too many specialized undertakings in trying to diversify the economy. Previous plans have been revised with the aim of securing self-sufficiency in food and obtaining equipment in exchange for exports rather than by borrowing. In the long run, success will depend on avoiding luxuries and expanding the local market.

Resources. Côte d'Ivoire is primarily noted for its forest resources; it is a major exporter of tropical wood, but timber now follows cocoa and coffee as an export from one of the more developed western African economies. Mineral resources are rather scanty, but offshore petroleum reserves are being exploited.

Agriculture. The forest, even in its present reduced state, remains the most considerable asset of the country. About 30 species of trees are of high commercial value. The most important timber types are sipo (utile) and sambu (obeche), and reforestation has begun at numerous locations.

The forest floor, after clearing, provides a rich soil for the cultivation of edible roots and bananas, as well as of such commercial tree crops as coffee, cocoa (cacao), and rubber. The savanna soils are good for rice and other cereals. Cotton and sugarcane grow in both areas.

Agriculture provides a livelihood for about 60 percent of the labour force, and locally grown subsistence crops meet most rural domestic needs, but urbanization and the growing use of hired labour throughout the country created a demand for foodstuffs other than yams, cassava, plantains, and corn. An acquired taste for bread and beer has led to significant imports of wheat.

Local and export crops

Cocoa has become the main export crop, cultivated by more than one-quarter of the population, but in the late 1980s, after overtaking Ghana in cocoa exports, Côte d'Ivoire faced a price collapse combined with increased competition from Southeast Asia. Coffee, though fallen to second place in export value, remains a favourite business for almost all families in the southeast. Though the local coffee is a robusta of low quality, it constitutes a safe investment, enjoying a privileged position on the French market thanks to low production costs and much publicity. Close to the sea, thousands of acres have been planted with coconut trees to increase the production of copra, the dried kernel from which coconut oil is extracted. The same area is also suitable for pineapples, a valuable export crop.

The southwest provides good soils and climate for oil palm and rubber trees. A South American species of hevea rubber tree was introduced in the early 1960s, and the

cultivation of palm trees for oil was promoted at about the same time. In the north, cotton planting was fostered by teaching the use of higher-yielding varieties and the practice of cotton-rice and cotton-yam crop rotation.

Livestock raising prospers in the northeast, but national needs are also met by imports from Mali and Burkina Faso; Côte d'Ivoire is the largest market for the former and the second largest for the latter country. Fishing is a traditional occupation in the lagoons, supplying one-half of the catch for the local market, but, though the country has become the largest African exporter of tuna, an old distaste for sea fish lessens local consumption.

Industry. Mining has not been regarded as a promising field for industrial development in Côte d'Ivoire. There are known reserves of copper, nickel, uranium, and manganese. Ivoirian territory extends up the slopes of Mount Nimba, well-known for iron-ore deposits that enrich both Liberia and Guinea, but in Côte d'Ivoire, mining was until recently limited to the exploitation of diamonds on the Bou tributary of the Bandama River. The extraction of some alluvial gold was started in the late 1980s, but even then petroleum remained the only major object of interest in mining, with three offshore fields being exploited that may satisfy local needs for the future.

The Ivoirian industrial sector retains much of the legacy of a colonial policy founded on export rather than the much-needed expansion of the local market. As independence meant the end of the French West African federation, Dakar, Senegal, lost its status as the federal capital, and many firms, mostly French but also Lebanese, shifted their headquarters to Abidjan. More than 700 industrial companies were registered in the mid-1980s. More than half of them were kept at a low level of activity, however, owing to the reluctance to invest capital locally and to the competing attractive offers of employment made by big firms to skilled labourers. Nevertheless, the country became one of the best-equipped in western Africa.

Processing and manufacturing

Ivoirian industry rests on agricultural progress, historically starting between the two world wars, in the development of timber, cotton, cocoa, and coffee for export. More crops were later added to these—among which pineapples became an outstanding success—as local canning and preserving facilities developed. Palm oil also benefited from progress in equipment, which led to the production of fine edible oils and of soap. Timber was used for furniture, cotton fabrics for garments, and sisal for string, and imported raw materials were shipped to local bakeries and breweries.

The increasing segregation of urbanites from the rest of the Ivoirian population created other new demands. Workers in Abidjan started drinking soluble coffee and cocoa with internationally known labels rather than letting the traditional female vendors roast local beans before their eyes. At an even more sophisticated level, refrigerating and air-conditioning units and automobiles are made or their parts assembled, although the high cost of fuel and electricity continues to restrict their use and their export.

Power is supplied in Abidjan by two hydroelectric and two oil-burning stations, but the main achievement in this field is the dam and power station erected at Kossou, near where the two branches of the Bandama merge and close to Yamoussoukro, the new political capital. All four main rivers of the country may be harnessed through existing projects.

Finance. Côte d'Ivoire stands on relatively firm financial ground. Many foreign and domestic banks and credit institutions and real-estate agencies are in business. The National Bank for Agricultural Development and the Côte d'Ivoire Bank for Industrial Development have headquarters in Abidjan, where there are also branches of major British and U.S. banks and the Central Bank of West African States. The monetary unit is the CFA (Communauté Financière Africaine) franc.

Trade. Exports are reasonably diversified—though mostly agricultural—with France, The Netherlands, and the United States almost even in the amount purchased, but the country depends on France for as much as one-third of its imports. Côte d'Ivoire benefits particularly from the Lomé agreements, which made it one of the associates of the European Economic Community (Common Market), but too much dependence on the former "mother country" has contributed to its balance-of-payment problems. The stability brought by a currency attached to the French franc sometimes hampers exports to less privileged neighbours.

Transportation. A single-track railway line connects Abidjan with Ouagadougou, the capital of Burkina Faso. Still vital to the latter country, the railway has lost much of its importance to Côte d'Ivoire, and its joint-management board was disbanded in 1987. The railroad had lost half of its traffic in goods in less than 15 years, mostly to the road network, one of the densest and best-maintained in sub-Saharan Africa, which focuses upon Abidjan. Paved roads have been extended to replace beaten-earth roads. A secondary system of dry-season roads feed the main roads. Daily local trade is still conducted along the innumerable tracks that crisscrossed the country long before the advent of Europeans.

Roads, harbours, and airfields

As western Africa's largest container port, Abidjan has separate docking accommodations for passengers, for goods requiring special care such as bananas, minerals, and petroleum, for fishermen, and for boatmen who transport goods by canoe. Other ports are Sassandra, Tabou, and San-Pédro.

Abidjan has a fully equipped international airport, located at Port-Bouët. It is used regularly by planes belonging to about a dozen foreign companies and by those of Air Afrique, a service maintained by several French-speaking African states. The national airline, Air Ivoire, serves airports and landing fields in the interior.

Administration and social conditions. *Government.* Côte d'Ivoire was proclaimed an independent republic on Aug. 7, 1960. As in most African countries, the president is vested with wide powers and may seek reelection at the end of each five-year term. The president appoints and presides over a Cabinet of ministers, several of whom are without portfolio and are styled simply "ministers of state." There is a single-house legislature, the National Assembly, with 175 members elected for five-year terms. An Economic and Social Council with 120 members acts in an advisory capacity. In 1983 Yamoussoukro was officially named the new national capital, but austerity measures have prevented the transfer of government functions and Abidjan remains the de facto capital.

Executive, legislative, and judicial branches

All judges are appointed by the president, and they render justice according to legal codes of French inspiration. There are trial courts located in Abidjan, Bouaké, and Daloa, and their judges may be assigned to 25 other towns or be called upon to constitute special labour and juvenile courts. The same three towns are visited by an assize court dealing with serious criminal offenses. Abidjan also has a court of appeals and a supreme court.

The old colonial subdivisions of the country were renamed *départements* and *sous-préfectures* ("subprefectures") as in France. The number of the former has increased from six to 34. Though all have elected assemblies known as general councils, they are headed by prefects who have extensive powers and the help of 163 subprefects at as many stations in the interior. Towns have elected municipal councils. In general, tribal authorities do not fit within such a regime, which is of French inspiration. Nevertheless, some chiefs, especially among the Akan group, have won elective positions.

The political system is controlled by the Democratic Party of Côte d'Ivoire (PDCI), the only authorized party. It originated as a league of African farmers founded at the end of World War II by Félix Houphouët-Boigny, who later founded the African Democratic Rally (RDA), a union of all parties with similar aims in the then French West African federation.

Education. Educational services expanded considerably after independence. In the late 1970s, a project of universal education, assisted by radio and television, brought attention from beyond the borders of Côte d'Ivoire. At that time almost all children had at least some schooling, but costs, maintenance, and, above all, staffing shortages created problems. In any case, the literacy rate is relatively high (the highest in western Africa). A university on the

French model has been open in Abidjan since 1964. There are a few private missionary schools, which enjoy great prestige, at the secondary level.

Health and welfare. Health services were comparatively good and leprosy and yaws were responding to strenuous control measures before the late 1980s, when the economic crisis made it hard to meet the needs of an exceptionally rapidly growing population. Modern hospitals are located in Abidjan, Bouaké, Daloa, and Korhogo.

Social problems have resulted from an urban population explosion, adding a new dimension to long-standing problems of unemployment and prostitution. Occasional shipments of illegal drugs are intercepted, but an average of 80 percent of such drugs have been found to be in transit from other countries. Juveniles roaming and peddling in the streets are a constant object of concern, but a thorough inquiry sponsored by Hortense Aka Anghui, the mayor of Abidjan–Port-Bouët, revealed a very low level of delinquency among them. The situation is complicated by the traditional solidarity, which makes petty thieving hard to control among relatives, tribal brothers, or people working together or living under the same roof.

Cultural life. The cultural milieu has remained split rather more completely than in other African countries between a maze of tribal cultures and a foreign intrusion that was sudden and almost exclusively French. Traditional arts continue to flourish everywhere, and the Abidjan museum offers a rich storehouse. The Senufo carve masks, decorate doors with esoteric signs, and dance to the slow, majestic rhythms of drums supported by xylophones. The mountaineers of the Man forest wear masks showing horrifying faces, and they dance at a quick pace governed by the sound of drums and led by stilt-walkers. Versatile Baule artists make fine gold jewelry and wooden sculptures. Their weavers use pulleys on their looms that are works of art in themselves.

Traditional arts

An Ivoirian literature in French was born in colonial times at the Ponty High School in Dakar, Senegal. One of its graduates, Bernard B. Dadié, became world-famous for autobiographical reminiscences in novel form. His schoolmates Goffi Jadeau and Amon d'Aby won a large local audience and many followers through their plays for a national theatre. A younger playwright, Zadi Zaourou, launched a chair in African literature at Abidjan University, and Ahmadou Kourouma, a Muslim, inaugurated a new era of the Ivoirian novel with *Les Soleils des indépendances* (1968; "The Suns of Independence"), first published in Canada. The stage reflects public opinion better than the press, which consists of about 20 periodicals, including two dailies, all published in French in Abidjan.

For statistical data on the land and people of Côte d'Ivoire, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (J.Co.)

HISTORY

Early history. Abundant archaeological evidence confirms the presence of early man in what is now Côte d'Ivoire. In the medieval era, groups in the north were drawn into the trans-Saharan trade networks of the Ghana and Mali empires. Islām arrived with Malinke merchants as trade expanded. Mali's collapse in the 16th century resulted in a great upheaval that sent waves of migrants southward, founding new kingdoms in the hinterlands of the forest zone. The original inhabitants were either displaced or assimilated by these new groups.

Precolonial kingdoms. Important kingdoms flourished in the precolonial period. In the savanna country, towns developed around communities of Dyula traders. Kong existed for several centuries before Sekou Ouattara and his sons established a new dynasty there in the early 18th century. In 1897, Kong was destroyed by Samory Touré, then involved in creating a new Muslim empire in northern Côte d'Ivoire. The Bouna kingdom was created in the late 17th century by Bounkani, an immigrant from Dagomba (Ghana). It, along with Kong, became a major centre of Islāmic learning.

The wars associated with the rise of Asante (Ashanti) in the late 17th century led to the migration of numerous Akan groups into the forest country of Côte d'Ivoire. Most powerful of the states established was the Abron kingdom of Gyaman founded by Tan Daté. It was conquered by Asante in the 1730s, and, despite numerous revolts, remained subject to it until 1875. In much the same circumstances the Anyi kingdoms of Indénié (Ndenye) and Sanwi were founded. A succession struggle in Kumasi, following the death of Asantehene Opoku Ware in 1750, caused the exodus of Queen Abla Poku (Aura Poku) and her supporters into north-central Côte d'Ivoire. They founded the Baule kingdom, remarkable for its blending of Akan and indigenous traditions.

Arrival of Europeans. Until the 19th century, European contact was confined to the coast, where French and Portuguese traders sought slaves and ivory. In the 1830s, Louis-Édouard Bouet-Willaumez began signing treaties with coastal chiefs, allowing France to build forts and trading posts. France withdrew in 1870, but private merchants remained. Arthur Verdier sent explorers north and imported the first coffee plants. By the 1890s, inland penetration by traders such as Marcel Triech-Laplène and military missions such as that of Captain Louis-Gustave Binger in 1887–89 resulted in treaties and French "protectorate" relationships with many groups.

As the European rush to divide Africa accelerated, France claimed Côte d'Ivoire as a colony in 1893. Borders were determined in 1898, following the capture of Samory Touré. Governor Gabriel Angoulvant began the military occupation in 1908. Imposition of forced labour and head taxes led to fierce resistance, especially among the Baule, Anyi, and Abe (Abbey), but France's superior weaponry triumphed, although the colony was not considered "pacified" until 1918. New revolts broke out as France conscripted thousands of Ivoirians to serve with 140,000 other western African soldiers in World War I.

French occupation and conquest

20th-century history. Following the war, concerted efforts toward economic development were taken. The railway was extended to Bobo Dioulasso, which, along with most of Upper Volta (now Burkina Faso), was attached to Côte d'Ivoire in 1933. Schools and health facilities were introduced, exploitation of the forests was intensified, and Africans were encouraged to plant cash crops for export. By 1939, Africans grew 90 percent of the cocoa and 80 percent of the coffee produced in the colony.

Forty thousand Ivoirians fought in the French army during World War II. Between 1940 and 1942 the colony, along with the rest of French West Africa, chose to remain under Vichy. Racist legislation, economic discrimination against African planters, increased forced labour, and a depression caused by Britain's naval blockade created enormous discontent. Educated Africans thus welcomed the subsequent Free French regime. In 1944 Félix Houphouët-Boigny and Auguste Denise formed the African Farmers Union (SAA), which, with Governor André Latrille's support, secured equal treatment for African planters. Houphouët-Boigny's all-African slate swept local elections in 1945. In 1946, with the Côte d'Ivoire part of the French Union, he was elected to the French Assembly, spearheading the law, which bears his name, to abolish forced labour throughout the empire. The present borders were set in 1947, when the north reverted to Upper Volta.

In 1946 Houphouët-Boigny helped found the African Democratic Rally (RDA), which sought equality for Africans. At first harshly repressed, the RDA achieved many of its goals. In the 1950s Houphouët-Boigny was a Cabinet minister in two French governments, and in 1960 he was elected president of the newly independent Côte d'Ivoire. Aside from two coup attempts in 1963 and 1973, the country, largely owing to Houphouët-Boigny's ability to reconcile opponents, has been peaceful and prosperous, maintaining good relations with France and with its neighbours.

For later developments in the history of Côte d'Ivoire, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (N.E.L.)

Equatorial Guinea

The Republic of Equatorial Guinea (República de Guinea Ecuatorial), on the west coast of Africa, is an independent

state. Formerly Spanish Guinea, a colony of Spain, it consists of Río Muni (also called Mbini), on the continent, and five islands: Bioko (formerly Fernando Po), Corisco, Great Elobey (Elobey Grande), Little Elobey (Elobey Chico), and Annobón. This fragmented republic has a total area of 10,830 square miles (28,051 square kilometres).

Continental Equatorial Guinea, with an area of 10,045 square miles, is a roughly rectangular territory bounded by Cameroon to the north and Gabon to the east and south. Of the small islands near the continental coast, Corisco measures six square miles, and Great and Little Elobey together are less than a square mile. Bioko, off the coast of Cameroon, has an area of 779 square miles. Annobón, a volcanic island, has an area less than seven square miles and lies south of the equator almost 400 miles (640 kilometres) to the southwest of Bioko.

The country obtained its independence on Oct. 12, 1968. The capital of the republic is Malabo on Bioko. Bata is the administrative capital of the mainland.

Equatorial Guinea is beset by regional differences, geographic isolation, a fragile economy, and a lack of trained personnel, in part a legacy from the colonial era.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief, drainage, and soils.* Half of the continental enclave is covered with forests. A coastal plain about 12 miles wide abuts on the coastal hills, which lead to inland plateaus (called *mesetas* in Spanish) that rise toward the frontier with Gabon. There are several ranges of hills. The central range divides the Benito River basin to the north from the southern basin of the Utamboni River. The Niefang-Mikomeseng range north of the Benito River is somewhat lower. All these ranges form segments of the Cristal Mountains in Gabon. The region is divided by the Benito River (known as the Woleu River in Gabon), which runs generally from east to west and is nonnavigable except for the first 12 miles inland. To the north the Campo River (called the Ntem in French-speaking Africa) marks part of the frontier with Cameroon.

In the south, the Muni is not itself a river but the estuary of various rivers of Gabon and southern Equatorial Guinea. To the east the de facto border with Gabon follows the meandering course of the Kié (Kyé) River rather than the legal frontier, which runs along a line 11° 20′ east of the Greenwich meridian. Except for limited hydroelectric power generation and the use of waterpower at some lumbering sites, the rivers of the mainland enclave are not exploited. The coast consists of a long stretch of beach with low cliffs toward Kogo to the south. Equatorial Guinea has no natural harbour, and Mbini and Kogo are no more than rudimentary ports of call for the ships that infrequently visit. Bata, on the other hand, has been enlarged artificially to become one of the country's main ports.

The coastal plain is overlaid by sedimentary deposits. The hinterland is composed primarily of ancient metamorphic rocks that have undergone a lengthy process of leaching and erosion, so that the resulting soils are relatively infertile. Exploration in the waters off the mainland has revealed some reserves of petroleum and natural gas, and there are prospects for their development. Gold, manganese, iron ore, and uranium are thought to exist in commercially exploitable quantities inland.

The main island, Bioko, is about 45 miles long and 22 miles wide. Its extinct volcanic cones, crater lakes, and rich lava soils form a contrast with the landscape of the mainland. In the north Mount Santa Isabel soars to a height of 9,878 feet (3,011 metres); this extinct volcano is the site of a television transmitter. In the centre of the island, Moca Peak and the Moca Heights present an alpine type of landscape. The southern part of the island, remote and scarcely developed, consists of the Gran Caldera range, which is rugged and indented by torrents and crater lakes.

Agricultural productivity

Despite its tortuous relief, Bioko can be productive agriculturally. Torrents are exploited for hydroelectric power; the Musola River provides electricity for much of Malabo. The coast is largely inhospitable, consisting for the most part of a cliff about 60 feet high, broken occasionally by small inlets and beaches. The southern coast is very steep and dangerous to shipping; San Antonio de Ureca, located along this stretch, is the most isolated settlement on the island. Malabo has a relatively good harbour, built on the partially sunken rim of a volcano.

Annobón is an isolated fragment of the republic, about 93 miles southwest of the island of São Tomé in São Tomé and Príncipe and about 400 miles southwest of Bioko. Like the latter, it is a volcanic island but is less high; it consists of a conglomeration of cones of which Mount Santa Mina (about 2,460 feet) is the highest. Not quite four miles long by two miles wide, it is a rugged island with only one settlement of note. The inhabitants are mostly fishermen who speak a Portuguese patois.

Climate. The climate of both the continental region and the islands is typically equatorial, with high temperatures, heavy rainfall, and much cloud cover most of the year. Local variations are due to differences in altitude and proximity to the sea. The wet seasons in the continental region are from February to June and from September to December. Rainfall is higher on the coast than inland. In Bata the rainiest months are September, October, and November, with rainfall averaging more than 94 inches (2,388 millimetres) a year. At Calatrava, farther south on the coast, it sometimes reaches 180 inches. Inland, however, rainfall diminishes; Mikomeseng, for example, receives only about 58 inches. The average annual temperature is about 79° F (26° C) and is fairly constant throughout the year. The temperature maxima are somewhat lower than in Bioko. The relative humidity, however, is higher than in Bioko.

Bioko has a rather debilitating climate. The so-called dry season lasts from November to March, and the rest of the year is rainy. The average annual temperature of about 77° F (25° C) varies little throughout the year. Afternoon temperatures reach the high 80s F (low 30s C) and drop to only about 70° F (21° C) at night. Most of the time the sky is cloudy and overcast. Extreme rainfall occurs in the south, with rain brought by monsoon winds amounting to about 450 inches a year around San Antonio de Ureca.

Plant and animal life. Continental Equatorial Guinea is covered by dense tropical rain forest that is exploited by the lumbering industry. More than 140 species of wood are found, of which the most important commercially are okume (*Aucoumea klaineana*), African walnut, and various mahoganies. Intensive exploitation close to the coast led the Spanish timber companies to venture deeper into the interior. Reforestation is minimal, and a secondary forest growth has replaced the virgin rain forest. African subsistence farmers clear the land by burning off the vegetation cover, after which they grow cassava, beans, yams, roots, and some cash crops such as coffee and cocoa, to the extent that the heavily leached soils allow. Mangroves fringe long stretches of the coast as well as riverbanks.

Bioko has a greater variety of tropical vegetation, including mangroves, and—at higher altitudes—vegetable gardens and pastureland.

Varied animal life

The continental region has a rich animal life that includes gorillas, chimpanzees, various rare species of monkeys, leopards, buffalo, antelope, elephants, hippopotamuses, crocodiles, and various species of snakes, including pythons. Insects, including the tsetse fly and the malaria-bearing *Anopheles* mosquito, as well as hosts of ants, beetles, spiders, and termites, abound. Bioko has no big game but has various monkeys, dwarf antelopes, and rodents, as well as the mosquito and other insects.

Settlement patterns. The mainland is sparsely settled by coffee and cocoa farmers who practice traditional methods of agriculture. During the colonial period, Roman Catholic missions did much to encourage the population to construct "corridor" villages by the sides of roads; in most villages the church and the school figure prominently. The region was never a settler colony, and the few European plantations—mostly Spanish or German—that survived the colonial era have been abandoned. After independence an exodus of Spanish technicians occurred, so that the mainland has largely reverted to a subsistence economy. The modern sector of the economy is represented by lumbering and coffee and oil palm cultivation.

Bioko, by contrast, is a plantation island; it retained

for several years a larger number of plantation owners and managers and consequently withstood longer than the continental region the effect of independence upon its economy. Before independence there were about 1,900 plantations (known as *fincas*), which ranged in size from 0.4 acre (one hectare) to more than 4,900 acres. The Bubi people, the indigenous population, live for the most part in mission villages in the northern part of the island on the lower slopes of Mount Santa Isabel, as well as in their traditional homeland, the Moca Heights. Unused to plantation labour, they gain their living mostly as small farmers and minor civil servants. The workers on the *fincas* were for many years migrant labourers from Nigeria, who served under contract. During the 1960s the Nigerian workers often brought their families, settling in numbers believed to have reached 50,000 to 80,000 by the end of the decade. Political and economic conditions after independence gradually reduced these numbers, despite an agreement with Nigeria in 1972 for the recruitment of new labourers. Reports of virtual slave-labour conditions on plantations and of repressive killings by authorities in the mid-1970s turned this gradual exodus into a flood, further impoverishing Equatorial Guinea's economy.

Nigerian migrant labour

Many of the formerly European-owned plantations were extensive; in 1962 more than 300 plantations occupied over 148,000 acres, leaving about 2,800 acres in the possession of 1,600 African farmers grouped in cooperatives. Some of the larger plantations, employing hundreds of Nigerian labourers and occupying the most fertile land, were local economic powers whose political influence, if subdued, was nevertheless felt by the local Bubi population.

Malabo, the national capital, is a small city standing behind its crater harbour. Created by the British in the 19th century, it was remodeled and developed by the Spanish. A rambling tropical city, it has a distinctly Spanish atmosphere—especially in the European district near the cathedral, the mission, and the government house. Farther inland, the African districts were inhabited mostly by Nigerian and other workers who first chose not to return to the mainland. Another town of some importance is Luba, on the southwest coast, linked with the capital by a good paved road that runs through a series of Bubi settlements. Basilé, on the slopes of Mount Santa Isabel, provides a cool refuge for heat-weary residents of the capital.

The continental region was settled much later by the Spanish, so that Bata, the main settlement, lacks the amenities of Malabo. Fang migrants from the interior have built new suburbs around this sprawling port city.

The people. *Ethnic composition.* The majority of the population is African, but its composition is complex for a political unit so small in size. The Fang people, who fought their way to the sea in the 19th and early 20th centuries by subjugating the weaker tribes in their path, form about 80 or 90 percent of the population of the mainland region. North of the Mbini River are the Ntumu Fang, and to the south of it are the Okak Fang. Holding political power on the mainland, the Fang tend to migrate to Bioko, where their leaders hold most of the levers of political control. Coastal tribes, such as the Kombe, Mabea, Lengi, Benga, and others, have been in contact with European traders much longer, and a limited amount of miscegenation has taken place, especially on the island of Corisco. Spanish ethnographers refer to these coastal tribes as *playeros* (literally, "those who live on the beach"). Both the Fang majority and the *playero* minority are Bantu.

The original inhabitants of Bioko are the Bubi, descendants of Bantu migrants from the mainland. Contacts with Europeans decimated them, and only a few thousand remained early in the 20th century. They became the most pro-Spanish element of the African population, viewing the end of Spanish rule as a signal for the invasion of their island by the Fang. Certainly, numbers of mainlanders, most of them Fang, have flocked to the island since the mid-1960s, seeking to join the civil or military forces or to receive political patronage. In addition to these two groups, there are Fernandinos, descendants of former slaves liberated by the British during the 19th century who mingled with other emancipated Africans from Sierra Leone and Cuba as well as with immigrants from other western African countries. Formerly constituting an influential bourgeoisie, they lost much of their status when the Spanish acquired the island and after independence. The inhabitants of Annobón are descended from slaves imported by the Portuguese when the island was a dependency of São Tomé; some of them now live on Bioko.

The Bubi

By about 1970, these different strata together constituted a minority on the island, the majority being formed by the Nigerian contract labourers, who lived in compact colonies in Malabo or on the plantations. The repatriation by Nigeria, however, of at least 45,000 workers beginning in 1975, following reports of repressive conditions in Equatorial Guinea, led to extensive realignment of the demographic, social, and labour structures of the island and, indeed, of the country. Additional communities on the island are formed by *crioulos* (of mixed Portuguese and African origin) from the islands of São Tomé and Príncipe; there are also some Cameroonians.

Linguistic composition. While each ethnic group speaks its own language, other linguistic influences are at work. The first is Spanish, the official language of the republic, which is taught in schools and used by the press and is the only means of communication common to both Bioko and the mainland. The second influence is pidgin English, which is used extensively in petty commerce and forms the lingua franca on Bioko. Third, as a result of closer economic association with Francophone countries begun in 1983, French became a compulsory subject in schools in 1988. A Portuguese patois is also spoken in both Bioko and Annobón.

Religion. While about four-fifths of Equatorial Guineans are nominally Roman Catholic, the Bubi and mainlanders often retain traditional forms of worship, which are emphasized in times of crisis. The Mbwiti cult on the mainland, banned by the Spanish authorities, still has adherents. Most churches were closed by presidential order in 1975, and the Roman Catholic church was banned in 1978. These orders were rescinded after the coup in 1979, but many denominations, notably Jehovah's Witnesses, were proscribed once again in 1986.

Demographic trends. The population was reduced by about one-third through the departure of some 110,000 citizens after 1968 and the repatriation of Nigerian workers by 1976. The rates of population increase, density, and life expectancy are lower than those of most other African countries.

The economy. *Agriculture and forestry.* Equatorial Guinea's economy has depended primarily upon three commodities—cocoa, coffee, and timber. Before independence, the Spanish subsidized cocoa and coffee exports to Spain. Cocoa, which is of high quality, was the mainstay of the economy of Bioko, which possesses the right soil and climate for its intensive cultivation. Most of Equatorial Guinea's cocoa is produced on the island. As with other commodities, production declined under the regime of Francisco Macías Nguema (1968–79) because Nigerian and local workers left the cocoa plantations and maintenance, output, and quality declined. Cocoa exports dropped to one-tenth of their former level, and exports of coffee almost ceased from island and mainland plantations, the small production of robusta coffee by Fang farmers in Río Muni alone being marketed. The timber companies are entirely European; there were at one time 30 timber concessions in operation on the mainland. The industry requires a large investment.

Cocoa, coffee, and timber exports

Bananas are grown on Bioko, where they are exported from the port of Luba. The output of palm oil from mainland plantations has been adversely affected by the unsettled state of the country.

Industry. There is little industry in the country. Some cocoa and coffee processing takes place locally on plantations and in African cooperatives. There is no mineral production.

Finance and trade. Following the economic collapse of the mid-1970s, imports came to exceed exports. The gap was narrowed only by external aid, which began to increase after 1979. Since independence the national budget has been balanced first by large Spanish subsidies and since 1981 by help from many nations and agencies.

A major economic reorientation took place in December 1983 when Equatorial Guinea joined the Customs and Economic Union of Central Africa (UDEAC). The country entered the Franc Zone in January 1985, acquiring a freely convertible currency and more French aid. French economic influence is growing.

Transportation. The road network on the mainland was adequate for the light traffic it was required to carry before independence, but it deteriorated in the 1970s. Bata is linked with Mbini by a tarred road. There is a cross-country road from Bata, branching at Niefang and Ncue, to Ebebiyin, Mongomo, and Nsoc near the Cameroon frontier. There are no railways. On Bioko the road system is of a higher standard, with a semicircular tarred road linking Malabo and Luba to the eastern Bubi villages.

The main port is Malabo, which is connected to Spanish and other European ports by regular steamer services. The harbour at Bata was enlarged and modernized in the 1980s and is expected to accommodate a growing share of the country's commerce. European Economic Community aid provided a small coaster in the mid-1980s to ply between Equatorial Guinean ports, Gabon, Cameroon, and São Tomé and Príncipe. There is an international airport at Malabo. In 1982 an international airline, Aerolíneas Guinea Ecuatorial, was formed, and construction began on a new international airport at Bata in 1983.

Administration and social conditions. *Government.* According to the constitution of 1982, the system of government is presidential, and all governing bodies are elected by universal adult suffrage. The Council of Ministers is appointed by the president and is responsible to him. The president is selected by universal suffrage for a seven-year term. The State Council includes the chairman of the House of Representatives, the president of the Supreme Tribunal, and the minister of defense. The Supreme Tribunal in Malabo is the highest judicial authority. There are also territorial high courts and courts of the first instance in Malabo and Bata.

In March 1970 all the political parties that existed before independence were merged into a Single National Party, which later changed its name to the Single National Party of Workers. The party was abolished after the coup of 1979, but exiled Equatorial Guineans formed several opposition parties.

In 1983, elections were held for membership in the 41-member House of Representatives. The president chose all candidates and each position was uncontested. In 1987 a new government party, the Democratic Party of Equatorial Guinea (PDGE), was formed. Some exiles were allowed to return in 1988, and a presidential election was envisaged.

Education. Education is compulsory for all children between the ages of six and 14, but only about half of the children attend school. The primary and secondary levels each consist of six years of schooling. Facilities for higher education are provided through Spanish assistance at Bata and Malabo. Efforts have been made to improve educational opportunities, but illiteracy remains a problem; only about one-third of the population is literate.

Cultural life. Despite a veneer of Spanish culture and of Roman Catholic religion that is thicker in Bioko than on the mainland, Equatorial Guineans live largely according to ancient customs, which have undergone a revival since independence. Among the Fang of the mainland, witchcraft, traditional music (in which the Fang harp, the xylophone, the great drums, and the wooden trumpet are used), and storytelling survive. Spanish aid is much oriented to educational and health services. Among the Bubi farmers of Bioko, some ancient customs are still followed.

Role of ancient customs

For statistical data on the land and people of Equatorial Guinea, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (R.P./R.J.H.-C.)

HISTORY

The island of Bioko (formerly Fernando Po) was sighted by Fernão do Pó, probably in 1472. At first it was called Formosa (Beautiful). Annobón was probably sighted by Ruy de Sequeira on a New Year's Day (hence the name, which means "good year") between 1472 and 1475, most likely that of 1474. By the Treaty of Tordesillas (June 7, 1494), the Portuguese had exclusive rights in Africa, and it was not until 1778 that they agreed to cede to Spain the islands of Annobón and Fernando Po as well as rights on the mainland coast between the Ogooué and Niger rivers. These cessions were designed to give Spain its own source of slaves in Africa for transport to Spanish America, where, in exchange, the Spanish confirmed the rights of the Portuguese west of the 50° W meridian in what is now Brazil. The Spanish were soon decimated by yellow fever on Fernando Po, and they withdrew in 1781. No occupation was made on the mainland.

British administration. After the British abolition of the slave trade in 1807, bases were required by the Royal Navy for the effective suppression of the trade. Fernando Po was unoccupied and lay in a strategic situation from which the Niger mouths and the Slave Coast could be watched for slavers. In 1827 the Spanish leased bases for this purpose to the British at Port Clarence (later Santa Isabel, now Malabo), a fine deepwater harbour on the north coast, and in San Carlos Bay on the west coast.

In the absence of the Spanish, the British also became responsible for administering the island. Thereafter, the British landed many freed slaves, in default of knowing their origin or of being able to repatriate them. Freed slaves also went to the island from Sierra Leone and Jamaica, and in the 20th century the descendants of these several groups continued to speak a form of English. Because of the existence of these freed slaves and the absence of any Spanish administration in the area, the United Kingdom made several unsuccessful offers to Spain for the purchase of Fernando Po—*e.g.*, from 1839 to 1841. In 1843 the Royal Navy concentrated its antislavery patrol at Freetown in Sierra Leone, and its buildings on Fernando Po were sold to a Baptist mission.

Arrival of freed slaves

Spanish Guinea. In 1844 the Spanish made a second effort at effective occupation of Fernando Po, and their first exploration of the mainland was carried out in the two decades ending in 1877. Meanwhile, the Spanish had expelled the British Baptists from Fernando Po in 1858, and in 1879 they began to use it as a penal settlement for Cubans. Following the Spanish-American War of 1898, Spanish Guinea remained as Spain's last significant tropical colony. Profiting from the weakness of Spain, France was able to confine mainland Spanish Guinea to its present limited extent. Economic development started only at that time and was concentrated on the richer and healthier Fernando Po. The mainland received significant attention from Spain only after the Spanish Civil War of 1936–39.

In 1959 the status of Spanish Guinea was changed, and the region was reorganized into two provinces of overseas Spain, each of which was placed under a civil governor. The citizens, including the Africans, were granted the same rights as those enjoyed by the citizens of Spain. In 1963 a measure of economic and administrative autonomy for the two provinces (which were henceforth known as Equatorial Guinea) was agreed on by plebiscite.

Independence. The movement toward independence began to take shape at the end of 1967. Early the following year the Spanish government suspended autonomous political control and, with the subsequent approval of the Organization of African Unity (OAU), proposed that a national referendum be held to approve the new constitution. The constitution was overwhelmingly approved on August 11 and was followed by parliamentary elections in September and by the proclamation of independence on Oct. 12, 1968.

The first president was Francisco Macías Nguema. After his election in 1971 he assumed wide powers and pushed through a constitution that named him president for life in July 1972. He assumed absolute personal powers in 1973 when the island of Fernando Po was renamed in his honour. He controlled the radio and press, and foreign travel was stopped. In 1975–77 there were many arrests and summary executions, which brought protests from world leaders and Amnesty International. The Nigerian government repatriated its nationals in 1976.

Macías was overthrown in 1979 by his nephew, Lieutenant Colonel Teodoro Obiang Nguema Mbasogo, who led a Supreme Military Council to which he added some

civilians in 1981. A less authoritarian constitution was instituted in 1982, followed by an election of 41 unopposed candidates to a House of Representatives in 1983.

Major efforts were made to revive political and especially economic activity, notably by admission to UDEAC in 1983, adoption of a convertible CFA (Communauté Financière Africaine) franc in 1985, acceptance of aid from many and varied sources, formation of a more open single party (PDGE), and the raising of hopes of more liberal government and revival of the economy and foreign trade.

For later developments in the history of Equatorial Guinea, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (R.J.H.-C.)

Ghana

Ghana is a western African republic situated on the coast of the Gulf of Guinea. It has an area of 92,098 square miles (238,533 square kilometres). Although relatively small in area and population, Ghana is one of the leading countries of Africa, partly because of its considerable natural wealth, and partly because it was the first black African country south of the Sahara to achieve independence from colonial rule and to embark on its own extensive educational and industrialization programs.

Ghana, which achieved independence on March 6, 1957, consists of the former British colony of the Gold Coast and the part of Togoland that was formerly a UN Trust Territory under British administration. It is bordered on the northwest and north by Burkina Faso, on the east by Togo, on the south by the Atlantic Ocean, and on the west by Côte d'Ivoire. The capital is Accra.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief and drainage.* Relief throughout Ghana is generally low, with altitudes nowhere exceeding 3,000 feet (900 metres). The southwestern, northwestern, and extreme northern parts of the country consist of a dissected peneplain—a land surface worn down by erosion to a nearly flat plain, later uplifted and again cut by erosion into hills and valleys or into flat uplands separated by valleys; it is made of Precambrian rocks (from 570 million to 3.8 billion years old). Most of the remainder of the country consists of Paleozoic deposits (from 245 to 570 million years old), which are thought to rest on older rocks. The Paleozoic sediments are composed mostly of beds of shales (laminated sediments consisting mostly of particles of clay) and sandstones in which strata of limestone occur in places. They occupy a large area called the Voltaian Basin in the north-central part of the country where the altitude rarely exceeds 500 feet. The basin is dominated by Lake Volta, an artificial lake that extends far into the central part of the country behind the Akosombo Dam and covers 3,275 square miles. Along the north and south, and to some extent along the west, the uplifted edges of the basin give rise to narrow plateaus between 1,000 and 2,000 feet high, bordered by impressive scarps. The most outstanding are the Kwahu (Mampong) Scarp in the south and the Gambaga Scarp in the north.

The Voltaian Basin

Surrounding the basin on all of its sides, except in the east, is the dissected Precambrian peneplain, which rises to elevations of 500 to 1,000 feet above sea level and contains several distinct ranges as high as 2,000 feet.

Along the eastern edge of the Voltaian Basin, and extending from the Togo border to the sea immediately west of Accra, is a narrow zone of folded Precambrian rocks running northeast to southwest, forming the Akwapim-Togo Ranges, which vary in height from 1,000 to 3,000 feet. The highest points in Ghana are found there, including Mount Afadjato (2,903 feet; 885 metres), Mount Djebobo (2,874 feet), and Mount Torogbani (2,861 feet), all situated east of the Volta River near the Togo border. These ranges are part of the Togo-Atakora Mountains, which extend northward into Togo and Benin.

The southeastern corner of the country, between the Akwapim-Togo Ranges and the sea, consists of the gently rolling Accra Plains, which are underlain by some of the oldest Precambrian rocks known—mostly gneisses (coarse-grained rocks in which bands containing granular minerals alternate with bands containing micaceous minerals); in places they rise above the surface to form inselbergs (prominent steep-sided hills left after erosion). The only extensive areas of young rocks less than 136 million years old are in the wide, lagoon-fringed delta of the Volta, about 50 miles (80 kilometres) east of Accra, and in the extreme southwest of the country, along the Axim coast.

In the east the predominant rocks are less than 65 million years old, though there is a patch of Cretaceous sediments (from 66.4 to 144 million years old) near the Ghana–Togo border. To the west of Axim, near the Côte d'Ivoire frontier, the rocks date to the Cretaceous period. The intervening coastal zone between eastern and western extremes contains patches of Devonian sediments (from 360 to 408 million years old). With the older and more resistant rocks of the Precambrian peneplain, these form a low, picturesque coastline of sandy bays and rocky promontories.

The drainage system is dominated by the Volta River basin, which includes Lake Volta. Most of the other rivers, such as the Pra, the Ankobra, the Tano, and a number of smaller ones, flow directly south into the sea from the watershed formed by the Kwahu Plateau, which separates them from the Volta drainage system. South of Kumasi, in the south-central part of the country, is Ghana's only true natural lake—Bosumtwi—lying in a deep basin without any outlet to the sea. It is believed to be of volcanic origin, although theories of a possible meteorite origin have also been advanced. Along the coast are numerous lagoons, most of them formed at the mouths of small streams.

The rivers

Over much of the surface of Ghana, the rocks are weathered, and great spreads of laterite (red, leached, iron-bearing soil) and lesser spreads of bauxite and manganese are found on the flat tops of hills and mountains. Although the movements of the Earth's crust that produced the basic geologic structure of the country have now virtually ceased, periodic earthquakes occur, especially near Accra along the eastern foot of the Akwapim-Togo Ranges, where there is a major fault line.

Soils. Throughout the country, weathering, leaching, and the formation of laterite hard pans (hard impervious layers, composed chiefly of iron and aluminum oxides cemented by relatively insoluble materials) by capillary movement (the movement of water containing mineral salts to the surface) and evaporation, are common processes that vary in importance according to the characteristics of each locality. Leaching is more pronounced in the wet south, while the formation of laterite is more widespread in the drier north. In general, most soils are formed in place from parent rock material that has been subjected to prolonged erosion and consequently has limited fertility.

Types of soil

In the forest zone the soils are mostly lateritic. They are subdivided into relatively fertile and less acidic ochrosols (red, brown, and yellow-brown, relatively well-drained soils) in areas of moderate rainfall and into more acidic and less fertile oxysols in the extreme southwest, where annual rainfall exceeds 65 inches. Ochrosols occur over considerable areas in the coastal and northern savanna zones. As in the forest zone, they are the best soils for agriculture.

The coastal savanna zone has an abundance of soil types, including tropical black earths, tropical gray earths, acid vleisols, and sodium vleisols. Except for the tropical black earths, known locally as Akuse clays, most of these soils are of little importance agriculturally. The Akuse clays fill a broad zone across the coastal savanna plains; although heavy and intractable, they respond well to cropping under irrigation and mechanical cultivation.

Because of their intrinsic poverty in nutrients, most of the soils are heavily dependent upon the humus supplied by the vegetation cover. There is, thus, a delicate balance between vegetation and soil fertility, which may be upset by uncontrolled burning or overuse.

Climate. Ghana's climate, like that of the rest of the Guinea Coast, is determined largely by the interplay of two air masses: a hot, dry continental air mass that forms over the Sahara and a warm, humid maritime tropical air mass that forms over the South Atlantic. Both air masses

move toward the equator with their hemispheric winds and meet at the Guinea Coast for several months each year. Continental air moves southward with the northeast trade winds, known in western Africa as the harmattan, and maritime tropical air moves northward with the southwest trades. The zone where these air masses converge is characterized by seasonal line squall rainfall. The convergence zone itself oscillates north and south, following the seasonal movements of the overhead sun and the thermal equator; it reaches its most northerly position in the central Sahara, about latitude 21° N, in August, and its most southerly position about 7° N, a few miles north of the Ghana coastline, in January. Rains occur when the dominant air mass is maritime tropical, and drought prevails when continental air and the harmattan dominate.

Rainfall and temperature

In the savanna country north of the Kwahu Plateau, there are two seasons—a dry season from November to March, with hot days and cool nights under clear skies, and a wet season that reaches its peak in August and September. The mean annual rainfall is between 45 and 50 inches (1,145 and 1,270 millimetres), but there is a marked moisture deficit because of the long, intensely dry season that follows. In the southern forest country, where the annual mean rainfall from north to south has a range of about 50 to 86 inches, there are two rainy seasons—one from April to July and a lesser one from September to November—and two relatively dry periods that occur during the harmattan season, from December to February, and in August, which is a cool, misty month along the coast. In the Accra Plains, anomalously low annual mean rainfall figures vary from 40 inches to less than 30 inches, and the rainfall variability and the vegetation bear close resemblance to conditions in the northern savanna zone.

Temperatures show much more regional uniformity. The annual mean temperature is from 78° to 84° F (26° to 29° C), and the daily range only some 10° to 15° F (6° to 8° C) along the coast, and some 13° to 30° F (7° to 17° C) in the north. Average relative humidities range from nearly 100 percent in the south to 65 percent in the north, although during the harmattan season figures as low as 12 percent have been recorded in the north and around Accra. Enervating conditions produced locally by the combination of high temperatures and high humidities are moderated by altitude in the higher parts and by land and sea breezes along the coast. In general, the hottest months are February and March, just before the rains, and the lowest temperatures occur in January or—along the coast—in August.

Plant and animal life. Although soils and biotic factors (*i.e.*, those pertaining to living organisms, including humans) are important, vegetation is primarily determined by rainfall. There are three principal types of vegetation from south to north occurring in the coastal savanna, in the forest zone, and in the northern savanna zone.

The coastal savanna in the southeastern plains around Accra consists of a mixture of scrub and tall grass (mostly Guinea grass), with giant anthills, often 10 to 14 feet high, providing an anchorage for thicket clumps that often include *Elaeophorbia* (a fleshy-leaved plant containing caustic latex) and other drought- and fire-resistant species such as the baobab (*Adansonia digitata*).

The forest zone

In the forest zone (the southern third of the country and the area along the Akwapim-Togo Ranges, where the mean annual rainfall exceeds 45 inches and is well distributed throughout the year without a pronounced dry season), the predominant vegetation is evergreen and tropical semi-deciduous forest. There are tall trees of varying heights, forming a closed canopy at the top, above which tower a few forest giants, such as the silk cotton tree, the wawa tree (African whitewood, a hardwood), and the African mahogany. The evergreen forest is in the extreme southwest, where the rainfall exceeds 65 inches a year, while there is a semi-deciduous forest farther north.

The dense forest zone formerly covered an area of about 30,000 square miles, but farming activities and timber exploitation have reduced it to less than 8,000 square miles, including about 6,000 square miles of reserved forest. To ensure the sustainable use of the country's rapidly diminishing forest resources, the government has embarked on a forestry policy involving the compulsory reforestation of cutover areas and more accurate measurements of exploitable timber and rates of extraction and regeneration.

The third vegetation type, the northern savanna, is found in the northern two-thirds of the country, where the low annual rainfall, between 45 and 30 inches, occurs in a single season and is followed by a period of intense drought. There the vegetation consists mostly of tall Guinea grass, together with a scattering of low trees, such as the shea butter tree, various species of acacia, and baobabs. Along the northern border the savanna gives way to a more open type of grassland that has developed largely as a result of prolonged human interference.

Ghana is relatively rich in animal life, although it has been reduced by hunting and the spread of human settlement. Large mammals include lions, leopards, hyenas, antelope, elephants, buffalo, wild hogs, chimpanzees, and many kinds of monkeys. Among the snakes are pythons, cobras, horned and puff adders, and green mambas. Crocodiles, the endangered manatees, and otters are found in the rivers and lagoons. Hippopotamuses are found in the Volta River. There are many species of lizards, tortoises, and giant snails.

Among the numerous birds are parrots, hornbills, kingfishers, eagles, kites, herons, cuckoos, nightjars, sunbirds, egrets, vultures, snakebirds, and plantain eaters.

Fish and other aquatic life

The ocean, rivers, and inland lakes are rich in fish and other forms of life. Sardines, locally called herring, arrive seasonally in the coastal waters in large shoals; other fish include mackerel, soles, skates, mullet, bonitos, flying fish, lungfish, elephant fish, sea bream, and sharks. Edible turtles, barracuda, and stingrays are fairly common; mussels, crabs, lobsters, and prawns also are found.

Insect life is particularly abundant. There are beetles, fireflies, ants, termites, butterflies, crickets, and bugs. Among the most dangerous insects are mosquitoes, tsetse flies, and blackflies (Simuliidae), which are responsible for transmitting the endemic diseases of malaria and yellow fever, trypanosomiasis (sleeping sickness), and onchocerciasis, or river blindness (a parasitic disease), respectively.

The Mole National Park in the western part of the Northern Region near Damongo is 1,900 square miles in extent. Other reserves have been developed farther south, notably on some of the islands in Lake Volta.

Settlement patterns. Ghana has three major geographic regions—coastal, forest, and northern savanna—the boundaries of which are not always clearly defined.

The coastal zone is traditionally a region of fishermen and small-scale farmers. The region was formerly occupied by a series of small kingdoms, the inhabitants of which were the first to be exposed to European contact—from the 15th century onward, perhaps even earlier. From east to west the principal ethnic groups are the Ewe, Adangme (Adangbe), Ga, Efutu, Fanti, Ahanta, and Nzima. The seaboard has made the region an important centre for commerce, causing population concentration in such urban centres as Accra, Cape Coast, and Sekondi-Takoradi. The coastal zone has more urban centres than any other region in Ghana.

Farther inland, occupying about a third of the country, is the forest region with its relatively large and prosperous traditional states and rich agricultural lands. West of the Volta these states consist mostly of Akan peoples; to the east the Ewe predominate. The forest environment and the economic activities and modes of life engendered by it, especially since the introduction of cocoa farming in 1879, have served to give the region a common stamp. Apart from the Ewe, the major ethnic groups are the Akwapim and Kwahu in the east, the Akim in the south, the Ashanti and Brong in the centre and north, and the Wasaw and Sefwi in the west. While all the peoples in the region have a relatively long history of settlement and political activity, those with the most impressive record are the Asante (Ashanti), who from the 17th to the late 19th century built a political empire centred on Kumasi, the present Asante capital, that included a large number of subject and satellite states spread throughout the region and in both the coastal and northern savanna zones.

The Asante

Almost all the timber, cocoa, and exploited mineral

wealth, as well as a number of minor cash crops grown for export, and a large part of the foodstuffs consumed in Ghana come from this region. Population density is relatively high, especially in the cocoa-growing areas. Except for Kumasi, there are few really large urban centres, although the region's other administrative centres—Ho, Koforidua, and Sunyani—form significant population concentrations.

The northern savanna region covers some two-thirds of the country but is economically the least developed region. The largest ethnic groups are the Dagomba and the Guang (Gonja), related to the Mossi people of Burkina Faso. The region has a harsh environment because of its low rainfall. The southern part immediately adjoining the forest zone, forming part of the disease-ridden "middle belt" of western Africa that combines the worst features of the forest and the savanna environments, is especially unattractive for settlement. In the past it was subject to extensive slave raiding from both north and south. Its distance from the sea and consequent insulation from active European contact over a long period retarded the development of the northern region.

Among the advantages of the northern region, especially in the most northerly part, which is relatively free from the tsetse fly so deadly to cattle, is an extensive savanna vegetation that is well suited to livestock breeding. Its relatively light soils and the rainfall regime favour the cultivation of yams and cereals. Although agriculture is mostly of the traditional subsistence type, irrigation and mechanized cultivation have opened up new prospects. Lake Volta, which extends far into the heart of the region, offers comparatively cheap communication with the south and serves as a reservoir of water for agricultural and other uses.

The pattern of rural–urban settlement

By the late 1980s only about one-third of Ghana's population was estimated to be urban, but there is a steady and increasing migration from the rural areas into the urban centres, some of which are expanding at about double the national population growth rate. Despite their rapid expansion in size and population, most of the urban centres remain small by world standards. The Accra–Tema agglomeration, with a population of more than one million, is the largest in the country, followed by Kumasi and Tamale.

Almost everywhere agriculture is extensive, rather than intensive, and rural settlements form scattered nuclei surrounded by land that is either under crops or is undergoing regeneration. Permanent or continuous cropping is encouraged throughout the country but is most common in the extreme northeast, where settlements consist of isolated compound houses, each surrounded by its own farm. Elsewhere, agriculture is based on a rotational system in which land is cropped for two or three years and then abandoned for from four to seven years, in order to allow it to regenerate. When cocoa or other tree crops are grown, however, cultivation is usually permanent.

The people. *Ethnic and linguistic groups.* Ethnically, the people of Ghana may be said to belong to one broad group within the African family, but there is a large variety of tribal, or subethnic, units. On the basis of language, it is possible to distinguish at least 75 different tribes. Many of these are very small, and only 10 of them are numerically significant. The largest groups are the Akan, Mole-Dagbani, Ewe, Ga-Adangme (Ga-Adangbe), and Gurma. Despite its tribal variety, there were no serious tribal dissensions when Ghana became independent. Tribal consciousness persists in many areas, however, and at times tensions have erupted, especially in northern Ghana, into violent clashes with many fatalities. At all levels in government and in public life, an effort has been made to play down tribal differences, a policy that has been helped by the adoption of English as the official language.

The largest tribal groups

Practically all the present tribes are believed to have moved into the country within the last 700 to 1,000 years in a series of migrations from the north, with the Ewe and Ga-Adangme, who occupy the southeastern corner of the country, entering from the east and southeast.

Religious groups. Nearly two-thirds of the population is Christian, about one-sixth is Muslim, and one-third adheres to the traditional tribal religions. Although the indigenous religions are widespread and deep-rooted, they lack a systematic body of doctrines. Though they are based, in general, on belief in the existence of a supreme being, a number of lesser deities associated with various natural phenomena are recognized. Considerable prominence is given to dead ancestors, who are considered to be ever-present, capable of influencing the course of events for the living, and capable of serving as intermediaries between the living and the gods.

Christianity has steadily gained ground at the expense of the indigenous religions. Christian influence is most dominant in the southern part of the country, while Islām is strongest in the extreme north and in the larger urban centres, which contain some immigrant populations from Muslim regions of western Africa. Since the 1950s many spiritualist churches claiming adherence to Christianity have appeared. The main divisions of the Christian church, however, are still the Roman Catholics, Methodists, Presbyterians, and smaller denominations.

Demographic trends. The country's population has, since 1970, maintained a high average annual growth rate of about 3 percent, with females slightly in excess of males. More than 60 percent of Ghanaians are under 25 years of age, assuring that the country's high growth rate will continue for some time. Life expectancy, which in 1960 was placed at about 46 years, has improved considerably and, at nearly 55 years in the late 1980s, was among the highest in western Africa.

Population fluctuations due to emigration became pronounced during the severe economic depression of the late 1970s and early 1980s. The expulsion of more than one million Ghanaian nationals, mostly young people without employable skills, from Nigeria in 1983 delivered a further shock to the economy but failed to cause major sociopolitical upheavals, owing largely to the built-in absorptive capacity of Ghana's indigenous social systems.

The economy. The economy is a mixture of private and public enterprise. National income is derived primarily from agricultural and mineral output and only to a limited extent from manufacturing and services. Most of the cash crops and mineral products are for export.

Before independence the government's role was confined mainly to the provision of such basic utilities as water, electricity, railways, roads, and postal services. Agriculture, commerce, banking, and industry were almost entirely in private hands, with foreign interests controlling the greater share in all of them except agriculture.

Shortly after independence, the government set out to extend its control over the economy by setting up a large number of state-owned enterprises in agriculture and industry. In order to make up for the local shortage of capital and entrepreneurial skills, measures were adopted to attract foreign investors to operate independently or in partnership with the government. These policies did not achieve the desired results because of poor planning and corrupt administration. By 1966, when the administration of President Kwame Nkrumah was overthrown, the heavy overseas borrowing upon which the government had relied to support its economic programs had dissipated almost all of the country's overseas reserves and had produced external and internal debts totaling some $1 billion.

Subsequent governments have sought to deal with the adverse balance of payments, to arrest inflation, to secure a rescheduling of overseas debts, to increase agricultural productivity, and to establish industrial development on a rational basis, as well as to save scarce foreign exchange by encouraging exports of locally manufactured goods.

Between 1966 and 1972 there was a marked contraction in governmental involvement in economic matters. The government continued to provide basic utilities and remained the largest single employer of labour. After the coup of 1972 policymakers returned to the concept of a centralized economy. The considerable debt owed to four British companies was repudiated, imports were cut, industrial projects abandoned after the fall of Nkrumah were resuscitated, and a policy of increased nationalization and state control was begun. In 1974, after a two-year suspension of foreign loans and aid, the government agreed

on a schedule for the repayment of its debts. This was accompanied by a more receptive policy toward investment by developed countries, though political instability resulted in a number of erratic economic policies. Ghana's external debt and balance of trade deficit increased and led to a devaluation of the cedi (the national currency) in 1978, a currency conversion in 1979, and a reduction of interest rates and demonetization of lower-value cedi notes in 1982. Under the restructuring program sponsored by the World Bank in the late 1980s foreign companies and private entrepreneurs were encouraged to invest in private or joint private and public ventures and to assist in the rehabilitation of the economy; the trend was toward increased privatization of the economy.

Taxation

A large part of government revenue is derived from various taxes, including a duty on cocoa, import duty, customs and excise duties, sales tax, income tax, property tax, and other taxes. Tax concessions are available to certain classes of business, and special incentives are offered to those generating foreign exchange through exports. In the late 1980s measures were instituted to widen the tax net so as to increase revenue, and subsidies on many goods—especially food items and imported fuel—and on public utilities were drastically reduced.

Despite attempts at reform, excessive bureaucratic controls and inefficiency continue to hamper trade and orderly economic development and to foster a flourishing underground economy. In addition, the massive devaluations of the cedi (from 1.02 cedis to the U.S. dollar in 1970 to 227 cedis to the U.S. dollar in 1989) have had mixed effects on both trade and the cost of living.

Although there is a minimum wage for workers, the gap in wages between the lowest-paid and well-paid workers is still wide. This disparity, coupled with rising living costs and instability in the national currency, impose severe hardships on a large section of the working population.

Trade unions

The trade union movement played a role in the struggle for self-government, and after independence the government, recognizing the importance of the movement as a political force, sought to make it a more direct instrument of policy. All trade unions in the country were brought under the authority of the Trade Union Congress, which was virtually an integral part of the government; this curtailed the freedom of workers to bargain with employers and the government. After the fall of the Nkrumah government, the monopoly of the Trade Union Congress was abolished and other unions were able to function. In 1972 the Trade Union Congress was revived, but the military government in 1982 once again suppressed its activity.

Resources. Although Ghana has a wide range of minerals, only a few—*e.g.,* gold, diamonds, manganese, and bauxite—are exploited. These minerals are found mostly in the southern part of the country. The gold industry, with an unbroken history dating from the 15th century, is the oldest; the others are of 20th-century origin—the working of manganese dating from 1915, diamonds from 1919, and bauxite from 1942. There are important reserves of limestone and iron ore, although they are not exploited. There is an aluminum rolling mill at Tema.

In 1970 oil was discovered offshore between Saltpond and Cape Coast. Although this discovery was initially classified as noncommercial, the steep world oil price increases of 1973–74 caused the government to reclassify it as commercial in 1974 and to undertake development. In 1974 and 1980 substantial amounts of natural gas were discovered offshore to the south and west of Cape Three Points. Oil production in the Saltpond area began in 1978, but it has proved disappointing; all crude oil is exported in order to reduce the country's foreign-trade deficit. Further explorations of a more comprehensive nature were begun in the late 1980s. Salt, in which the country is self-sufficient with a surplus for export, is obtained from the sea and lagoons. There are also extensive supplies of building stone, gravel, and sand.

Biological resources

Biological resources are extensive. The soil and climate favour a wide range of crops. The most important is cocoa, of which Ghana is a leading world producer. Timber and the crops of the forest zone constitute additional important biological resources. Yams and such cereals as rice and millet are produced primarily in the northern savanna zone; cattle are also raised there. The forests yield shea nuts and kola nuts. Ghana's offshore waters are rich in fish, and the creation of Lake Volta added another important source of fish for the domestic market.

Many of Ghana's rivers have the requisite regimes and rates of flow to permit exploitation for hydroelectric power. The Akosombo Dam on the Volta River and a second dam a few miles downstream at Kpong have a combined electrical capacity of more than one million kilowatts. Electricity from Akosombo meets most of the domestic requirements, leaving a surplus for sale to Togo and Benin.

Agriculture, forestry, and fishing. Apart from providing the bulk of national income, agriculture, forestry, and fishing employ more than half of the population. The annual output of cocoa—which is cultivated on more than one-half of Ghana's arable land—provides between three-fifths and three-quarters of the total revenue from exports. Consequently, the world price paid for cocoa directly determines Ghana's economic fortunes. Cocoa production fell sharply during the 1970s. It was undermined by aging and diseased trees, drought, bush fires, poor transport facilities, lack of adequate price and other incentives to farmers, and widespread smuggling across Ghana's borders. The Cocoa Marketing Board was abolished in 1979 following charges of corruption but was reconstituted in 1985 as the Ghana Cocoa Board. The government has repeatedly raised the production price of cocoa in an effort to stimulate production and to decrease the country's balance of payments deficit. These efforts and slight improvements in the world price have had a favourable effect on the economy, but the outlook remains far from promising.

Timber is also an important foreign exchange earner. Ghana's timber marketing is controlled by the Timber Export Development Board. Since the beginning of 1973 all foreign-owned timber export firms have been required to incorporate locally, so that the purchasing of timber is conducted only through this board.

The Ghana Oil Palm Development Corporation built a mill for the production of palm oil on its plantation near Kade. One of the largest in western Africa, the mill is designed to fulfill industrial and domestic consumption needs.

The Ghanaian domestic market is important. The value of food produced for local consumption is considerable. Successive governments have strongly supported diversification of food production to reduce reliance on a few crops and to cut the need for imported foodstuffs, but their measures have often been contradictory because of the emphasis on exports capable of earning foreign exchange. Besides cocoa, timber, and palm oil, other agricultural products that are exported include sugar, coffee, palm kernels, copra, and various fruits and vegetables.

Fishing

The various types of fish caught include cape hake, grunt, sea bream, tilapia, herring, mackerel, barracuda, and tuna. Most of the catch is sun-dried or smoked and consumed locally, but an increasing proportion is refrigerated; certain fishes, especially tuna, are mainly directed toward the overseas market. Government and private research agencies are working to increase the cash income of rural fishing communities and also to modernize the processing, distribution, and marketing of the catch. The State Fishing Corporation and a number of private companies operate a fleet of deep-sea trawlers.

Industry. A policy of industrialization has resulted in the establishment of a wide range of manufacturing industries, producing food products, beverages, tobacco, textiles, clothes, footwear, timber and wood products, chemicals and pharmaceuticals, and metals, including steel and steel products. These are produced mostly for local consumption. Among the announced program directives of a five-year plan (1975–80), however, was the maintenance of a reasonable balance on external trade, and a number of industrial projects were directed to the export market in either the short or long term. Ghana's industrial development has been hampered by a lack of capital, and official industrial development policy in the early 1980s recognized the importance of attracting foreign capital for the purpose of an effective economic take-off.

The mining industry

During the 1960s there was a decline in the mining industry, but in the 1970s increased world market prices for gold and manganese, new discoveries in the Prestea goldfields, and the formation of a National Manganese Corporation in 1975 to carry out a five-year plan to rehabilitate the manganese mines at Nsuta improved prospects for these minerals. The production of bauxite, on the other hand, has been well below capacity, and other minerals, notably diamonds, are close to depletion.

The development of Ghana's mineral industry has been hampered by a shortage of equipment, skilled personnel, and capital in the form of foreign exchange. The mineral industry failed to regain the economic preeminence it enjoyed before the 1960s. In the late 1980s, however, renewed efforts at revitalization with massive foreign investments and the encouragement of local and foreign private entrepreneurs were begun. Ghana possesses substantial bauxite reserves, though the output, all of which is exported, is less than half of capacity. High-quality sand in the Tarkwa mining area provides the basis for a small but important glass industry. Cement factories have been developed at Tema and Takoradi.

Revenue from tourism has increased gradually, with most of the tourists coming from Nigeria, the United Kingdom, Côte d'Ivoire, the United States, and West Germany. The Ghana Tourist Board and Ghana Tourist Development Company supervise the regulation, financing, and development of the tourist industry. Hotels are located at Accra, Tema, Takoradi, and Kumasi, and there is a hotel at Akosombo overlooking Lake Volta.

Transportation. The principal means of transport, in order of importance, are motor vehicles, railways, and aircraft. Animals are scarcely used except in the extreme north, where horses and donkeys are sometimes employed. In recent years there has been a marked rise in the number of privately owned vehicles, especially automobiles.

The density of roads and railways is much greater in the southern part of the country than in the north, and even in the south the cocoa-growing areas and the coastal zone tend to be favoured at the expense of other parts. Only about one-fifth of the country's roads are paved. With foreign aid, rehabilitation of the road network was undertaken in the mid-1980s.

The road system

Motor transport, now widespread and popular, was introduced in the towns about 1912 and spread quickly to the cocoa-producing areas. While the railways are owned by the state, motor transport is almost entirely in private hands. The state operates municipal bus services and express coach and freight services between the larger towns.

The Ghana Highway Authority oversees road maintenance and improvement. Road quality ranges from first-class paved (asphalt-surfaced) roads to third-class unsurfaced roads. First-class roads run between the large urban centres; they include the coastal Accra–Sekondi-Takoradi–Axim road; the Accra–Kumasi road, which continues northward to Tamale; and the Accra–Ho, Accra–Keta, Cape Coast–Kumasi, and Sekondi-Takoradi–Kumasi roads. A concrete-surfaced highway, with two lanes in each direction, runs from Accra to Tema. Second-class roads are narrower than first-class roads and have a base of swish (sun-dried earth) rather than quarried stone.

Rail transport

Rail transport was introduced in the early 20th century. The rail system forms a triangle joining Sekondi-Takoradi, Kumasi, and Accra. In addition, a central line runs from Huni Valley, on the Takoradi–Kumasi line, to Kade in the centre of the triangle, and an extension joins Achiasi on this central line with Kotoku on the Accra–Kumasi line. The port of Tema is also linked to the system by a short extension running from Achimota, near Accra. The Takoradi–Kumasi line is also joined by two other branch lines, one running from Tarkwa out to the gold-mining town of Prestea and the other running from Dunkwa to the bauxite-mining town of Awaso. Rail transport is much less popular than road transport. Railways are primarily used for the transport of freight, especially minerals and logs, although even this has declined since the early 1970s as the result of deteriorating facilities.

Small airports located at Takoradi, Kumasi, Sunyani, and Tamale are used solely for domestic services, while the Kotoka International Airport at Accra handles both domestic and international flights. Domestic air services are operated by a state-owned corporation, Ghana Airways, which also operates a western African service linking the coastal states and an international service to the Middle East, Europe, and the United Kingdom. While air transport is popular in Ghana, the maintenance of the national airways is costly and requires a large annual governmental subsidy. Air transport is used predominantly for passengers, but its use for cargo is rapidly expanding.

The importance of sea transport has dwindled with the expansion of air services. Most goods entering and leaving the country, however, are carried by sea. There are two modern harbours, Takoradi (opened in 1928) and Tema (opened in 1961). Takoradi specializes in exporting timber, manganese, and bauxite, while Tema specializes in cocoa. Both ports also handle passengers. In terms of tonnage, Tema and Takoradi handle about equal amounts of cargo. The national shipping company, the Black Star Line, Ltd., was plagued by debts, obsolete vessels, and labour disputes, and management was turned over to a West German concern. The line was renationalized by the military government in 1982. Ships from many other countries also use Ghanaian ports; traffic is mostly with Europe, the United States, and the Far East. The Ghana Railway and Ports Authority is responsible for the country's port operations under the Ministry of Transport and Communications.

The creation of Lake Volta occasioned an interruption on the Kumasi–Tamale road, where the Yeji ferry, formerly only a few yards wide, became a crossing of almost seven miles. A number of other ferries across the Volta have completely disappeared. Launch service on Lake Volta has been expanded, however. A launch service is also maintained on the lower reaches of the Volta between Ada and Amedika.

Administration and social conditions. *Government.* The constitution of 1979 provided for a unicameral parliamentary form of government with a president as head of state and a vice president. The president was elected for a term of four years (with the possibility of reelection for one further term) by universal adult suffrage. There was a broadly based Council of State with deliberative and advisory functions.

The executive branch

Flight Lieutenant Jerry Rawlings led a military coup in December 1981 and became the head of state as chairman of the Provisional National Defense Council (PNDC). The Parliament was disbanded and political parties were prohibited. In January 1983 the PNDC consisted of four members in addition to Rawlings but was subsequently increased to nine. The appointed Cabinet, comprising secretaries of state with ministerial portfolios or other responsibilities, has averaged between 20 and 25 members.

Ghana is divided into 10 regions: Western, Central, Greater Accra, Eastern, Volta, Ashanti, Brong-Ahafo, Northern, Upper West, and Upper East. After 1972, when the first constitution was suspended, each of the then existing eight regions (Greater Accra had not yet been created, and Upper West and Upper East formed the Upper region) was administered by a regional commissioner, who was an army officer. The constitution of 1979 revived the local, district, and regional councils, but the military government in 1982 appointed regional secretaries and district secretaries to take charge of the regions and districts and abolished the local councils as previously known. Under the PNDC's proposals for the development of participatory grass-roots democracy, district elections were held in late 1988–early 1989 for the establishment of district assemblies composed of elected and government-nominated members.

The judicial system is based chiefly on the English model, but Ghanaian customary law is recognized as well as English common law. The administration of justice is handled by various courts divided into two groups: the superior courts, consisting of the Supreme Court, the Court of Appeal, and the High Court; and inferior courts, consisting of the circuit courts, the district courts, and other courts provided by law, such as the juvenile courts. After the military government took control the Supreme

Court was abolished, and the Court of Appeal became the highest court. The constitution of 1979 restored the Supreme Court, and the Rawlings government retained it as the final court of appeal. A special tribunal investigates corruption. In 1982 the PNDC established a special military tribunal to hear cases against members of the armed forces and people's tribunals to handle corruption and abuse of public office. The adjudicating authorities in chieftaincy and purely traditional matters are the regional and national houses of chiefs. Appeals from decisions of the National House of Chiefs are made directly to the Supreme Court.

Dating to the period of British colonial rule, chieftaincy and the traditional political authorities have tended to run along parallel lines with the central government. Since independence, this tendency has persisted or even expanded, and the institution of chieftaincy has become increasingly divorced from the exercise of real political power at almost all levels of government; its role now is largely ceremonial.

Education. Ghana has one of the best-developed educational systems and one of the highest adult literacy rates in tropical Africa, but the cost is high. In April 1974 the government began implementation of a new educational system. It consisted of a pre-primary cycle for ages four to six; a basic first cycle, including six years of primary education and three years of junior secondary; and a second cycle of variable length. The second cycle leads to secondary vocational or commercial programs or to a senior secondary course preparing students for university studies or other third-cycle courses in high-level polytechnics and specialized institutions.

The first cycle is free and compulsory; for the first three years education is in the predominant local language, with provision for education in at least one other Ghanaian language and English, the latter being the language of instruction from the fourth year of the primary cycle.

The universities

Teacher training and technical education are approximately equivalent to secondary education, though they tend to attract pupils who are not aiming at university careers.

University education is provided at three institutions: the University of Ghana, at Legon, near Accra, established as a university college in 1948 and granted full university status in 1961; the University of Science and Technology at Kumasi, established in 1951 and granted full university status in 1961; and the University of Cape Coast, established in 1962 for the training of science teachers and granted full university status in 1972. All three institutions are financed by the government; there are no private universities. The Tarkwa School of Mines, which is also a government-financed institution, is affiliated with the University of Science and Technology and offers diploma courses in mining and related subjects.

Under the restructuring program of the late 1980s it was decided that more drastic reductions should be made in student subsidies for food and accommodation in third-cycle institutions.

The enrollment in all schools, especially in secondary schools, has soared dramatically since Ghana achieved self-government. There are a number of private schools at both elementary and secondary levels. The number of available places in second- and third-cycle institutions, especially the universities, however, is still far short of the demand from qualified applicants.

Despite the heavy national expenditure on education and the large school population, Ghana still has a relatively low literacy level by world standards. Thanks to the extensive use of the sound and visual media, however, illiteracy is not as serious a handicap as it formerly was. English is widely spoken, especially in the urban areas.

Health and welfare. Major health problems are communicable diseases, poor sanitation, and poor nutrition. The main emphasis of government health policy is on improved public health, and since independence many improvements have been made in nutrition and in maternal and child care. Many of the endemic diseases, such as malaria, pneumonia, and diseases of the gastroenteritis group, which formerly took a heavy toll of life, have been brought under a measure of control as a result of improved hygiene, better drugs, and education. The infant mortality rate has shown a steady decline, especially in urban areas, as a result of improved health care facilities and dietary habits; it is among the lowest in western Africa but remains high by world standards.

There are hospitals and clinics provided by the government and by various Christian missions in most parts of the country. Supplementary services consist of health centres, dispensaries, and dressing stations. Considerable progress has been made in the quantity and quality of health facilities and medical personnel, but rapid population growth continues to impose great pressures on the available facilities. In addition to the large number of doctors in the public service, many private practitioners operate their own clinics and hospitals. Registered doctors and dentists are supported by a paramedical staff of nurses, midwives, and pharmacists, as well as by auxiliaries. There are two medical schools in the country: one in the University of Ghana in Accra and the other in the University of Science and Technology in Kumasi.

Rural community development

Government programs of rural community development are assisted by village improvement projects undertaken with the participation of the residents. Welfare and economic aspects outside urban areas are handled by the national government. In the urban areas, welfare services concentrate on casework, probation work, youth activities, and guidance through voluntary organizations. There is a government-sponsored pension plan for wage employees.

With the rapid growth of population and the movement of large numbers of people from the rural to urban areas, housing has become an acute problem, especially in the large cities where the problem is both quantitative and qualitative. In the rural areas the problem is mainly qualitative. There is distinct overcrowding in the urban areas, where the number of persons per house is about twice that in rural areas. All but a small proportion of housing is provided by private individuals, and the main role of the Ministry of Works and Housing is to supplement these efforts by handling standard government housing needs, developing a few housing estates in the large towns through the Housing Corporation, and piloting demonstration projects in such areas as low-cost housing and the development of suitable building materials. However, town planning procedures and human settlements policies are still largely outdated or inadequate.

Cultural life. *The cultural milieu.* Ghana has a rich indigenous culture. Culturally, the peoples of Ghana have many affinities with their French-speaking neighbours, but each tribal group has distinctive cultural attributes. In all parts of the country the cultural heritage is closely linked with religion and the institution of chieftaincy. Various festivals and rites are centred on chieftaincy and the family and are occasioned by such events as harvest, marriage, birth, puberty, and death.

Ghanaian society

Ghanaian society is without sharp class distinctions. Insofar as traditional authority is based on a system of hereditary chieftaincy, it is possible to speak of aristocratic classes within the tribal groups, but the institution of chieftaincy is essentially democratic in operation and the authority of chiefs is broadly based. Land is usually owned by families, militating against the emergence of a small, powerful landed class wielding economic control over a landless class. These inherent egalitarian tendencies of the society have been heightened by economic and social mobility, depending on education and individual enterprise.

Daily life. Although the bonds of family life are an important factor in the social norms of Ghanaians as a whole, they tend to be much less pronounced among the urban population, where the trend is toward the nuclear family, especially among the professional classes and scattered immigrant groups.

The same differences between the urban and rural populations are found in dress and eating habits, with the urban dwellers being distinctly more Westernized and sophisticated. The national cuisine reflects the country's agricultural wealth and varied historical connections, and Ghana is one of the few countries in tropical Africa that can be said to possess a rich indigenous cuisine.

Traditional social values, such as respect for elders and

the veneration of dead ancestors, are generally more evident among the rural than the urban population. However, a revival in the importance of these values and a closer identity with traditional social roots, as expressed in the institution of chieftaincy, is gaining ground among the urban diaspora drawn from different parts of Ghana.

The arts. Ghana's arts include dance and music, plastic art (especially pottery and wood carving), gold- and silverwork, and textiles, most notably the richly coloured, handwoven *kente* cloth of the Akan and Ewe.

Indigenous art is in keen competition with various art forms of foreign origin, especially in those areas in which the end product is intended for practical household or personal use, such as pottery, carving, gold- and silversmithing, and weaving. Consequently, only the unique and most indispensable of these forms have managed to survive without special public support or patronage. The increased national self-consciousness generated in Ghana and in other African countries by the independence movement, however, has been instrumental in fostering and popularizing many art forms. An equally important factor has been the government's effort to promote tourism.

Of growing importance are literary and dramatic works produced by Ghanaian authors, although most of these are in English. Ghanaian works have attracted world attention in the fields of popular music, painting, sculpture, and film production. Important innovations in traditional dance have taken place since the mid-1960s, when the University of Ghana's Institute of African Studies embarked on the systematic study and organization of indigenous dance forms.

Cultural institutions. Apart from small groups of craftsmen who provide the chiefs' stools and skins throughout the country—a stool is the traditional symbol of office for chiefs in southern Ghana, and a skin is the equivalent symbol in the north—there are few established cultural institutions. The most outstanding are the National Cultural Centre, based in Kumasi, and the Arts Council of Ghana, based in Accra and with branches throughout the country. The National Cultural Centre is primarily concerned with the cultural heritage of Asante, while the Arts Council is concerned with the preservation of indigenous Ghanaian culture in all of its forms and with its development and improvement in light of contemporary local and world trends. Dance, music, drama, painting, and sculpture all come within the purview of the council. The Ghana Museum and Monuments Board is based in Accra. The board has an ethnological museum and a science museum in Accra and is responsible for the maintenance of buildings and relics of historical importance, such as the forts and castles, and for the preservation of important art treasures throughout the country. The forts, built by various European powers, mostly between the 14th and 18th centuries, are all, except the Kumasi fort (1897), located on the coast.

Robert Frerck—Odyssey Productions

The University of Ghana, at Legon, near Accra.

For statistical data on the land and people of Ghana, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (E.A.B.)

HISTORY

Prehistoric era. As elsewhere in Africa, the climate of Ghana varied during the Pleistocene epoch. With greater rainfall, the forest spread northward and humans retreated toward the Sahara; when rainfall diminished, they occupied even the present forest. Apart from some pebble tools from high river terraces, the first industry is Late Chellean in the southeast. In the succeeding pluvial era, the Acheulean culture is lacking save from the extreme north.

With increases in aridity, humans reappeared, bringing Late Acheulean and Sangoan cultures, probably successively. They moved along the Togo mountain range from the Niger River. Sangoan tools abound in Transvolta and around Accra and extend to Kumasi; the west remained forest and was rarely visited. The Sangoan culture waned in the Gamblian pluvial era. At its close there appears a Lupemban culture, probably from the desiccating Sahara; it occurs in basal gravels of valleys carved during the preceding pluvial period. In central Ghana its tools are shapely, near the coast crude and formless.

Mesolithic traditions lingered into the succeeding subpluvial era. Thereafter excavations at Legon yielded quartz microliths made on small pebbles. Up-country these occur on silt terraces deposited in the preceding wet phase as far as the Niger. This culture is independent of the Saharan Mesolithic.

The latest Mesolithic period has stone hoes, quartz beads, and other Congo types; pottery seems absent. This stage dates to the post-Flandrian marine regression (end of 2nd millennium BC?).

Several Neolithic cultures seem identifiable. They contain polished axes and usually coarse pottery. The most distinctive appears around Kintampo and in the Accra plains; it had clay houses, Saharan chert microliths, shale arm rings, and scored terra-cottas like flattened cigars. A Neolithic culture more in Mesolithic tradition was excavated near Abetifi. Neolithic cultures

Evidence is lacking for the introduction of iron. Polished stone was commonly used until the 16th century, especially in the forest. Trade in greenstone for ax manufacture flourished. In Transvolta and the west greenstone hoes are common. No satisfactory chronology has been established, nor can existing tribes be identified before the 17th century. Of excavated sites, Nsuta, with decorated pottery and bobbin beads, should be early medieval; the Sekondi village and cemetery, with fine pottery, stone axes, and quartz and shell beads, lasted until Portuguese times. In the north heavily decorated pottery continued later on open sites and mounds indicating clay houses. European imports are unknown before the 17th century. (O.D./Ed.)

Early traditions. The modern state of Ghana is named after the ancient African empire that flourished until the 13th century and was situated close to the Sahara in the western Sudan. The centre of ancient Ghana lay about 500 miles to the northwest of the nearest part of the modern state, and it is tolerably certain that no part of the latter lay within its borders. The claim that an appreciable proportion of modern Ghana's people derived from emigrants from the ancient empire cannot be substantiated with the evidence available at present. Written sources relate only to the period since European contact with the Gold Coast—*i.e.*, modern Ghana—began in the 15th century, or to Muslim contacts with ancient Ghana from about the 8th to the 13th century. Many modern Ghanaian peoples possess well-preserved oral traditions, but even though some of these may reach as far back as the 14th century, this is after the final disappearance of ancient Ghana, and such very early traditions often present considerable problems of interpretation. Little progress has so far been made in linking the surviving traditions with the available archaeological evidence.

Trade routes of Islām. More archaeological research,

especially into the Iron Age, will undoubtedly do much toward resolving present uncertainties about the early history of modern Ghana, but for the moment little more can be said than that the traditions of many of the states into which the country was divided before it came under British rule refer to their people having immigrated within the last 600 years either from the north or northwest or from the east or northeast. Such traditions link up with other evidence to suggest that the area which is now Ghana was for many centuries a meeting place for two great streams of western African history. Ultimately these streams stemmed from the existence of two major trans-Saharan routes, a western one linking the headwaters of the Niger and Sénégal rivers to Morocco and a more central one linking the region between the Niger Bend and Lake Chad with Tunisia and Tripoli. At the end of the western route arose the great Mande states, notably ancient Ghana and Mali, while around the more easterly termini developed Songhai, the Hausa states, and Bornu. There is evidence that parts of modern Ghana north of the forest were being reached by Mande traders (seeking gold dust) by the 14th century and by Hausa merchants (desiring kola nuts) by the 16th century. In this way the inhabitants of what is now Ghana were influenced by the new wealth and cross-fertilization of ideas that arose in the great empires of the western Sudan following the development of Islāmic civilization in northern Africa.

Trans-Saharan routes

Migrations. It is against this background that the traditions of origin of the Ghanaian states must be viewed. It would seem that the first states of the Akan-speaking peoples who now inhabit most of the forest and coastlands were founded, in about the 13th century, by the settlement just north of the forest of migrants coming from the direction of Mande; that the dominant states of northern Ghana, Dagomba, Mamprussi, and their satellites were established by the 15th century by invaders from the Hausa region; that a little later the founders of the Ga and Ewe states of the southeast began to arrive from what is now Nigeria by a more southerly route; and that Gonja, in the centre, was created by Mande conquerors about the beginning of the 17th century.

Tradition tends to present these migrations as movements of whole peoples. In certain instances—for example, Dagomba, Mamprussi, and Gonja—it can be shown that the traditions relate in fact to comparatively small bands of invaders who used military and political techniques acquired farther north to impose their rule on already established populations whose own organization was based more on community of kin than on allegiance to political sovereigns. It is probable that the first Akan states—*e.g.,* such influential states as Bono and Banda north of the forest or the smaller states founded on the coast by migration down the Volta River—were also established in this way. The later Akan infiltration into the forest, which then was probably sparsely inhabited, and the Ga and Ewe settlement of the southeast may have been more of mass movements, though in the latter case it is known that the immigrants met and absorbed earlier inhabitants.

Contact with Europe and its effects. A revolution in Ghanaian history was initiated by the establishment of direct sea trade with Europe following the arrival on the coast of Portuguese mariners in 1471. Initially Europe's main interest in the country was as a source of gold, a commodity that was readily available at the coast in exchange for such European exports as cloth, hardware, beads, metals, spirits, arms, and ammunition. This gave rise to the name Gold Coast, by which the country was known until 1957. In an attempt to preserve a monopoly of the trade, the Portuguese initiated the practice of erecting stone fortresses (Elmina Castle dating from 1482 was the first) on the coast on sites leased from the native states. In the 17th century the Portuguese monopoly, already considerably eroded, gave way completely when traders from the Netherlands, England, Denmark, Sweden, and Prussia—Protestant seapowers antagonistic to Iberian imperial pretensions—discovered that the commercial relations developed with the Gold Coast states could be adapted to the export of slaves, then in rapidly increasing demand for the American plantations, as well as of gold. By the mid-18th century the coastal scene was dominated by the presence of about 40 forts controlled by Dutch, British, or Danish merchants.

The presence of these permanent European bases on the coast had far-reaching consequences. The new centres of trade thus established were much more accessible than were the Sudanese emporia, and this, coupled with the greater capacity and efficiency of the sea-borne trade compared with the ancient overland routes, gradually brought about the reversal of the direction of the trade flow. The new wealth, tools and arms, techniques and ideas introduced through close contact with Europeans initiated political and social as well as economic changes. The states north of the forest, hitherto the wealthiest and most powerful, declined in the face of new combinations farther south. At the end of the 17th century the Akan state of Akwamu created an empire that, stretching from the central Gold Coast eastward to Dahomey, sought to control the trade roads to the coast of the whole eastern Gold Coast. The Akwamu empire was short-lived, but its example soon stimulated a union of the Asante states of the central forest, which union, after establishing its dominance over other neighbouring Akan states, expanded north of the forest to conquer Bono, Banda, Gonja, and Dagomba.

The Asante states

Having thus engrossed almost the whole of the area that served as a market and source of supply for the coastal trade, Asante turned toward the coastlands. There traditional ways of life were being increasingly modified by contact with Europeans and their trade, and when, beginning in the latter part of the 18th century, Asante armies began to invade the coastal states, their peoples tended to look for leadership and protection to the European traders in the forts. But between 1804 and 1814 the Danes, English, and Dutch had each in turn outlawed their slave trades, and the gold trade was declining. The political uncertainty following the Asante invasions impeded the development of new trades, and in these circumstances the mutually suspicious European interests were not always keen to embark on new political responsibilities. However, during 1830–44, under the outstanding leadership of George Maclean, the British merchants began to assume an informal protectorate over the Fanti states, much to the commercial benefit of both parties. As a result of this the British Colonial Office finally agreed to take over the British forts, and in 1850 it was able to buy out the Danes. However, trade declined under the new regime, which was averse to assuming formal control over the territory influenced from the forts, and in the 1860s, as a result of this influence and of the growth, from the 1820s onward, of Christian missionary education, the Fanti states attempted to organize a European-style confederacy. Further Asante incursions and the final evacuation of the coast by the Dutch (1872) combined to reverse this British policy, and in 1874 a punitive expedition sacked Kumasi, the Asante capital, and the Gold Coast was declared a British colony.

Colonial period. French and German activity in adjacent territories and the demand of British mining and commercial interests for better protection led to a further active period of British policy during 1895–1901, during which Asante was conquered and its northern hinterland formed into a British protectorate. The 56 years of British rule that followed went far toward welding into one state the three elements of the territory, the colonies of the Gold Coast and Asante and the protectorate of the Northern Territories, to which after World War I was added a fourth, under mandate from the League of Nations, the western part of former German Togoland. But this was hardly the result of deliberate policy. The ever increasing assimilation of European ways by the people on the Gold Coast had already made possible there the introduction of such organs of government as a legislative council (1850) and a supreme court (1853), but for many years Asante and the Northern Territories remained the sole responsibility of the governor, whose officials were from the 1920s onward encouraged to work with and through the authorities of the indigenous states. Attempts to introduce similar elements of indirect rule in the Gold Coast served mainly to stimulate a nationalist opposition among the educated professional classes, especially in the growing

towns, which aimed at converting the legislative council into a fully responsible parliament.

What really brought the country together was the great development of its economy following the rapid expansion of cocoa-growing by farmers in the forest. By the 1920s the Gold Coast, while continuing to export some gold, was producing more than half of the world's supply of cocoa; timber and manganese later became additional exports of note. With the wealth created by this great increase of trade, it was possible to provide modern transport facilities—harbours, railways, roads—and social services, especially education (to the university level), all of which tended toward the conversion of the traditional social order, of groups bound together by kinship, into one in which individuals were linked principally by economic ties.

Independence. Political advancement tended to lag behind economic and social development, especially in the south (for the role of the Northern Territories was principally the supply of cheap labour for the Gold Coast and Asante). World War II and its aftermath tended to accentuate this lag, and in 1948 there were riots in the larger towns. The Watson Commission of Inquiry reported that the Burns constitution of 1946, which had granted Africans a majority in the legislative council, was "outmoded at birth." An all-African committee under Justice (later Sir Henley) Coussey was appointed to work out a new constitution in which some executive power would be transferred to African ministers responsible to an African assembly. Meanwhile, a radical politician, Kwame Nkrumah, had established the Convention People's Party (CPP), which, with wide popular support, campaigned with strikes and other actions under the slogan "Self-government now." In 1951 the CPP won almost all the elective seats in the post-Coussey legislative assembly, whereupon Governor Sir Charles Arden-Clarke invited Nkrumah to lead the new administration. A partnership developed between the two so that power was rapidly transferred to an all-African cabinet responsible to a popularly elected national assembly. In 1956 the trust territory of British Togoland chose by United Nations plebiscite to integrate with the Gold Coast. Securing more than 70 percent of the assembly seats at general elections in 1954 and 1956, Nkrumah and the CPP government were able in 1957 to obtain the recognition of their country, renamed Ghana, as an independent self-governing member of the Commonwealth and a member of the UN.

Kwame Nkrumah and the Convention People's Party

Nkrumah saw independent Ghana as a spearhead for the liberation of the rest of Africa from colonial rule and the establishment of a socialist African unity under his leadership. After the establishment of a republic in 1960, the state became identified with a single political party (the CPP), with Nkrumah, as life president of both, taking ever more power for himself. On the pan-African front Nkrumah's messianism was increasingly challenged by other, often more stable, leaders of an ever-growing number of independent states. By 1966 his dream of African socialism was foundering under haphazard and corrupt administration, massive foreign debts, and declining living standards. In February, while he was in Peking, army and police leaders rose against him, and his regime was replaced by a National Liberation Council chaired by Lieutenant General Joseph A. Ankrah. The machinery of government was overhauled and conservative financial policies introduced. But Ankrah failed to redeem a promise to restore parliamentary democracy, and in 1969 he gave way to the dynamic young brigadier Akwasi Amankwaa Afrifa, a principal leader of the coup. A constituent assembly produced a constitution for a second republic, and a general election was held in August 1969. This resulted in a substantial victory for the Progress Party, led by Kofi Busia, a university professor who had consistently opposed Nkrumah. Busia became prime minister, and a year later a former chief justice, Edward Akufo-Addo, was chosen as president.

But the civilian regime, handicapped by the great burden of foreign debt it had inherited and the low prices then obtained by cocoa on the world market, was slow to produce the results expected of it. In January 1972 impatient army officers intervened again, and the government was taken over by a National Redemption Council (NRC) of military men chaired by Colonel Ignatius Kutu Acheampong. The national assembly was dissolved, public meetings prohibited, political parties proscribed, and leading politicians imprisoned. In July 1972, a retroactive Subversion Decree was enacted under which military courts were empowered to impose the death penalty for offenses such as subversive political activity, robbery, theft, and damaging public property, and, from 1973, for the spreading of rumours and profiteering. The military regime was clearly failing to maintain good order or anything approaching a prosperous or stable economy. In 1975, the NRC was reorganized to include some civilians, but ultimate power was given to a Supreme Military Council.

The coup of 1972

In 1978, Acheampong (now a general) proposed a "Union Government to which everybody will belong," with no political parties and the military sharing in government with civilians. But a national referendum held to approve this served mainly to show the unpopularity of the NRC. He was replaced as chairman by Lieutenant General Frederick W.K. Akuffo, who had even less idea of what might be done than his somewhat more energetic predecessor. Eventually, in 1979, the government of the generals was overthrown by young officers and noncommissioned officers, led and inspired by an air force flight lieutenant, Jerry Rawlings. Acheampong and Akuffo were executed, and a quick return to parliamentary government was organized. But under President Hilla Limann this failed to produce the radical improvements in the political and economic life of Ghana looked for by Rawlings and his colleagues. At the end of 1981, Rawlings decided that he and those who thought like him must take the lead in all walks of life. His second military coup established a Provisional National Defense Council as the supreme national government; at local levels, people's defense committees were to take the campaign for national renewal down to the grass roots.

Initially older Ghanaians doubted that Rawlings and his colleagues could provide more effective and less self-interested government than the old politicians or generals, while other young soldiers thought that they could themselves engineer coups to secure the fruits of power. But Rawlings easily snuffed out two countercoups in 1982 and 1983, and it was apparent that there was wide and genuine approval of his purpose of reforming Ghana's political and economic life. This continued even when he decided that there was no alternative but to follow conservative economic policies that would secure International Monetary Fund support and other foreign aid.

For later developments in the history of Ghana, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (J.D.F.)

Guinea

The Republic of Guinea (République de Guinée) is an independent nation of western Africa. It is bordered by Guinea-Bissau, Senegal, and Mali to the north and east; by Côte d'Ivoire to the southeast; by Liberia and Sierra Leone to the south; and by the Atlantic Ocean to the west. Its area of 94,926 square miles (245,857 square kilometres) supports a largely rural population. The national capital of Conakry is the country's main port.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* There are four geographic regions: Lower Guinea, the Fouta Djallon, Upper Guinea, and the Forest Region. Lower Guinea includes the coast and coastal plain. The coast has undergone recent marine submergence and is marked by rias, or drowned river valleys, that form inlets and tidal estuaries. Numerous offshore islands are remnants of former hills.

Immediately inland the gently rolling coastal plain rises to the east, being broken by rocky spurs of the Fouta Djallon highlands in the north at Cape Verga and in the south at the Kaloum Peninsula. Between 30 and 50 miles (48 and 80 kilometres) wide, the plain is wider in the south than the north. Its base rocks of granite and gneiss (coarse-

grained rock containing bands of minerals) are covered with laterite (red soil with a high content of iron oxides and aluminum hydroxide) and sandstone gravel.

The Fouta Djallon highlands rise sharply from the coastal plain in a series of abrupt faults. More than 5,000 square miles of the highlands' total extent of 30,000 square miles lie above 3,000 feet (900 metres). Basically an enormous sandstone block, the Fouta Djallon consists of level plateaus broken by deeply incised valleys and dotted with sills and dikes, or exposed structures of ancient volcanism resulting in resistant landforms of igneous rock, such as the Kakoulima Massif, which attains 3,273 feet northeast of Conakry. The highest point in the highlands, Mount Loura, rises to 4,970 feet (1,515 metres) near the town of Mali in the north.

Upper Guinea is composed of the Niger Plains, which slope northeastward toward the Sahara. The flat relief is broken by rounded granite hills and outliers of the Fouta Djallon. Composed of granite, gneiss, schist (crystalline rock), and quartzite, the region has an average elevation of about 1,000 feet.

The Forest Region, or Guinea Highlands, is a historically isolated area of hills in the country's southeastern corner. Mount Nimba (5,748 feet; 1,752 metres), the highest mountain in the region, is located at the borders of Guinea, Liberia, and Côte d'Ivoire. The rocks of this region are of the same composition as those of Upper Guinea.

Three major western African rivers

Drainage and soils. The Fouta Djallon is the source of western Africa's three major rivers. The Niger River and several tributaries, including the Tinkisso, Milo, and Sankarani, rise in the highlands and flow in a general northeasterly direction across Upper Guinea to Mali. The Bafing and Bakoy rivers, headwaters of the Sénégal River, flow northward into Mali before uniting to form the main river. The Gambia River flows northwestward before crossing Senegal and The Gambia.

The Fouta Djallon also gives rise to numerous smaller rivers, such as the Fatala, Konkouré, and Kolenté, which flow westward across the coastal plain to enter the Atlantic. The Forest Region generally drains to the southwest through Sierra Leone and Liberia. The St. Paul River enters the Atlantic at Monrovia, Liberia, and the Moa River has its mouth at Sulima, Sierra Leone.

The most common soils are laterites formed of iron and hydrated aluminum oxides and other materials that often concretize into hard iron-rich conglomerates. In the northeast, sandy brown soils predominate, while along the coast black, heavy clay soils accumulate in the backwaters. There are alluvial soils along the major rivers. Soil conservation is extremely important because most soils are thin, and the heavy rainfall causes much erosion.

Climate. The climate of Guinea is tropical with two alternating seasons—a dry season (November through March) and a wet season (April through October). The arrival of the migratory intertropical convergence zone (ITCZ) in June brings the heaviest rainfall of the wet season. As the ITCZ shifts southward in November, the hot, dry wind known as the harmattan blows from the northeast off the Sahara.

On the coast a period of six months of dry weather is followed by six months of rain. The average rainfall at Conakry is 170 inches (4,300 millimetres) a year, and the average annual temperature is about 81° F (27° C). In the Fouta Djallon, January afternoon temperatures range between 86° and 95° F (30° and 35° C), while evening temperatures dip to 50° F (10° C). Rainfall varies between 63 and 91 inches annually, and the average annual temperature is about 77° F (25° C).

In Upper Guinea rainfall drops to about 59 inches a year. During the dry season temperatures of more than 104° F (40° C) are common in the northeast. In the Forest Region at Macenta, there may be 106 inches of rain annually. Only the months of December, January, and February are relatively dry, each having less than one inch of precipitation. At low altitudes, temperatures resemble those of the coastal areas.

Plant and animal life. The coast is fringed with mangrove trees, and the coastal plain supports stands of oil palms. The Fouta Djallon is mostly open, with trees growing along the wider stream valleys. In Upper Guinea, the savanna grassland comprises several species of tall grasses that reach heights of five to 10 feet during the rainy season. Deciduous trees grow in scattered clumps, but few have commercial value; baobabs and shea trees furnish fruit and oil. The Forest Region contains several extensive patches of rain forest, with teak, mahogany, and ebony trees; agriculture, however, has diminished the forests and resulted in a shift largely toward open savanna.

Vanishing animal life

Guinea is not rich in African big game. Baboons and hyenas are common, while an occasional wild boar, several types of antelope, and a rare leopard may be sighted. A few hippopotamuses and manatees inhabit the rivers of both Lower and Upper Guinea. Poisonous snakes include mambas, vipers, and cobras, along with pythons and a variety of harmless snakes. Crocodiles and several varieties of fish are found in most rivers.

Settlement patterns. Until recent urbanization and movement toward regional towns, the Fulani (Fulbe, Fula, or Peul) of the Fouta Djallon tended to live in small hillside hamlets of 75 to 95 persons each, with the lower classes occupying the valleys. In the heart of the highlands the countryside was thickly settled with hamlets every few miles, while in the east the land was less settled. In Lower Guinea, villages were grouped together at the bases of hills, in the open plain, or in a valley floor. Village solidarity was more marked in this area than in the highlands, and each village contained between 100 and 200 people.

The majority of the Malinke (Mandingo) people of Upper Guinea lived in moderately large villages of about 1,000 inhabitants located near permanent water sources, the adjacent soils of which were used for cultivation. The villages were tightly grouped; there were empty brush areas in which farming was unprofitable.

In the Forest Region the effects of human occupation, especially in the southwest, have become apparent only since the mid-20th century. Among the Kissi people on the Sierra Leone and Liberian borders, rice was grown on most hillsides and in every low-lying and swampy area. Villages tended to be small and rarely contained more than 150 people; they were often tucked inside groves of kola, mango, and coffee trees. Farther east among the Loma and Kpelle people, fire-cleared land was used to plant vegetables and rice. Larger villages were usually located on remote hillside terraces that are often surrounded by secondary forest growth.

Guinea's urban centres

Guinea's main urban centre is Conakry. The old city, located on Tumbo Island, retains the segregated aspect of a colonial town, while the Camayenne (Kaloum) Peninsula community, which has grown up since the 1950s, has a few buildings of the colonial period. From the tip of the peninsula, an industrial zone has a growing salaried population that is truly urbanized.

The second largest town, Kankan, in Upper Guinea, is a commercial, educational, administrative, and Muslim religious centre of some importance. Labé, located well into the Fouta Djallon, serves as a market town and an administrative and educational centre. Nzérékoré, in the Forest Region, serves the same functions as Labé. Other important towns are the trading centres of Kindia and Mamou and the industrial settlements of Boké, Fria, and Kamsar.

The people. *Ethnic and linguistic composition.* The four major geographic regions largely correspond to the areas inhabited by the major linguistic groups. In Lower Guinea the major language of Susu has gradually replaced many of the other indigenous languages and is a lingua franca for most of the coastal population. In the Fouta Djallon the major language is Pular (a dialect of Fulfulde, the language of the Fulani), while in Upper Guinea the Malinke (Maninkakan) language is the most widespread. The Forest Region contains the linguistic areas, from east to west, of Kpelle (Guerzé), Loma (Toma), and Kissi.

Besides the diplomatic community and a growing number of expatriate teachers and technical advisers, the number of non-Guinean residents has increased considerably since 1984. This community includes Lebanese and Syrian traders and a growing number of French engaged in agriculture, business, and technical occupations.

Religions. Nearly three-quarters of the population is

Muslim, and a small but influential percentage is Christian, mostly Roman Catholic. A substantial minority of Guineans continue to follow local traditional religious practices.

Demographic trends. Life expectancy has improved only slightly since independence and, at 39 years for men and 42 years for women, ranks as one of the lowest in sub-Saharan Africa.

Immigration increased slightly after 1984, and emigration, which was high in the 1970s and early 1980s—especially from the Fouta Djallon and Upper Guinea—decreased in the 1980s. At its peak this out-migration consisted of one-sixth of the working-age male population, leaving an imbalance of aged, children, and women. Emigration was directed toward the neighbouring countries, with a small percentage going to Europe or North America.

The heaviest regional population concentration is in the Fouta Djallon. Conakry, the Camayenne Peninsula, and, to a lesser extent, the industrial enclaves of Boké, Fria, and Kamsar suffer from rapid population growth caused primarily by continuing migration from the rural areas to the urban centres. Except for the Fouta Djallon, population poses no serious immediate threat to development because there is no pressure on the land and no landholding class.

The economy. Agriculture and other rural activities account for 80 percent of the country's employment, with less than 10 percent in industrial employment (including mining). The rest of the wage and salary earners are in the service and governmental sectors. In general, salaries are low, and the need for extra-salary means in order to eke out a livelihood remains the norm.

The shortage of trained personnel is serious, and finances suffer from misappropriation and tax evasion. Many of the processing industries have been held back by inadequate supplies of raw materials. Internal production is not sufficiently high, in agriculture particularly, and the shortage of investment capital is persistent.

Bauxite and diamonds

Resources. Guinea has from one-third to one-half of the world's known reserves of bauxite, plus significant reserves of high-grade iron ore at Mount Nimba and the Simandou Mountains. Alluvial gold is taken from the Niger and its tributaries, and diamond production is substantial and largely of gem-quality stones. The southeastern rain forest has some valuable species of tropical hardwoods, and both river and ocean fisheries yield large catches of food fish. Hydroelectric potential is considerable because of the high rainfall and deep gorges of the Fouta Djallon but has been only partially developed, largely to meet the demands of the alumina sector.

Agriculture, forestry, and fishing. Guinea is an agricultural nation. The high plateaus of the Fouta Djallon are little more than part-time pastures, with hillsides given over to the growing of peanuts (groundnuts) and fonio (a sorghumlike grain). Along the streams and rivers, rice, bananas, tomatoes, strawberries, and citrus fruits are grown commercially. Most families have truck gardens, and tsetse-resistant Ndama cattle, sheep, goats, horses, donkeys, chickens, and Muscovy ducks are raised.

In Lower Guinea, oil and coconut palms, rice, bananas, vegetables, salt, and fish are important elements of trade. A number of large-scale plantations produce a good quantity of bananas and pineapples. Except for poultry and a few goats, there are relatively few domestic animals. In Upper Guinea, grains and cassava (manioc) are important food crops; vegetables, tobacco, and karite (shea butter) are traded locally; and domestic animals are common.

In the Forest Region, rice is the chief food crop, along with cassava, peanuts, and corn (maize). Gardens of tomatoes, peppers, eggplants, and tobacco are scattered in the shade of fruit trees, and coffee trees, kola nuts, and oil palms are important cash crops. Goats and fowl are the most common domestic animals.

Experiments conducted in the early 1970s with large-scale cooperative agricultural production were unsuccessful. Relatively low government farm prices and the high cost and scarcity of consumer goods caused many producers to return to subsistence agriculture or to resort to smuggling. The production of coffee, formerly the major cash crop, declined. Food imports of staples such as rice, once exported, remain necessary. The production of other cash crops, such as palm kernels, peanuts, pineapples, bananas, and citrus fruit, has improved only marginally since 1984, though considerable potential for expansion exists.

Commercial fishing continues to grow with the introduction of U.S., French, Japanese, and other internationally financed and operated fishing ventures. Individual small-scale riverine and marine fishing, producing fresh, dried, and smoked fish for local markets, remains important.

Forestry is hampered by the lack of adequate transportation. Mixed government and private-investment sawmills and plywood plants function below capacity because of insufficient supplies of timber, transportation difficulties, and inadequate capital and managerial input.

Mining and quarrying

Industry. Guinea depends heavily upon mineral exports to maintain a favourable trade balance. The bauxite deposits at Fria, Kindia, and Sangaredi in the Boké region are exploited by international consortia in which the Guinea government holds major shares. Similarly mixed foreign and domestic plants produce the bauxite and alumina that provide nine-tenths of Guinea's export earnings.

The iron-ore deposits of Mount Nimba are exploited under a shipping agreement with the government of Liberia. Mining of gem-quality diamonds has increased greatly since 1984, and gold production has risen substantially as well.

Food-processing plants run at less than full capacity because agricultural production is insufficient and capital and managerial input are inadequate. Most industry consists of the manufacturing of light consumer goods and the primary processing of agricultural products. Heavy industry and manufacturing is not part of the economic planning for Guinea.

Finance. Since 1984 the government of Guinea has pursued a slow process of economic reform aimed at reestablishing a free-market system. In 1986 Guinea began a process to link its currency with the French franc again after having maintained a nonconvertible currency since 1960. The government has also actively sought closer economic ties with France and other Western nations.

In mid-1985 a new banking law was passed allowing the establishment of new commercial banks to replace the publicly owned institutions (with the exception of an Islāmic bank established in 1983) that had existed under the Touré government. In December 1985 three new banks involving French participation began operation. They are the Banque Internationale pour l'Afrique en Guinée (BIAG), the Banque Internationale pour le Commerce et l'Industrie de la Guinée (BICI-GUI), and the Société Générale de Banques en Guinée (SGBG). The central bank is the Banque Centrale de la République de Guinée.

The Guinean investment code follows fairly classical lines, offering a variety of inducements to domestic and international investors in productive sectors. Benefits include waivers of import duties on capital equipment and deductions of various peripheral tax liabilities such as statutory employers' contributions. Compared with many other African nations, the extent of these investor benefits is modest.

Government revenue is derived chiefly from mining concessions, import and export duties, excise taxes, a petroleum-products tax, and taxes on commercial transactions and production. There are also various other surtaxes, stamp duties, and registration fees. Business and other licenses and personal-property, building, dwelling, and vehicle taxes are handled by the prefecture administrations. Taxes on salaries and wages contribute little revenue because few people are salaried and because many wage earners work within the government.

Trade. Trade figures are limited and sketchy. During the Touré regime smuggling of both imports and exports brought on by an unrealistic exchange rate and poor returns to agricultural producers selling in the domestic market made accurate trade figures impossible. Estimates prepared by the World Bank show that, while Guinea recorded a substantial surplus on visible trade in the 1980s, its export base was narrow and largely dependent on the mining industry for international trade earnings.

Exports of gold and diamonds, in particular, have shown substantial growth since 1984.

The principal markets for Guinean bauxite and alumina are the United States, France, and the Soviet Union, while France is the largest supplier of imports. A growing balance of payments deficit has characterized the economy in spite of merchandise surpluses in the 1980s. With virtually no private capital inflow, the government must simply fall behind in debt service payments and cut back on both investments and the yearly operating budget.

Transportation. Guinea's transportation system is largely based upon the road and railway from Conakry to Kankan. This forked axis is intersected at Mamou by a road north to Senegal. East of Kouroussa the road branches northeastward through Siguiri to Bamako, Mali. The main road continues northeast of the railhead at Kankan to Sikasso, Mali. The regional centres, like pods strung out on a vine, lie along thin lines of communication that, in turn, radiate feeder routes.

The railroad from Conakry to Kankan is a single-track metric line. Two other railways serve the bauxite mining areas, including a line linking Conakry to the Fria bauxite mines. The Boké Railway runs between Kamsar and Sangaredi.

The port facilities of Conakry are extensive. There is a channel 26 to 66 feet deep and dock space with modern loading equipment. The Sangaredi bauxite mine company maintains its own ore-exporting port at Kamsar. Coastal shipping, however, is limited.

The Gbessia international airport at Conakry serves jets of all sizes. Air Guinée operates a somewhat irregular schedule of weekly domestic flights to the hard-surfaced airports at Kankan, Labé, and Faranah and maintains occasional service to Bamako, Mali; Dakar, Senegal; Freetown, Sierra Leone; Monrovia, Liberia; and several small, unsurfaced domestic landing strips.

Administration and social conditions. *Government.* For more than 25 years under President Sékou Touré Guinea was a one-party state, in which the Democratic Party of Guinea (PDG) and the government were one. In April 1984, after Touré's death, a military group abolished the PDG and all associated revolutionary committees and replaced them with the Military Committee for National Recovery (CMRN). The CMRN supervises a government consisting of the president and various numbers of ministers, secretaries of state, and commissioners, including a number of civilians appointed by the CMRN. Each of the four supra-regions of Upper Guinea, Forest Guinea, Middle Guinea, and Maritime Guinea is headed by an official appointed by the CMRN.

The CMRN

Education. The government has put a major emphasis on improving the provision of education. In the early 1980s only about one-quarter of primary-school-age children were enrolled in school. Educational facilities at all levels had shown a marked decline in the last decade of the Touré government. Despite improvements, perhaps three-quarters of the population is illiterate in French, and the overall adult literacy rate is below average for western Africa.

Primary education is compulsory for six years beginning at age seven. Secondary education is also offered as a six-year program. Instruction is offered in French and in local languages. Private schools, previously banned, were allowed to reopen in 1984.

The country's post-secondary institutions, particularly the University of Conakry, function on an irregular basis, with assistance from France and other countries for personnel and materials.

Health and welfare. Since independence the government has made an effort to improve health care services, but infant and child mortality rates remain among the highest in sub-Saharan Africa. Equipment and supply shortages and an inadequate number of medical personnel hamper the health care system. Government-run infant clinics and child-care centres are inadequate. Most such social welfare services are either provided by the extended family or are absent.

A severe housing shortage exists in the urbanized areas, though mud and straw construction reduces the problem in rural areas. It is estimated that one-fifth of the country's population lives in Conakry and its environs, where the housing shortage is especially serious.

Cultural life. Until 1984 artistic and literary expressions were limited largely to African themes by the single political party and its leader. As a result Guinean intellectuals exhibit a strong sense of nationalism and a decolonialized state of mind. As greater openness of expression returns, a distinctly Guinean literature is gradually emerging.

A truly autonomous free press has yet to emerge as a major force in Guinean life. One French-language newspaper, *Horoya,* formerly controlled by the PDG, is published. A number of informal newsletters are also published in indigenous languages. A television service was begun in 1977, and transmissions continue on an irregular basis for short periods each day.

Press and broadcasting

The telephone network is limited and obsolescent; the few telephones in service are almost all in Conakry and other major urban centres. A program to upgrade the telecommunications system was implemented in the late 1980s.

The professional National Guinean Ballet, which emerged after independence, has retained some of the dance and music of the distinct ethnic and regional groups. Creative accomplishments in modern dance and popular music have given Guinean musicians and singers an international reputation.

Handicrafts in Guinea, as elsewhere in Africa, declined sharply during the colonial era with competition from manufactured consumer goods. The lack of tourism and creative marketing since independence has limited the amount of change and innovation in local crafts, so that the leatherwork, wood carving, and jewelry produced in Guinea tend to be more genuinely ethnic than elsewhere in western Africa.

For statistical data on the land and people of Guinea, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

HISTORY

Early history. Hunting and gathering populations occupied the area of Guinea at least 30,000 years ago, and farming has been practiced there for about 3,000 years. About 1,000 years ago Susu and Malinke (Maninka) people began to encroach on the Baga, Koniagi (Coniagui), and Nalu (Nalou) populations who had been living in the area for more than 1,000 years. The towns and villages of Upper Guinea were incorporated into the Mali empire from the mid-13th century, and by the 16th century the Fulani (Fulbe) had established domination over the Fouta Djallon. The Portuguese presence on the coast dates from the 15th century, and the slave trade continued to affect Guinea until the mid-19th century. British and French trading interests on the coast played minor roles in the historical evolution of the Guinean interior until the *almamy* (ruler) of Fouta Djallon placed his country under French protection in 1881. The independent Malinke state ruled by Samory Touré resisted the French military until 1898, and isolated small groups of Africans continued to resist the French until the end of World War I.

Colonial era. The French protectorate of Rivières du Sud was detached from Senegal as a separate colony in 1890. As French Guinea it became part of the Federation of French West Africa in 1895. Treaties with Liberia and Great Britain largely established the present boundaries by World War I.

Under the 1946 constitution of the French Fourth Republic a small number of French-educated Africans in Guinea were allowed to vote for deputies to the French National Assembly. In the 1958 referendum on the constitution for the French Fifth Republic only Guinea, under the influence of Sékou Touré, who later became the country's first president, voted against membership in the French Community and became independent.

Independence. Guinea came to occupy a special position among African states for its unqualified rejection of neocolonial control. Touré's rule grew increasingly more repressive, however. Denied French assistance, Guinea contracted loans and economic and trade agreements with

the Soviet Union and the People's Republic of China. When it failed to become a full economic partner in the Soviet bloc, Guinea turned to France and the West for capital and technical assistance in the waning years of Touré's regime. Under Touré's uncertain economic leadership, however, the potentially wealthy country did not prosper.

Throughout Touré's rule, difficulties of economic adjustment and political reorganization caused him to become increasingly obsessed with what he perceived as opposition. Probably the event that had the most negative effect was the Portuguese-backed invasion of Conakry by Guinean dissidents. Such real conspiracies, together with a myriad of imaginary ones, led to show trials, imprisonments, and executions of dissidents and other suspects. Gradually power was concentrated in the hands of Touré and his predominantly Malinke associates. Members of his own family occupied leading government posts, from which illicit earnings were drawn on a large scale. Though the Democratic Party of Guinea (PDG), which Touré had led since 1953, retained control, it ceased to enjoy the mass support it had had in the late 1950s and early 1960s. Touré's death in 1984 left party leaders with little grassroots support. The ensuing military coup began with fairly strong support from the general public.

The Military Committee for National Recovery under Colonel Lansana Conté, Guinea's second president, radically altered the composition of the government. Since 1984 Guinea has accelerated its return to the Western camp. The new regime endorsed the concept of a pluralist society. Private ownership and international investment were actively supported, while the role of the state in the economy was reduced. In the late 1980s Guinea sought reintegration into French-speaking western Africa and the Franc Zone.

For later developments in the history of Guinea, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (T.E.O'T.)

Guinea-Bissau

The Republic of Guinea-Bissau (República da Guiné-Bissau) became an independent nation on Sept. 10, 1974, after more than a decade of war. The country is bounded by Senegal to the north, by Guinea to the east and south, and by the Atlantic Ocean to the west. Including the Bijagós (Bissagos) Archipelago and other islands lying off the coast, the total area is 13,948 square miles (36,125 square kilometres). The capital is Bissau.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief.* Almost all of Guinea-Bissau is low-lying and bathed daily by tidal waters reaching as much as 62 miles (100 kilometres) inland. In the southeast, the Fouta Djallon plateau rises about 600 feet (180 metres). The Boé Hills extend from the western slopes of the Fouta Djallon to the Corubal Basin and the Gabú Plain.

The coastal area is demarcated by a dense network of drowned valleys, called rias. The Bafatá Plateau is drained by the Geba and Corubal rivers. The Gabú Plain occupies

Drainage

the northeastern portion of the country and is drained by the Cacheu and Geba rivers and their tributaries. The interior plains are part of the southern edge of the Sénégal River basin. The uniform elevation of the mature floodplain allows rivers to meander and renders the area susceptible to flooding during the rainy season.

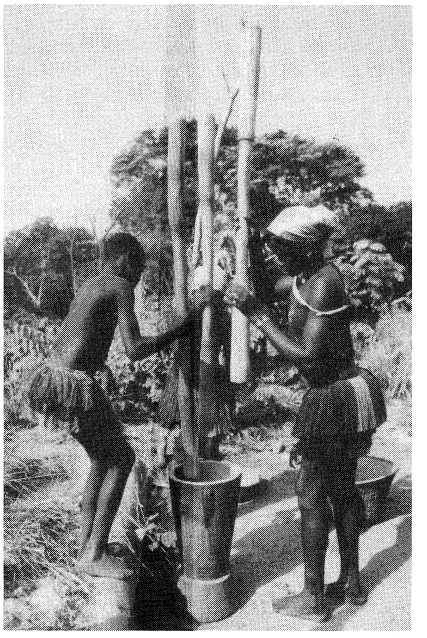
Climate. Rainfall occurs between May and October, followed by a dry season. April and May are the hottest months, with afternoon temperatures in the high 90s F (mid-30s C) at most locations. The coast has a monsoonal climate with abundant rainfall, amounting to 60 to 120 inches (1,500 to 3,000 millimetres), whereas the interior is influenced by the tropical savanna climate, with greater variation in rainfall and temperature.

Plant and animal life. The three zones of vegetation are the coastal swamps and plains that are covered with mangrove and palm trees, the heavily forested interior plain, and the savanna found in the north.

Guinea-Bissau has a great variety of aquatic birds, including the pelican and the flamingo. Crocodiles, snakes, gazelles, apes, parrots, hyenas, and leopards abound.

Settlement patterns. The large majority of Guineans live in villages. In general, the ecology places severe limitations on agriculture, but Guineans have adapted to the environment in a number of complex ways. At least 19 microecologies have been identified with as many farming systems, the common feature being rice cultivation. Coastal areas specialize in saltwater varieties grown in flooded paddies.

© Dave G. Houser



Villagers pounding rice, Bijagós Islands, Guinea-Bissau.

The Balanta Brassa, who occupy the central and southern parts of the country, are renowned as paddy rice growers. Another coastal people, the Bram (Brame), are known for their skill in farming without irrigation. The Bijagós people and the Mandyako (Manjaco) specialize in processing palm wine and palm oil, while inland Fulani (Fula) raise cattle. Fulani and Malinke (Mandinga) grow the largest volumes of peanuts (groundnuts) and cotton. Almost all peoples farm a variety of terrains, growing a number of different kinds of produce and using multicropping techniques to ensure one or more successful crops.

The war of independence against Portugal (1963–74) uprooted many peoples. Some estimate that fully half the population was affected. The Portuguese disrupted production by superimposing a form of temporary settlement—the "strategic hamlet"—to isolate as many people as possible from the nationalist forces. In the nationalist-held territories, regroupings also took place in the forests to escape air attacks from the Portuguese. More than 100,000 people fled to Senegal or Guinea to escape the war.

The people. *Ethnic and linguistic composition.* The population includes more than 20 ethnicities; the main groups are the Balanta, Fulani, Mandyako, Pepel (Papel), Bram, and Malinke. In theory, each ethnic group originally had a territory (known as *chao*), but conquest—first by the Malinke and then by the Fulani and their expansion toward the coast—the movement of the Balanta southward, and the war and postwar migrations have tended to complicate the settlement pattern.

The Balanta Brassa

The largest and most widely spread group, the Balanta Brassa, belong to a relatively egalitarian society in which patrilineage, household, age group, and gender are important divisions. They were the most receptive to nationalist slogans of emancipation from Portuguese rule. A fiercely independent people, largely animist in belief, they constituted the most notable body of guerrilla forces during the war against Portugal. A recent grass roots movement sought renovation and change in the social structure.

The Fulani, who may be divided into at least three subgroups, were originally pastoralists, but in the 19th century they conquered large sections of western Africa. Their society is Muslim and hierarchical. Largely impervious to Portuguese culture—but not to collaborating with the colonial administration—many of their leaders became tactical allies of the Portuguese army against the guerrillas, whom they saw as a threat to their religion, society, and traditions. Some smaller groups, however, joined the nationalists in order to emancipate themselves from the authority of elders and lords.

The Malinke, the ancient rulers of the Senegambia, live in stratified societies of noble families; craftsmen, traders, and other professional groups; and descendants of former captives. They also converted to Islām.

Mandyako and Pepel in the northern coastal region were among the first peoples to establish trading relations with the Portuguese. Some intermarried with them; others worked for them, adopting European customs and dress and helping to create and spread the trading language Crioulo. The Pepel, however, fiercely defended their landlord rights against the Europeans. Some of the smaller groups, such as the Biafada, Felupe, Bayot, Nalu, Susu, and Bijagós, are coastal farmers and appear to have been little influenced by either Portuguese or Islāmic culture.

Urbanized Guineans, formerly called *assimilados* and numbering only a few thousand, adapted many aspects of European culture and became chiefly civil servants or white-collar workers, professions they still occupy. A formerly significant colony of Cape Verdeans immigrated during colonial times as farmer-traders, soldiers, and administrators for the Portuguese. They played a prominent role within the nationalist leadership, seeking to unite Cape Verde and Guinea-Bissau, but lost most of their influence after the coup in 1980. Non-Africans include Portuguese, eastern and western European, Cuban, and Brazilian technical experts and Lebanese merchants.

Among the African languages spoken are two categories: the first includes the Mande-tan grouping and the Mande-fu grouping. The second, the Atlantic (West Atlantic) language group, includes all other African languages spoken in Guinea-Bissau. Apart from this mixture of some 20 languages and dialects, the lingua franca is Crioulo. It exerts a unifying influence in the rural areas and, along with Portuguese, is used in schools. Portuguese is the official language. Some Arabic is known by Muslim scholars.

Religions. Traditional animist beliefs have remained strong, even among those who have formally adopted Christianity or Islām. Christianity made only a few inroads during the Portuguese colonial period. There remain a small number of Roman Catholics and a few Protestants.

Role of Islām

Most adherents of Islām belong to the Qādirīyah or Tijānīyah orders. Portugal supported the expansion of Islām to help counteract the influence of nationalist leaders. Since independence, the government of Guinea-Bissau has joined the Islāmic Conference and receives aid from Saudi Arabia and Kuwait. The Libyan government, moreover, supports the spread of Islām in a variety of ways.

Demographic trends. The population is growing at a comparatively moderate rate for sub-Saharan Africa. It is young, with 40 percent of Guineans less than 15 years of age, portending a higher rate of growth in the near future. The average life expectancy is low, owing largely to a high infant mortality rate. About half the population is active in the work force; only about one-fourth is urban. There is a significant seasonal and permanent migration, mainly to Senegal, The Gambia, and France. (R.P./R.E.Ga.)

The economy. *Resources.* There has not been a comprehensive survey of mineral resources, but large deposits of bauxite in the east along the Guinean border and phosphates in the centre and northwest have been found. A lack of infrastructure has slowed mining. There is offshore prospecting for petroleum. The Corubal River has immense hydroelectric potential, and a project is planned on the Corubal at Saltinho.

Agriculture, fishing, and forestry. The economy is largely agricultural, with good prospects for forestry and fishery development. Rice is the main staple, and export crops include peanuts, cashews, palm products, timber, and seafood. Indigenous hardwoods are the predominant forestry export. Fishing potential is estimated at nearly 250,000 metric tons per year. Most of the fishing is done by foreign vessels under license.

Restoration of prewar levels of production has been hampered by government neglect of rural development and by trade policies that primarily benefit urban areas. Less than half the arable land is in use. Urban areas in particular have suffered food shortages, and scarce foreign exchange is used to import food.

Industry. The amount of industry is small. Most factories produce light consumer goods.

Finance and trade. The banking system is rudimentary but in the process of change. The Banco Nacional da Guiné-Bissau controls the money supply and has financed government deficits by expanding the supply. There are plans to restructure the system to include a central bank, a private commercial bank, and an agricultural credit bank. External financing of the national budget has been at the rate of nearly one-half the gross domestic product, and in the mid-1980s Guinea-Bissau went into arrears in servicing its national debt. Balance-of-payments difficulties increased along with the external debt. Since 1983 the government has made an effort to reorient investment priorities, reduce public spending, privatize the market, and raise producer prices. In 1987 it began a structural adjustment program under the supervision of the International Monetary Fund. Portugal has agreed to finance limited convertibility of the country's currency, the peso, and the government has applied to join the Franc Zone.

Transportation. River transport accounts for most of the commerce in the south, whereas northern and eastern areas are served mainly by roads. Both networks centre on Bissau and need upgrading. There are few paved roads and no rail lines. Bissau is the only port for oceangoing vessels, but several other coastal and river ports are regularly served by barges. The airport at Bissalanca (near Bissau) handles international air traffic, while several smaller airports and landing strips serve the interior.

Administration and social conditions. *Government.* Guinea-Bissau is divided administratively into eight *regiões* ("regions"), 37 *sectores* ("sectors"), and the autonomous sector of Bissau. The most basic unit of government is the village or neighbourhood committee; it is also the basic unit of the party apparatus. These units are linked with committees at section, sector, and regional levels that respond to the national levels of government and the party. A new provincial level of government has been inserted between the regional and national levels. The highest authority of government is the 150-member National People's Assembly or, when not in session, the Council of State. The supreme authority of the party, the African Party for the Independence of Guinea and Cape Verde (PAIGC), is its congress and, in the interim, the political bureau.

Local administration

Education. The government has concerned itself with providing a compulsory universal basic education consisting of six years. For those children who show scholastic promise there are five years of secondary education. There are schools for teacher training, nursing, and vocational training. Only some two-fifths of school-age children attend school, however, and adult illiteracy remains high.

Health and welfare. The state of health in Guinea-Bissau is poor. The health care system aims at providing basic medical care for all citizens and is engaged in training health workers in every village. The goal is a health post in each section, a sectoral referral system, and a regional hospital within the reach of every citizen. There are two national hospitals and one centre for psychiatric care in Bissau. The financing of the health care system is heavily dependent upon foreign assistance.

Cultural life. Cultural life in Guinea-Bissau is mainly organized by the government. A state radio station exists, and an experimental television program is run in conjunction with the Portuguese broadcasting system. The government publishes its own newspaper, *Nô Pintcha.* The National Institute of Studies and Research (INEP) sponsors social and scientific investigation and publishes *Soronda,* a journal of Guinean studies. The national arts

institute maintains a school of music and dance and sponsors frequent concerts. There is a public library and museum.

For statistical data on the land and people of Guinea-Bissau, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (R.E.Ga.)

HISTORY

Early history. For more than a thousand years the coast of Guinea-Bissau has been occupied by iron-using agriculturists. They were particularly skilled in the production of irrigated and dry rice and were suppliers of marine salt to adjacent areas of western Sudan. From the 13th century coastal farmers came increasingly into contact with the outside world, first from the landward side and later from the seaward. The earliest recorded influences are associated with the breakup of the Ghana empire when displaced peoples sought refuge near the coast. Later the region was loosely drawn into the sphere of the Mali empire, and regional governors called *farim*s were appointed to impose some form of allegiance to the great Mande ruler.

Portuguese contact

Overseas contacts with the Guinea Coast were opened by the Portuguese, starting in the 1440s. Guinea played an important role in the colonization of the Cape Verde Islands from this period. Slave labour was first used to establish plantations of cotton and indigo, and then skilled Guinea craftsmen were introduced to establish a weaving and dyeing industry. Much of the cloth was sent back to the mainland for the purchase of slaves destined for the Americas. The transatlantic slave trade was facilitated by Portuguese and mulattoes called *lançados* who acted as intermediaries between the Guinean rulers and the visiting slave ships. In the 16th century the expansion of Mande-speaking peoples into the upper Guinea Coast area caused wars that greatly increased the number of prisoners available for export as slaves. In addition to the slave trade the country conducted some trade in salt, kola nuts, and food to the interior and ivory, wax, dyewood, and hides overseas. The main overseas buyers came from Portugal, Britain, Holland, and France.

During the next four centuries, when the slave trade was the main economic activity of the country, the people of Guinea had little difficulty in preventing or restricting the attempts of foreign powers to establish territorial claims. A post established at Cacheu by Cape Verde traders in 1588 was given periodic support by the Lisbon government during the 17th century but did not expand. In 1687 a Portuguese post was established at Bissau in an attempt to limit French commercial competition by political, diplomatic, and military means, but that too failed to survive. In 1792 the English briefly and disastrously held a settlement at Bolama. Meanwhile the Portuguese had reestablished a base at Bissau and during the 19th century increasingly came to regard the coast on either side as sovereign territory.

The colonial era. The Portuguese territorial claim in Guinea was disputed by both the British and the French. Periodic negotiation first of all excluded the British (1870) and then settled the boundaries with the French-claimed territories (1886 and 1902–05). These frontier agreements were followed by the slow and sometimes violent imposition of Portuguese colonial rule. The final "pacification" campaigns were fought by João Teixeira Pinto in 1913–15. These wars were followed by nearly half a century of predominantly peaceful Portuguese administration. But with the rise of African nationalism after World War II and the gaining of independence by the neighbouring territories, Guineans again began to challenge their colonial rulers. Nationalist attacks on Portuguese administrative and military posts were instigated in July 1961 by guerrillas of the PAIGC, led by Amílcar Cabral. In August Cabral declared at Conakry, capital of the French-speaking Republic of Guinea, that political endeavours to obtain the liberation of Portuguese Guinea and the Cape Verde Islands from Portuguese domination would be replaced by armed struggle. Bitter guerrilla warfare ensued between the PAIGC National Liberation Army (about 10,000 men) and the Portuguese armed forces (about 30,000 strong). The guerrillas were unable to occupy the coastal towns and river ports, but by 1971 they were firmly established in the interior, especially in the areas adjacent to the republics of Senegal and Guinea.

Independent Guinea-Bissau. By 1974, one year after Cabral's assassination, a military stalemate divided the African-ruled provinces of Guinea from the European-ruled towns. The disillusioned Portuguese army overthrew the civilian dictatorship in Lisbon and chose the former military commander in Guinea, General António Ribeiro de Spínola, to govern Portugal and negotiate independence for the African colonies. Cabral's Cape Verdian half-brother, Luís de Almeida Cabral, became president of Guinea-Bissau, but relations between the creolized middle class from Cape Verde and the poorer, less educated indigenous population of the coast created rising political tension. Eventually a coup d'état overthrew the government of Luís Cabral and severed the PAIGC party links between mainland and islands. The new head of state, João Bernardo Vieira, faced the problem of restoring cultural, educational, and political ties with Portugal in order to build up the economy, satisfy the material aspirations of the military and the bureaucracy, and spread commercial opportunity to the farmers of the isolated rural zones. Sharing meagre resources between rival claimants created a climate of friction that was not always conducive to social development or political maturation.

End of links with Cape Verde

For later developments in the history of Guinea-Bissau, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (D.Bi.)

Liberia

Liberia is the only black state in Africa never subject to colonial rule, and it is the oldest republic on the continent. Located on the western African coast in the tropical rain forest, with an area of about 38,250 square miles (99,067 square kilometres), Liberia is bounded by English-speaking Sierra Leone to the northwest; French-speaking, Guinea and Côte d'Ivoire to the north and to the east, respectively; and the Atlantic Ocean to the south and west. Monrovia, a port, is the capital. In 1973 Liberia and Sierra Leone organized the Mano River Union for economic cooperation; Guinea joined in 1980.

PHYSICAL AND HUMAN GEOGRAPHY

Physical regions

The land. *Relief.* The four physiographic regions of Liberia parallel the coast. The coastal plains are about 350 miles (560 kilometres) long and extend up to 25 miles inland. They are low and sandy, with miles of beaches interspersed with bar-enclosed lagoons, mangrove swamps, and a few rocky promontories—the highest being Cape Mount (about 1,000 feet [305 metres] in elevation) in the northwest, Cape Mesurado in Monrovia, and Cape Palmas in the southeast. Parallel to the coastal plains is a region of rolling hills some 20 miles wide with an average maximum elevation of about 300 feet; a few hills rise as high as

Shostal Associates/Superstock

The Executive Mansion in Monrovia, Liberia.

500 feet. It is a region suitable for agriculture and forestry. Behind the rolling hills, most of the country's interior is a dissected plateau with scattered low mountains ranging from 600 to 1,000 feet in elevation; some mountains rise to 2,000 feet. A striking feature of the mountainous northern highlands along the Guinea frontier is Mount Nimba.

Drainage. The Mano and Morro rivers in the northwest and the Cavalla in the east and southeast are major rivers and form sections of Liberia's boundaries. Other major rivers are the Lofa in the north and, moving southward, the St. Paul, St. John, and Cestos, all of which parallel each other and flow perpendicular to the coast. The Farmington River is a source of hydroelectric power. Waterfalls, rapids, rocks, and sandbanks occur frequently in upstream sections of most rivers, inhibiting river traffic, and limiting navigation inland to short distances. During the rainy season there is often severe flooding in the coastal plains.

Liberia forms part of the West African Shield, a rock formation 2.7 to 3.4 billion years old, composed of granite, schist, and gneiss. In Liberia the shield has been intensely folded and faulted and is interspersed with iron-bearing formations known as itabirites. Along the coast lie beds of sandstone, with occasional crystalline-rock outcrops. Monrovia stands on such an outcropping, a ridge of diabase (a dark-coloured, fine-grained rock).

Soils. Four types of soil are found in Liberia. Latosols of low to medium fertility occur in rolling hill country and cover about 75 percent of the total land surface. Shallow, coarse lithosols, in the hilly and rugged terrain, cover about 16 percent of the land. Infertile regosols, or sandy soils, are found along the coastal plains. Highly fertile alluvial soils represent only about 3 percent of the land area and are utilized largely for agriculture.

Climate. The climate, especially on the coast, is warm and humid year-round, dominated by a dry season from November to April and by a rainy season from May to October. The dusty and dry harmattan (desert winds) blow from the Sahara to the coast in December, bringing relief from the high relative humidity. Deforestation and drought in the Sahel have affected the climate, lengthening the dry season by almost a month in some areas.

Temperature and rainfall

Mean annual temperatures range between 65° F (18° C) in the northern highlands to 80° F (27° C) along the coast. Rainfall is irregular, and the rainy season varies in intensity and begins earlier at the coast than in the interior. The greatest amount of rainfall, 205 inches (5,200 millimetres), occurs at Cape Mount and diminishes inland to about 70 inches on the central plateau. The interior has hot but pleasant days and cool nights during the dry season.

Plant and animal life. Liberia has year-round evergreen vegetation. Many trees—such as red ironwood, camwood, whismore, teak, and mahogany—are valuable, but occur with other species, preventing easy harvest. Other trees of value are rubber, cocoa, coffee, and the raffia palm.

Liberia's rain forest abounds with animals such as the monkey, chimpanzee, small antelope, pygmy hippopotamus, and anteater. Elephants, bush cows (short-horned buffalo), and leopards are gradually disappearing. There are many reptiles, including three types of crocodiles and at least eight poisonous snakes. There are several unique species of bats and birds, and scorpions, lizards, and fish are numerous.

Settlement patterns. The present pattern of population distribution in Liberia is both a reflection of its migration history and a response to such social, economic, and cultural factors as war, employment, and superstition.

Migrants from north-central Africa, who began to arrive in the 13th century, originally settled in the hinterlands but were driven by overcrowding to the coast. Immigrants from the United States and the West Indies, and from neighbouring African countries, also settled on the coast. The former migrated mostly to selected areas such as Monrovia (the oldest immigrant settlement), Buchanan, Edina, Greenville, Harper, Robertsport, and Marshall. Scattered settlements were created along newly constructed or improved roads, while plantation and mining activities encouraged larger settlements in a few interior and coastal areas. There are more than 2,000 villages, the majority of which are concentrated in central Liberia, in the northwest, and in the coastal region near Monrovia. The predominantly forested regions of south-central and northern Liberia have remained sparsely populated. The trend toward urbanization has had little impact on these villages. The result has been the segmentation of Liberian society into two coexisting subsystems—traditional-rural and modern-urban.

Monrovia

Monrovia, founded in 1822, is the focal point of political, economic, and cultural activities. Situated on the left bank of the St. Paul River on the ridge formed by Cape Mesurado, it commands an imposing view of the Atlantic Ocean and the coastal plains. The city and its outlying districts and suburbs occupy five square miles. The old style of architecture that once characterized it, reminiscent of that of the southern United States before 1860, is giving way to contemporary styles. All of the ethnic groups of Liberia are represented in its population, as are refugees, African nationals from other countries, and Europeans.

The people. *Ethnic and linguistic composition.* The people of Liberia are classified into three major groups: the indigenous people, who are in the majority and who migrated from the western Sudan in the late Middle Ages; black immigrants from the United States (known historically as Americo-Liberians) and the West Indies; and other black immigrants from neighbouring western African states who came during the anti-slave-trade campaign and European colonial rule. The Americo-Liberians are most closely associated with founding Liberia. Most of them migrated to Liberia between 1820 and 1865; continued migration has been intermittent. Americo-Liberians controlled the government until a military coup in 1980.

The 16 ethnic groups may be classified into three linguistic groups: the Mande, Kwa, and Mel (West Atlantic). The Mande are located in the northwest and central regions of Liberia and also in Senegal, Mali, Guinea, and Sierra Leone. Prominent among them are the Vai, who invented their own alphabet and who, in addition, use Arabic and English; the Kpelle, the largest Mande group, who are also found in Guinea; Loma (also found in Guinea); Gbandi; Gio (Dan); Mano; Mende; and Malinke. The Kwa include the Bassa, the largest group in this category and the largest ethnic group in Monrovia; the Kru and Grebo, who were among the earliest converts to Christianity; the De; Belle (Belleh); and Krahn. The Kwa-speaking group occupies the southern half of the country. The Mel group includes the Gola and Kissi, who are also found in Sierra Leone and are known to be the oldest inhabitants of Liberia. These people live in the north and in the coastal region of the northwest.

Religion. Liberians are a religious people. About two-thirds of the people are Christian, about 15 percent are Muslim, and almost one-fifth profess other religions, primarily traditional beliefs. The largest number of Christians are the Kpelle, followed by the Bassa. The Muslims are found predominantly among the Mande peoples in the northwest region of the country.

Demographic trends. More than two-fifths of the population of Liberia is under 15 years of age; only about 5 percent is older than 65. The annual rate of natural increase is one of the highest in sub-Saharan Africa, and, because emigration is negligible, the country's annual rate of population growth is one of the highest in the world. Life expectancy, about 54 years for males and 56 years for females, is high by African standards. About 40 percent of the population lives in urban communities, and there is a high rural-to-urban movement, especially to Monrovia. Other destinations include enclaves around rubber plantations and mines.

The economy. The Liberian economy is predominantly agrarian, and raw materials, equipment, and consumer goods are imported. Production for export is carried out on a large scale through foreign investment in rubber, forestry, and mining. Foreign ships registering under a Liberian "flag of convenience" have made Liberia the world's foremost nation in registered shipping tonnage. Liberia nevertheless remains a primarily agricultural and underdeveloped country. The distribution of wealth is uneven, the coastal districts receiving a greater share of

economic benefits than the hinterland, after which the administrative centres are the next beneficiaries.

After the mid-1970s the once-vibrant economy took a sharp downturn. Between 1976 and 1980 sluggish demand and low prices stagnated the economy and the annual growth rate plunged. But gradual signs of recovery have appeared, especially in agriculture and forestry. There continues to be a heavy reliance on iron ore and forest products as major export commodities.

Liberia's economy is mixed and there is no nationalization of industry. The government, which is the largest single employer, operates several public corporations. There is a national Federation of Labour Unions, a federation of trade unions, and several other employees' unions, but no employers' association.

About 70 percent of the work force is employed in agriculture; the rest work in manufacturing, sales, services, and administration and management. About 40 percent of the total labour force is made up of women. More women than men are employed in agriculture.

The U.S. dollar, previously legal tender in Liberia, is no longer in circulation. The value of the local Liberian dollar retains parity with the U.S. dollar, however. Government revenues are derived from income, profits, property, domestic transaction, foreign trade, and maritime taxes. About one-third of economic development funding is derived from foreign sources, both bilateral and multilateral.

Liberia is a member of two regional economic unions—the Mano River Union, a free trade group to which Sierra Leone and Guinea also belong, and the Economic Community of West African States (ECOWAS).

Resources. Liberia is rich in natural resources. It is among the leading producers of iron ore, which it has produced since 1951, in Africa, and it is one of the principal exporters of iron ore in the world. Its sizable reserves are found primarily in four areas: the Bomi Hills, the Bong Range, the Mano Hills, and at Mount Nimba, where the largest deposits occur. Other minerals include diamonds, gold, lead, manganese, graphite, cyanite (a silicate of aluminum, with thin bladelike crystals), and barite. There are possible oil reserves off the coast.

Minerals

There is vast potential for the development of hydroelectric power. About half of Liberia's electricity is from hydroelectric sources. The Mount Coffee hydroelectric station outside Monrovia on the St. Paul River is the country's largest hydroelectric installation.

Water supplies have been improved in both rural and urban areas so that some 40 percent of the population has access to potable water. Surface water is abundant, and groundwater reserves are ample and regularly replenished by the country's heavy rainfall.

Agriculture, forestry, and fishing. Agriculture is the fastest-growing sector of the economy. About half the land area is suitable for cultivation, though less than 5 percent is actually cultivated. Commercial farms are often operated by foreigners. Traditional farms, which comprise the largest number, are usually cultivated by slash-and-burn methods. Traditional farming, though less capital-intensive, contributes nearly twice as much to Liberia's gross domestic product as commercial farming.

Traditional farmers practice mixed cultivation of rice, cassava, and vegetables and keep goats, sheep, chickens, and ducks. Cultivation of cash crops such as coffee, cocoa, oil palm, sugarcane, and swamp rice is increasing. Rice production meets about 75 percent of the country's needs. The rest is imported, principally from the United States.

Liberia's climate favours rubber production, which thrives on the country's poor soils. In 1926, the Firestone Tire and Rubber Company of the United States obtained a concession for rubber cultivation. Rubber has become by far the country's most valuable commercial crop, with coffee and cocoa increasing in importance. Kola nuts, peanuts, and cotton are also produced, and cattle and pigs are raised.

Rain forests produce fine hardwood timber, especially in the east of the country, but also in the centre and in the west. Timber concessions operate in the southeast and northwest. Substantial amounts of timber are produced, but exploitation of the forest resources is difficult because of poor roads and shortage of labour. Of the approximately 250 species of forest trees about 90 are marketable. Forest depletion continues despite government reforestation regulations.

Deep-sea fishing is important, and the catch is largely mackerel, barracuda, and red snapper. Kru and Fanti fishermen, the latter from Ghana, have traditionally been the suppliers of fish to coastal areas but are supplemented by Liberian fishing companies. Inland fish-breeding ponds provide a source of protein.

Industry. To export the ores, iron interests have built railroads connecting the mines with Monrovia and Buchanan. Iron ore is extracted by open-pit mining, while gold and diamonds are extracted by placer mining. Traditional, small-scale mining for gold and diamonds continues.

Manufacturing enterprises have increased greatly since 1960. Predominantly private and foreign-owned, most serve the local market. Near Monrovia there is a petroleum refinery as well as a cement plant. There are also explosives, paint, pharmaceutical, and cosmetics plants. Bricks, tiles, cement blocks, lumber and furniture, soap, and footwear are also manufactured, and there are several distilleries. The problem of foreign exchange and the high cost of raw materials for these industries cause frequent market shortages, and the failure rate among manufacturing businesses is high.

Manufacturing

Liberia's tourist industry has much potential that has yet to be developed. Tourist facilities are concentrated near beaches in Monrovia and Robertsport and near Lake Piso. The Lakpazee Zoo and the National Cultural Center at Kendeja, Providence Island near Monrovia, and the Kpatawee Waterfalls on the Zor River near Suakoko are the principal tourist attractions.

Finance and trade. Among the several government-sponsored banks are the National Bank of Liberia, the National Housing and Savings Bank, the Agricultural and Cooperative Development Bank, and the Liberia Bank for Development and Investment. In addition there are private banks, insurance companies, and credit unions.

Countries of the European Economic Community, especially West Germany, and the United States are the principal markets for Liberian exports. Iron ore and rubber account for almost 75 percent of Liberian export earnings, followed by logs and timber, diamonds, coffee, and cocoa. Food is the primary import; others include machinery and transport equipment, beverages, tobacco, manufactured goods, fuels, lubricants, and chemicals. The United States and West Germany are the largest suppliers of imports.

Transportation. Only a small percentage of Liberian roads are paved. Primary roads connect administrative and economic centres and provide access to the road systems of neighbouring countries.

Monrovia is the principal commercial port, and it also has facilities for transshipping iron ore and liquid latex. Nimba Range iron ore is shipped from Buchanan, while the ports at Greenville and Harper are used primarily for the shipment of rubber and forest products. All ports are administered by the National Port Authority.

Liberia has two major airports, Robertsfield International, and James Spriggs Payne Airport, both near Monrovia. More than 100 airfields and airstrips dot the country's interior.

Administration and social conditions. *Government.* Liberia's government is patterned after that of the United States with the executive, legislative, and judicial branches. The military regime that followed the 1980 coup drafted a new constitution that replaced the one of 1847. It became effective in 1986 with the inauguration of Samuel K. Doe as the first president of the Second Republic. The 1986 constitution provides for direct election by popular vote of the president and vice president for a four-year term, with a limit of two consecutive terms, and grants the president emergency powers. A minister for presidential affairs serves as acting president in consultation with the vice president and the cabinet in the absence of the president from Monrovia.

Members of the bicameral National Legislature, who serve six-year terms in the House of Representatives and nine-year terms in the Senate, are also elected directly.

The constitution calls for a multiparty system. There are five political parties—the ruling National Democratic Party, the Unity Party, the Liberian Action Party, the Liberia Unification Party, and the United People's Party. The Supreme Court is headed by a chief justice assisted by five associate justices, all of whom are appointed by the president.

Each of the country's 11 counties and two territories is headed by a superintendent, who is the direct representative of the president and has a council. The mayor of the capital city of the county also has a council. The counties are divided into districts, chiefdoms, and clans administered by county commissioners and paramount and clan chiefs. The ethnic peoples are allowed, as far as possible, to govern themselves according to customary law. County Development Associations assist in the delivery and improvement of government services.

Education. Since 1939 education has been compulsory for children between the ages of six and 16 and is free at the primary and secondary levels. In 1974 Liberia became a full member of the West African Council in order to provide an international yardstick for measuring the quality of its education.

The government provides for the education of teachers and sponsors the employment of foreign teachers. International aid has also enabled the government to expand the quality and availability of education. There are several vocational schools, including the Booker Washington Agricultural and Industrial Institute at Kakata, a government school. Advanced training is provided at the University of Liberia (1951) in Monrovia, at Cuttington University College (1889) in Suakoko (Episcopalian), and at the William V.S. Tubman College of Science and Technology (1978) in Harper. Several community colleges have also been established in the Monrovia area. The Monrovia Torrino Medical College trains paramedical students. Liberians who study abroad receive advanced training under a government foreign scholarship program and from donor agencies.

Vocational schools, colleges, and universities

Health and welfare. Conditions remain poor, although much progress has been made in providing better health facilities. Malaria and leprosy are major health problems, and influenza, hernias, intestinal worms, trypanosomiasis (sleeping sickness), schistosomiasis, and elephantiasis are also prevalent. The government conducts inoculation campaigns to combat smallpox and yellow fever. Yaws is no longer a serious problem, following a World Health Organization control program, but tuberculosis has been difficult to eradicate. Dysentery, malaria, and diarrhea are the most important causes of infant mortality, which, at about 120 per 1,000 births, is high by world standards but about average for western Africa.

Modern hospitals include the John F. Kennedy Memorial Hospital in Monrovia, sponsored by the U.S. Agency for International Development, the Roman Catholic Hospital at Monrovia, and Phebe Hospital in Bong county, all of which have nursing schools and medical teaching units attached to them.

The Ministry of Public Works supervises low-cost housing projects. Housing is expensive, and most building equipment is imported.

Intermarriage and economic progress have been important in breaking down social divisions. Mines and plantations located throughout the country have helped the slow trend toward a more equitable distribution of income. In the coastal districts, government jobs, foreign businesses, and local markets provide greater opportunities for economic and social advancement.

Cultural life. Traditional and Western life-styles coexist; however, traditional values, customs, and norms influence the Western type considerably. In cities both Western and African music and dancing styles are in vogue, but in rural areas, traditional rhythms are favoured. Schools instruct students in the legends, traditions, songs, arts, and crafts of African culture, and the government promotes African culture through such agencies as the National Museum in Monrovia, the Tubman Center for African Culture in Robertsport, and the National Cultural Center in Kendeja, which exhibits architecture of the 16 ethnic groups of Liberia. Mask making is an artistic pursuit that is also related to the social structure of some ethnic groups. Musical festivals, predominantly religious, are held in most communities. The University of Liberia has an arts and crafts centre. There are several libraries, including a children's library in Monrovia and a National Public Library.

Association football (soccer) is the most popular sport. An intercounty football competition is held for the annual championship. The University of Liberia and Cuttington University College hold annual sports competitions.

Monrovia has five daily newspapers including the *Daily Observer,* the largest and most prestigious. A few magazines are published annually. Officially, there is press freedom, but newspapers are banned occasionally for violating government policies on information.

There are four radio stations and one television station. International telecommunication services are available through direct satellite links between Liberia, the United States, Italy, and France.

For statistical data on the land and people of Liberia, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (A.B.J.)

HISTORY

Outsiders' knowledge of the west of Africa began with a Portuguese sailor, Pedro de Cintra, who reached the Liberian coast in 1461. Subsequent Portuguese explorers named Grand Cape Mount, Cape Mesurado (Montserrado), and Cape Palmas, all prominent coastal features. The area became known as the Grain Coast because grains of Melegueta pepper, then as valuable as gold, were the principal item of trade.

The Grain Coast

In the beginning of the 19th century the tide started to rise in favour of the abolition of slavery, and the Grain Coast was suggested as a suitable home for freed American slaves. In 1818 two U.S. government agents and two officers of the American Colonization Society (founded 1816) visited the Grain Coast. After abortive attempts to establish settlements there, an agreement was signed in 1821 between the officers of the society and local African chiefs granting the society possession of Cape Mesurado. The first American freed slaves landed in 1822 on Providence Island at the mouth of the Mesurado River. They were followed shortly by Jehudi Ashmun, a white American, who became the real founder of Liberia. By the time Ashmun left in 1828 the territory had a government, a digest of laws for the settlers, and the beginnings of profitable foreign commerce. Other settlements were started along the St. John River, at Greenville, and at Harper. In 1839 Thomas Buchanan was appointed the first governor. On his death in 1841 he was succeeded by Joseph Jenkins Roberts, a black man born free in Virginia in 1809; Roberts enlarged the boundaries of the territory and improved economic conditions.

Joseph Jenkins Roberts

The early republic. When the American Colonization Society intimated that Liberia should cease to be dependent upon it, Roberts proclaimed it an independent republic in 1847. Independence was recognized in 1848–56 by most of the great powers, though formal recognition by the United States did not come until 1862.

At the time independence was declared, a constitution based on that of the United States was drawn up. Roberts, who had been elected the first president of the republic, retained that office until 1856. During this period the slave trade, hitherto illicitly carried on from various nominally Liberian ports, was ended by the activity of the British and U.S. navies.

In 1871 the first foreign loan was raised, being negotiated in London nominally for £100,000. The loan was unpopular, and still more unpopular was the new president, Edward J. Roye, who was deposed and imprisoned at Monrovia. Roberts was called back to office. He served until 1876.

During the early days of Liberia there were constant frontier troubles with the French on the Ivory Coast and the British at Sierra Leone. The Liberians tried to extend their authority inland, although they were still unable to control all the coastal area they claimed. Efforts to end the frontier disputes resulted in treaties with Great Britain in

1885 and with France in 1892. In 1904 President Arthur Barclay, who was born in Barbados, initiated a policy of direct cooperation with the tribes. Having obtained a loan from London in 1907, he made real efforts at reform. The foreign debt, however, was a burden, and the government was unable to exert effective authority over the interior for more than 20 miles inland. In 1919 an agreement was signed transferring to France 2,000 square miles of hinterland that Liberia had claimed but could not control.

Outside intervention. In 1909 a commission appointed by President Theodore Roosevelt of the United States investigated political and economic conditions in Liberia and recommended financial reorganization. A loan of $1.7 million, secured by customs revenue, was raised by an international consortium of bankers in 1912, and a receivership of customs was set up, administered by appointees of the British, French, and German governments and a U.S. receiver-general. A frontier police force was organized by officers of the U.S. Army, with the result that Liberian authority was better maintained. Unfortunately, this promising new regime was upset by World War I. Revenues dropped to one-fourth of normal, and the financial situation steadily deteriorated.

Firestone rubber concession

In the 1920s the Firestone Tire and Rubber Company investigated the possibility of growing rubber in Liberia and, finding the environment favourable, obtained a concession of one million acres (400,000 hectares) in 1926. At the same time, a loan of $5 million was arranged through the Finance Corporation of America, a Firestone subsidiary. With this private loan, only half of which had been taken up by 1945, the Liberian government consolidated and bonded all its external and internal debts and placed the country's finances on a relatively stable basis. Administration of the customs and internal revenue was placed in the hands of a U.S. financial adviser. In 1952 the government was able to liquidate its foreign debt for the first time since accepting the English loan of 1871.

An investigation by the League of Nations of forced labour and slavery in Liberia, involving the shipment of Africans to the Spanish plantations in Fernando Po, brought about the resignations of President Charles King and Vice President Allen Yancy and the election of Edwin Barclay to the presidency in 1931. Liberia appealed to the Council of the League of Nations for financial aid, and a commission of inquiry was established. The next three years were marked by unsuccessful attempts to work out a plan of assistance involving the appointment of foreign administrators, the declaration of a moratorium on the Firestone loan, and the suspension of diplomatic relations with Great Britain and the United States. After the League Council had finally withdrawn its plan of assistance, the Liberian government reached an agreement with Firestone along lines similar to the League's recommendations.

World War II and after. The new significance of Liberia became apparent after the outbreak of World War II in 1939. During the war Liberia's rubber plantation was the only source of natural latex rubber available to the Allies, apart from plantations in Ceylon. In 1942 Liberia signed a defense agreement with the United States for the operation of strategic airports and military protection of the installations. This resulted in the undertaking of a program of strategic road building, the construction of the airport now known as Robertsfield International Airport, and the completion of a deepwater harbour at Monrovia. U.S. money was declared legal tender in Liberia in 1943, replacing British West African currency. In 1943 William V.S. Tubman was elected to his first term as president. Liberia declared war against Germany and Japan in January 1944 and in April signed the declaration of the United Nations. In December 1960 Liberia became a member of the UN Security Council and from that time took an active part in African and international affairs. In 1963 Liberia became a member of the UN Trusteeship Council as well as a member of the Organization of African Unity at its inception.

In 1963 Tubman was elected to serve his fifth term as president. In May 1964 the United States and Liberia signed an agreement providing for the transfer of the free port of Monrovia, built with U.S. lend-lease funds during World War II, to the government of Liberia. Tubman was again elected president in 1967, the only candidate for the office. Tubman died in London on July 23, 1971, shortly after his election to a seventh term as president. He was immediately succeeded by Vice President William R. Tolbert.

A decline in world prices for Liberia's chief exports, iron ore and natural rubber, brought financial hardship to the country during the 1960s and early 1970s. Foreign loans, especially by the United States, helped sustain the economy during that period.

On April 12, 1980, Tolbert was killed in a coup led by Master Sergeant (later General) Samuel K. Doe, who became head of state and chairman of the People's Redemption Council (PRC). The PRC promised a new constitution, which became effective in 1986, and a return to civilian rule. Elections were held in 1985 with the participation of several parties but were widely criticized as fraudulent. Doe was inaugurated as the first president of the Second Republic in January 1986.

For later developments in the history of Liberia, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (D.R.P./S.E.H.)

Nigeria

Nigeria—in full, Federal Republic of Nigeria—is the largest of the western African coastal states. Its population is the largest of any country in Africa and one of the largest in the world. With an area of 356,669 square miles (923,768 square kilometres), it is the 14th largest state on the continent. It is bordered on the north by Niger, on the east by Chad and Cameroon, on the south by the Gulf of Guinea, and on the west by Benin. The national capital of Lagos, located on the southwest coast, is scheduled to be replaced by the centrally located new town of Abuja, in the Federal Capital Territory, which was created by decree in 1976.

Modern Nigeria dates from 1914, when the British protectorates of Northern and Southern Nigeria were joined. The country became independent on Oct. 1, 1960, and in 1963 adopted a republican constitution but elected to stay a member of the Commonwealth. English remains the official language, since more than 200 different languages are spoken by the many groups living in the country. From 1980, however, Igbo (Ibo), Hausa, and Yoruba, the languages spoken by the three largest ethnic groups, could also be used for official transactions at the federal level. Cultural contacts with the more distant English-speaking countries of Ghana and Sierra Leone, and more recently with Liberia, remain stronger than those with the adjacent French-speaking Benin, Niger, and Cameroon.

PHYSICAL AND HUMAN GEOGRAPHY

Plateaus and undulating plains

The land. *Relief.* Nigeria consists of several eroded surfaces, occurring as plateaus, at elevations of 2,000 feet (610 metres), 3,000 feet, and 4,000 feet above sea level. The coastal areas, including the Niger Delta, as well as the Lake Chad Basin and the western parts of the Sokoto region in the far northwest, are underlain by young, soft sedimentary rocks. Gently undulating plains, which become waterlogged during the rainy season, are found in these areas. In much of the southwest and in most of the north central part of the country, the underlying rocks are crystalline, and the characteristic landforms are high plains with broad, shallow valleys dotted with numerous hills or inselbergs (steep-sided masses of rock left after erosion). Dome-shaped isolated hills and elongated ridges occur in the sandstone plateau region near the confluence of the Niger and Benue rivers.

The Nsukka–Okigwe cuesta, northeast of the Niger Delta, contains the Enugu (Enugu–Okigwe) escarpment, which rises abruptly for about 600 feet above the Cross River plains. Other prominent relief forms include the Jos Plateau and the Biu Plateau, both of which consist of extensive lava surfaces dotted with many extinct volcanoes.

Drainage. There are three major drainage areas in Nigeria—the Niger–Benue Basin; the Lake Chad Basin; and the coastal, or Gulf of Guinea, basin. The Niger River,

for which the country is named, and the Benue, its largest tributary, are the principal rivers. The Niger has many rapids and waterfalls, but the Benue is not interrupted by either and is navigable throughout its length except during the dry season. Rivers draining the area north of the Niger–Benue trough include the Sokoto, the Kaduna, the Gongola, and the rivers draining into Lake Chad. The coastal areas are drained by short rivers that flow into the Gulf of Guinea. River basin development projects have resulted in the creation of many large man-made lakes, including Lake Kainji on the Niger and Lake Bakolori on the Rima.

The Niger Delta is a vast low-lying region through which the waters of the Niger River drain into the Gulf of Guinea. Oxbow lakes, river meander belts, and prominent levees are characteristic landforms in this region. Large freshwater swamps give way to brackish mangrove formations near the seacoast.

Soils. There are four major soil groups: (1) the ferruginous tropical soils (red and brown), which cover the largest area and are found primarily in the regions underlain by crystalline rocks, (2) the ferralsols, which are localized in areas of older sedimentary rocks, (3) the regosols and brown soils, developed on loose sandy materials consisting of wind-borne deposits and riverine sands and found primarily in the dry north, and (4) the hydromorphic and organic soils, confined largely to areas of young sedimentary rocks along the coast and river floodplains. The most productive soils agriculturally belong to the first group. Along the coast and in the forest belt, where the rainfall is greatest and the soils are heavily leached, the soils readily lose their fertility. In parts of the north, where there is a marked dry season, a dense surface layer of laterite develops, making these soils difficult to cultivate.

Climate. Nigeria has a tropical climate with rainy and dry seasons. It is hot and wet throughout the year in the southeast but dry in the southwest and farther inland. Three climatic patterns are distinguished: (1) a tropical wet climate in the southeast, with high temperatures and heavy rainfall throughout the year; (2) a tropical wet and dry, or savanna, climate in the north and west; and (3) a dry, or steppe, climate in the far north.

In general, the length of the rainy season decreases from south to north. In the south the rainy season lasts from March to November; in the far north it lasts only from mid-May to September. In August there is a marked interruption in the rains in the south, resulting in a short dry season often referred to as the "August break."

Rainfall is heavier in the south, especially in the southeast, which receives more than 120 inches (3,050 millimetres) of rain a year, as compared with 70 inches in the southwest. Rainfall decreases progressively away from the coast; the far north receives no more than 20 inches a year.

Temperature and humidity remain relatively constant throughout the year in the south. In the north considerable seasonal variation occurs, and, during the dry season, the daily temperature range becomes great as well. On the coast the mean monthly maximum temperatures are steady throughout the year, remaining at a nearly constant 90° F (32° C) at Lagos and at about 91° F (33° C) at Port Harcourt; the mean monthly minimum temperatures remain approximately at 72° F (22° C) for Lagos and at 68° F (20° C) for Port Harcourt. In the northeastern city of Maiduguri, on the other hand, the mean monthly maximum temperature may exceed 100° F (38° C) during the hot months of April and May, while in the same season frosts may also occur at night. In general, mean maximum temperatures are higher in the north, while mean minimum temperatures are lower. The humidity is high, but it falls during the harmattan (the hot, dry, northeast trade wind), which blows for more than three months in the north but rarely for more than two weeks along the coast.

The harmattan season

Plant and animal life. Like rainfall patterns, the main vegetation patterns run in broad east–west belts, parallel to the equator. Mangrove and freshwater swamps occur along the coast and in the Niger Delta. A few miles inland, swamps give way to dense tropical rain forests. The oil palm, which is economically valuable, grows wild and is usually preserved when the forest is cleared for cultivation. In the more densely populated parts of the southeast, the original forest vegetation has been replaced by open palm bush. In the southwest large areas of forest have been replaced by cacao and rubber farms.

The savanna (tropical grassland) occupies the area north of the forest belt and is studded with baobab, tamarind, and locust bean trees. The savanna becomes more open in the far north and is characterized by scattered, stunted trees and short grass. Semidesert conditions appear in the Lake Chad region, where various species of acacia and the doum species of palm are common. Gallery forests (narrow forest zones along rivers) are also characteristic of the open savanna in the north.

In densely populated areas of the savanna, such as around Sokoto, Kano, and Katsina, the vegetation has been depleted as a result of continuous cropping, overgrazing, and bush burning. The near-total decimation of plant life has resulted in the gradual southward advance of the Sahara in the far northern districts.

Camels, antelope, hyenas, lions, and giraffes inhabit the savanna region, and the red river hog, the forest elephant, and the chimpanzee live in the rain forest belt. Animals found in both forest and savanna include leopards, golden cats, monkeys, gorillas, and wild pigs. Rodents such as the squirrel, porcupine, and cane rat constitute the largest family of mammals. The northern savanna abounds in Guinea fowl. Other common birds include quail, vultures, kites, bustards, and gray parrots. The rivers contain crocodiles, hippopotamuses, and a great variety of fish.

Settlement patterns. Marked differences exist between north and south not only in physical landscape, climate, and vegetation but also in the social organization, religion, literacy, and agricultural practices of the people. These differences form the basis of the division of Nigeria into three main regions—the south, or Guinea coastlands; the middle belt; and the north, or Nigerian Sudan.

Regional differences

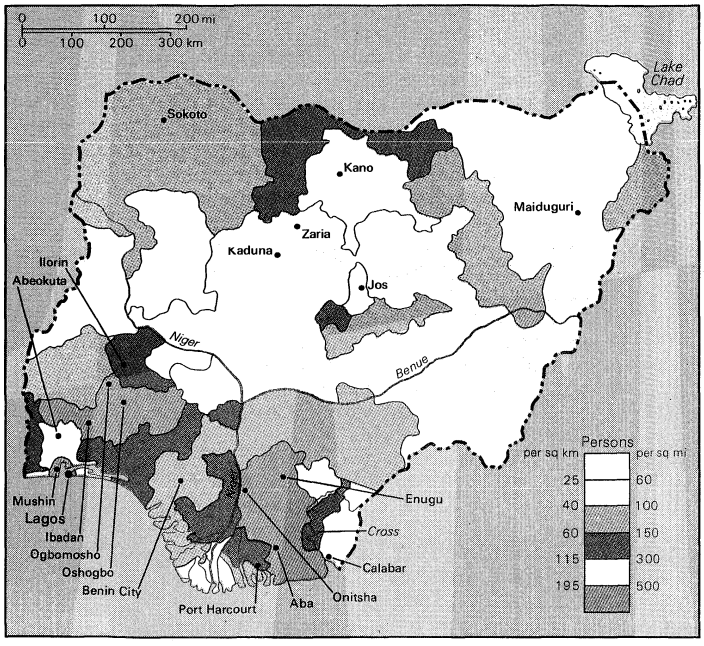
The south is the most economically developed part of Nigeria. Its forest resources are intensively exploited, and its tree crops are harvested on peasant farms and commercial plantations. All of the major industrial centres and oil fields, as well as the seaports, are concentrated in the region. The south has several cultural regions, the most important of which are the Yoruba in the west, the Benin in the central part, and the Igbo–Ibibio in the east.

The middle belt is the most sparsely settled and least developed part of Nigeria. It covers about two-fifths of the land area of the country but supports less than one-fifth of the total population. The inhabitants of the middle belt belong to more than 180 linguistic groups. Before 1970 large-scale development was restricted to a few government-supported projects, such as the Kainji Dam and the Bacita sugar project, and a few industries in the new towns of Jos and Kaduna. After the national administrative reorganization of 1975, the region gained importance because seven of the 19 state capitals as well as the 2,800-square-mile Federal Capital Territory are located there. In addition, during the early 1980s a giant iron and steel complex was built at Ajaokuta, near Lokoja.

Until the beginning of the 20th century, when a new economic pattern was created by the construction of a railroad to the coastal ports, the Nigerian Sudan maintained regular trans-Saharan contacts with the Mediterranean and the Middle East. Except in the Lake Chad Basin, where the Kanuri people established the state of Bornu, the Nigerian Sudan is dominated by a blend of Hausa–Fulani culture. It is a cattle zone inhabited by the nomadic cattle-owning Fulani and by the Hausa, who are settled cultivators. Islām is the predominant religion.

The main concentrations of people are in the forest belt west of the Cross River and in the western half of the extreme north. Parts of the Igbo and Ibibio areas in Nigeria are the most densely settled areas in Africa south of the Sahara. This concentration of agricultural people occurs in a region that has heavily leached and impoverished soils, and there is a food deficit. Many people leave the region to seek employment in other parts of the country. The second region of dense population in the forest belt occurs in the cacao-growing Yoruba areas, which attract many migrants from the congested Igbo and Ibibio areas.

Population distribution



Population density of Nigeria.

In the extreme north there are two regions of dense population—the Sokoto area and the Kano–Katsina area. The Kano concentration is based on intensive agriculture in an area of relatively fertile soils, but the densely settled areas around Sokoto and Katsina have impoverished soils and do not produce enough food for the local population.

Smaller pockets of dense population occur in the tin fields of the Jos Plateau, in the southern Tiv region, and in the Okene area. The remaining, and by far the greater, part of the country is sparsely settled; vast areas of the middle belt, the Lake Chad Basin, and the eastern Cross River area are virtually uninhabited.

Rural settlement

About 80 percent of the people live in villages. Closely nucleated settlements occur along the coast, in the Yoruba area in the southwest, and in the Hausa and Kanuri areas of the far north. In parts of the Igbo and the Anang–Ibibio regions in the southeast and of the Tiv area in the middle belt, settlements consist of dispersed homesteads called compounds. Each compound houses a man, his immediate family, and some relations. A number of compounds make up the village, usually inhabited by people claiming a common ancestor—often the founder of the village.

House types change as one moves inland. In the coastlands the walls consist of bamboos tied together with ropes, and roofs are made of bamboo leaf mats. Bamboos, ropes, and mats are made from the raffia palm, which abounds in the region. Rectangular mud houses with mat roofs are found in the forest belt; the houses of the more prosperous have corrugated iron roofs. In the savanna areas of the middle belt and in parts of the north, houses are round mud buildings roofed with sloping grass thatch; flat mud roofs are used in the drier areas of the extreme north.

Each village has a chief, or headman, who usually is one of the oldest men in the community and rules by consent of the people. This is particularly true in the eastern states. In Yoruba areas, and in most parts of the northern states, the chief is chosen by, or with the consent of, the region's traditional ruler. The chiefs are usually powerful and highly respected. A characteristic feature of village life is the age-grade system of social organization, which groups together people of the same age group.

Urban settlement

Only the Yoruba, Hausa, Bini, Kanuri, and coastal peoples were town dwellers before the 20th century. The Yoruba, of whom about half live in towns of more than 5,000, are the most urbanized people in tropical Africa. Their towns, most of which are several hundred years old, were originally administrative and trading centres; most have retained these functions. The principal Yoruba towns are Ibadan, Ogbomosho, Abeokuta, Ife, and Oyo.

The northern towns of the Nigerian Sudan, including Kano, Zaria, Sokoto, and Katsina, are much older than the Yoruba towns. Owing their growth to the trans-Saharan trade, as well as to the agricultural wealth of the Sudan, these ancient towns were unplanned, and they consist of an amorphous assemblage of mud buildings.

Coastal fishing and salt-trading villages grew into towns in response to the slave trade and the later trade in palm oil between the coastal peoples and Europeans. These towns include Bonny, Opobo Town, Okrika, Buguma, Brass, Forcados, Creek Town, and Calabar (Duke Town). At the beginning of colonial rule, these port towns had a more cosmopolitan population than the towns of the Yoruba area and the far north, but they were much smaller.

Ancient and modern towns

During the period of British rule, new towns grew up and the older ones grew larger, except along the coast, where all the older towns, excluding Lagos, declined in size or even ceased to exist. Many towns originally were primarily administrative centres, but—like Port Harcourt, Lagos, and Ibadan, as well as the Sudanese towns of Kano and Kaduna—they have become industrialized.

Lagos is the largest conurbation in Nigeria. Primarily a Yoruba town, it was founded in the late 17th century as a small fishing and trading settlement. Lagos retains the characteristics of a preindustrial city, even though it is by far the most industrialized city in the country. The creation of several states since 1967 diverted some of the industries and job-seeking migrants from Lagos to the new state capitals, especially the older and larger ones such as Ibadan, Kaduna, Kano, and Enugu. Some small towns, notably Minna, Uyo, Makurdi, Maiduguri, and Bauchi, experienced remarkable population and economic growth after becoming state capitals.

The people. Virtually all the native races of Africa are represented in Nigeria, hence the great diversity of peoples and cultures. It was in Nigeria that the Bantu and Semi-Bantu, migrating from southern and central Africa, intermingled with the Sudanese Negro. Later smaller groups such as the Shuwa Arabs, the Tuareg, and the Fulani, who are concentrated in the far north, entered northern Nigeria in waves of migration across the Sahara.

The earliest occupants of Nigeria settled in the forest belt and in the Niger Delta, into which they retreated from the invaders of the northern savanna zone. During the period of the slave trade the social organization of coastal peoples, including those of the Niger Delta, was altered radically, in part because of the forced migrations of people from the interior into the area and in part because of contact with European traders. The early colonial period witnessed more cultural intermingling in the coastal towns, notably Calabar, Warri, and Abonnema, where Syrian, Lebanese, and European traders settled.

Ethnic composition. There are estimated to be more than 250 ethnic groups. Each group occupies a territory that it considers to be its own by right of first occupancy and inheritance. Nonmembers of a given group who have lived and worked for several decades in the territory of

Robert Frerck—Odyssey Productions

Court musicians at the Emir's Palace, Kano, Nigeria.

the group are still considered to be aliens. In most rural areas, such aliens may not acquire outright title to land; yet there has been considerable migration of people from one ethnic territory to another in search of farmland.

The Hausa

There are three major ethnic groups in the country, of which the Hausa are the most numerous. The Hausa who live in the far north have become integrated with the minority Fulani who conquered Hausaland in the early 19th century, and the great majority of both groups are Muslims. Although the town Fulani intermarry freely with the Hausa and other groups, they continue to control the administration of the Hausa towns. The cattle Fulani generally do not intermarry. They speak the Fulani language—Fulfulde—rather than Hausa.

Another large and politically dominant group consists of the Yoruba-speaking peoples of southwestern Nigeria, who, like the Hausa and the Fulani, have ancient connections with the Middle East. The Yoruba, although they are farmers, often live in large preindustrial cities. Each Yoruba subgroup is ruled by a paramount chief, or oba, who is usually supported by a council of chiefs. The *oni* of Ife, who is the spiritual leader of the Yoruba, and the *alafin* of Oyo, who is their traditional political leader, are the most powerful rulers; their influence is acknowledged throughout the Yoruba areas.

The Igbo-speaking peoples are one of the largest ethnic groups in Nigeria. In the southeastern region, where the core area of Igbo territory is located, the great majority of people live in small dispersed settlements. A much smaller proportion (about 10 percent) of the Igbo, whose territory is in Bendel state, live in large preindustrial towns and are culturally much closer to the Edo of neighbouring Benin City than to the Igbo east of the lower Niger valley. Traditional Igbo society has been democratic; and, outside the western Igbo districts of Bendel state, the largest political unit has been the village group, ruled by a council of elders rather than by a chief.

Less numerous groups include the Ibibio, who live near the Igbo, with whom they share many traits, and the Edo people of Benin City, whose culture has been influenced by their Yoruba neighbours. In the middle belt, where the greatest concentration of ethnic groups (more than 180) occurs, the Tiv and the Nupe are the largest groups. Both are settled cultivators, but while Nupe society is hierarchical, that of the Tiv tends to be decentralized.

Linguistic composition. Each ethnic group has its own language, of which there may be several dialects. The major languages of Nigeria are Hausa, Yoruba, Igbo, Ibibio, Kanuri, Tiv, and Ijo, in that order. Hausa is by far the most widely spoken language, partly because between 1951 and 1967 Hausa was an official language of the northern states. The continued political domination of the country by the Hausa/Fulani has also contributed to the spread of the Hausa language. The creation of more states has resulted in the proliferation of written languages, since radio stations relay news in all the major languages of each state, some of which are also used in local newspapers and magazines. Since about 1950 the translation of the Bible into a number of Nigerian languages has also led to an increase in the number of written languages.

Religions. The three main religions are traditional beliefs, Islām, and Christianity. Adherents of traditional religion worship the supreme God (Olorun Olodumare in Yoruba, Chukwu in Igbo, Allah in Hausa, and Abasi Ibom in Ibibio) through a number of intermediaries or lesser gods. At the beginning of the 20th century, most Nigerians were followers of traditional religion, but during the 1963 census, 47 percent of the people were recorded as Muslims and 35 percent as Christians. However, many professing Muslims and Christians now openly perform certain rites or rituals of traditional religion that are no longer condemned as they were during the early colonial period.

Traditional religions

Religious freedom is guaranteed by the constitution, and Muslims and Christians live and work together. However, the country has since the early 1970s witnessed a number of religious riots sparked by Muslim fundamentalists. The greatest concentration of Muslims is in the northern states, where three-quarters of the people profess the religion. Islām has also gained strength in the southern states of Lagos, Ogun, and Oyo, where Muslims outnumbered Christians by the early 1980s. Christians make up more than three-quarters of the population in the eastern states, and Christianity has made great inroads since 1960 in the middle belt, which is now the main area of religious conflict.

The main established Christian groups are Roman Catholic, Methodist, Anglican, and Baptist. There is an increasing number of separatist Christian groups, of which the largest are the Cherubim and Seraphim, the Celestial Church of Christ, and the Brotherhood of the Cross and Star. The popularity of these spiritualist sects established by indigenous peoples is a threat to the older established churches, which have introduced drumming and dancing—formerly the distinguishing features of the separatist churches—into their services to avoid losing more members. The rapid growth in the membership of these sects, which, like Islām, allow a member to marry more than one wife, has also posed a threat to Islām and is thought to be partly responsible for the recent religious conflict.

Demographic trends. Nigeria, like other developing nations, has high birth and mortality rates, although since about 1950 there has been a considerable decline in the rate of infant mortality and an increase in life expectancy. There has consequently been a rapid growth in the population.

Internal migration

A considerable movement of population takes place within Nigeria. There has been a south-to-north movement of migrants who settled in the northern cities of Kano, Sokoto, Kaduna, Jos, and others, as well as a north-to-south movement of seasonal migrants from the Sokoto and Kano areas into the cacao-growing Yoruba areas. A larger number of people migrate from the Igbo and Ibibio areas and from the Niger Delta to the more industrialized and urbanized western states of Lagos, Oyo, and Ogun and to the western states of Ondo and Bendel, which are rich in agricultural resources. Many of these migrants work as labourers in the cacao- and rubber-producing districts or as tenant farmers.

Before the end of the civil war in 1970, many Nigerians migrated to work in Benin, Ghana, Equatorial Guinea, Cameroon, and Sierra Leone. The expulsion of aliens (mostly Nigerians) from Cameroon in 1967 and from Ghana in 1969, as well as the relative buoyancy of the country's economy, attracted foreign western Africans into Nigeria from about 1972. This trend was encouraged in 1978 by the establishment of the Economic Community of West African States, under which the free movement of citizens of member states is guaranteed. In the early 1980s, the downturn in the Nigerian economy and religious riots in which foreigners were allegedly involved prompted the government to reverse its immigration policy. In early 1983, about two million aliens were expelled from Nigeria and in 1985 another 700,000 were expelled. A small but increasing number of Nigerians emigrate to Great Britain, West Germany, Canada, and the United States.

The economy. The Nigerian economy still exhibits colonial traits. One such trait is the coexistence of a modern market economy and a dwindling traditional economy that features the use of group and age-grade labour and of payments in kind. Government, federal and state, is the major employer.

Since independence the government has implemented four comprehensive development plans. Despite these, the great majority of the population is still engaged in farming; yet there exists a serious food deficit resulting in an increasing food import bill. A series of world oil price increases from 1973 produced rapid economic growth in transportation, construction, manufacturing, and government services; this led to a great influx of rural people into the larger urban centres. Agricultural production stagnated to the extent that Nigeria, once a major exporter of palm oil, peanuts, and cotton, became a net importer from about 1975. The rapid increase in population has also contributed to the continued low income per capita in spite of sizable oil revenues.

The indigenization decrees of 1972 and 1977, which stipulated an equity participation by Nigerians of 40 percent and 60 percent, respectively, in most economic ventures,

led to a considerable increase in the profits that accrue to Nigerians, especially from the manufacturing sector. Outright nationalization of industries is not contemplated in view of the dependence of some key industries on foreign equipment and personnel and the need to create an atmosphere attractive to foreign investment.

Resources. Nigeria has a variety of both renewable and nonrenewable resources, some of which have not yet been effectively tapped. Solar energy is probably the most extensive of the underutilized renewable resources, but it is likely to remain untapped as the vast reserves of natural gas produced along with crude oil have yet to be fully harnessed. Water power from the many rivers, another renewable resource, is not fully developed but provides much of the power used in industry and in homes. The main sources of hydroelectric power are the Kainji Dam, the Shiroro Dam, and the Jebba Dam power stations. Other smaller plants are in operation, and large hydropower stations are planned for Makurdi, Zungeru, and Mambilla.

The land, like the forests and fisheries of the country, is renewable if properly managed but can be rendered nonrenewable through careless exploitation. Nigeria has no shortage of arable land. There is an extreme shortage of farmland in the very densely settled districts of the southeastern states and around Kano, Katsina, and Sokoto, however, resulting in the large-scale migration of land-hungry Igbo, Ibibio, and Hausa people to other parts of the country. Access to farmland is restricted in some localities that appear to have abundant cultivable land, and, in the far north, the desiccation of the Sahel has severely limited the land area available for cultivation.

Minerals

The most important minerals are crude oil, natural gas, coal, tin, and columbite. Oil, first discovered in 1956, has since 1970 been the most important source of government revenue and foreign exchange. Most of the oil output comes from onshore fields in the Niger Delta, although an increasing proportion of the crude produced since 1975 comes from offshore locations. Large reserves of natural gas exist, but most of the gas produced as a by-product of crude oil has had to be flared for lack of a market. Gas is used to generate electricity at Sapele and Afam, and, since 1984, oil companies have been required to reinject into the ground some of the natural gas produced in the course of pumping crude oil. The earliest mineral extraction (since the early colonial period) was of tin and coal. Tin and columbite, an iron-bearing mineral that accompanies it, occur in the Jos Plateau, and coal is found in the Udi district of Anambra state and in parts of Kwara and Benue states. The coal reserves consist of low-quality bituminous coal, a third of which is located in Anambra state. Deposits of lignite, or brown coal, occur in Bendel, Anambra, and Imo states. There are iron-ore deposits in the Lokoja area, which is close to the Ajaokuta steel complex in the lower Niger Valley. Limestone occurs in many areas and is widely exploited for manufacturing cement and for use in the steel industry.

Agriculture, forestry, and fishing. About two-thirds of all Nigerians obtain a living from agricultural production. Most are small-scale farmers who produce food crops for home consumption, with a little surplus for sale, and who derive additional income from one or more cash crops and from the sale of local crafts. Farms are small, usually less than two and one-half acres (one hectare) in the south and about seven and one-half acres in the open grassland areas of the north. The continued dominance of the primitive hoe and matchet (machete) as farm implements and the shortage of farmland in some localities and limited access to land in others are among the factors that restrict the size of farmland cultivated per family. Overcropping, soil impoverishment, repeated damage by drought, and lack of investment capital have contributed to low productivity and general stagnation in agriculture. In the face of rapid population growth and an accelerating rate of urbanization, the food deficit continues to worsen.

Root crops—notably yams, taro, and cassava—are the main food crops in the south, while grains and legumes—such as sorghum, millet, cowpeas, and corn (maize)—are the staple crops of the drier north. Tree crops—notably oil palm, cocoa, and rubber—are the main industrial crops of the south, while peanuts (groundnuts) and cotton are produced in the north. As with staple food crops, the production of industrial crops is dominated by small-scale farmers. In the late 1980s only cocoa was exported, since the local demand for palm oil, peanuts, cotton, and rubber exceeds production.

The first major step taken by the government to halt the decline in industrial crop production was to disband in 1982 the produce marketing boards, which paid prices set by the government. Many farmers have since been motivated to cultivate tree crops, and the federal and state governments have established plantations of oil palm, rubber, and cocoa. Programs to alleviate the food shortage have featured the direct purchase and distribution of foodstuffs by government agencies and the production by government parastatals of various staples on large commercial farms. The Operation Feed the Nation program of 1976–80 sought to increase local food production and thereby reduce imports. Citizens were encouraged to cultivate any empty plot of land; urban dwellers were encouraged to garden undeveloped building plots. In 1980 this program was replaced by the more elaborate National Council on the Green Revolution, which has had little impact.

Livestock production has remained largely undeveloped during the mid-20th century. The nomadic cattle Fulani are still the main producers of beef cattle, although some of the cattle under the care of these nomads belong to settled farmers and city dwellers.

Forests

Nigeria's permanent forest reserves consist of savanna woodland (76 percent), tropical rain forest (20 percent), and swamp forest (4 percent). Outside the reserves, much of the forest cover has been destroyed through regular burning when preparing farmland or during hunting. Forest destruction is most complete in the very densely settled areas and in the drier savanna, where overgrazing, bush fires, and the great demand for fuelwood has prevented normal regeneration of plants on fallow land. Fuelwood (firewood and charcoal) is the main source of domestic fuel even in urban areas and accounts for about 95 percent of domestic consumption of forest products.

There are many large plantations of exotic species, such as gmelina and teak, established by the government to provide electric and telegraph poles and fuelwood. In the arid zone of Sokoto, Kano, and Borno states, forest belts have been established to help arrest the southward advance of the Sahara. Forest plantations have been established in many watersheds to protect water catchment areas of rivers and to reduce the incidence of soil erosion.

Fishing is the fastest-growing agricultural subsector and has assumed greater importance as a source of protein following the loss of thousands of livestock animals as a result of recurring drought in the Sahel since the early 1970s. The domestic catch supplies about 60 percent of the fish demand. Lake Chad and the coastal waters are the main sources of fish, but large quantities are caught every year in the pools of the seasonal rivers of the northern states; and there are opportunities to develop large scale fish farms in the reservoirs at Kainji Dam and a few smaller dams associated with multipurpose river-basin development projects.

Industry. Mining is the most important and the fastest-growing sector of the economy. Its share of the gross domestic product rose from about 1 percent in the 1950s to about 25 percent in 1977 because of the rapid increase in crude oil output. From an output of 9,000 tons in 1958, oil production increased to more than 19 million tons in 1966 and in 1979 the record level of 114 million tons was achieved. Two years after Nigeria joined the Organization of Petroleum Exporting Countries (OPEC) in 1971, a huge price increase occurred, more than doubling Nigeria's oil revenue from 1973 to 1974. Thereafter oil prices continued to rise, reaching a peak in 1980 before the price crash of 1981. There are oil refineries at Port Harcourt, Warri, and Kaduna.

Tin mining in the Jos Plateau started in 1905. Since 1962, when a smelting plant was built to produce tin ingots, tin has not been exported in its raw form. The production of tin ore and columbite has always been controlled by private companies. Production is by open-pit mining.

Coal mining began in 1915 and has always been under the direct control of the central government. In the oldest mines, located near Enugu, the coal is obtained from adits driven into the scarp slope of the Enugu cuesta. Production was restricted to Anambra state until 1968, when the mines near Enugu were closed because of the civil war; production then began at Okaba in Benue state.

Traditional industries carried out in homes or in makeshift workshops include the making of iron implements such as hoes and matchets, door hinges, bolts, and dane guns (firearms of obsolete design, originally of European manufacture). Traditional soap- and salt-making workshops appeared in large numbers after the near collapse of the Nigerian economy in 1983, when most wage earners were unable to pay for factory-made soap and imported table salt. Pottery making and wood carving are widespread, as are canework and the making of bags and mats from raffia.

The manufacturing sector is the second fastest-growing sector after mining. Revenue from mining has enabled the central government to embark upon such capital-intensive industries as the Ajaokuta and Aladja steel mills, pulp and paper mills at Oku Iboku and Iwopin, and petrochemical plants at Kaduna and Port Harcourt. Before the large increase in government revenue, large-scale manufacturing—dominated by the production of textiles, tobacco, beer, soft drinks, and cement—was controlled by foreign investors. The indigenization decrees have altered the ownership situation, although the management and effective control of most large factories are still in the hands of expatriate representatives of multinational corporations. The greatest weakness of this sector has been its dependence on foreign raw materials, a situation that changed in 1987 when the prohibition on importation of a wide range of raw materials forced local breweries to substitute sorghum and rice for barley, and flour mills to process corn instead of wheat for bread. The largest concentration of large factories is in the Greater Lagos area. Each state capital has a number of large manufacturing industries, but a few major industries, such as the paper mills and the steel mills, are located in remote areas where new towns have grown up to serve the factories.

Finance. The naira is the basic unit of Nigerian currency. The main sources of government finance consist of direct taxes and revenue from a petroleum profits tax and from mining royalties. Unexpected and drastic fluctuations in oil prices since 1981 therefore created considerable difficulty in the implementation of the 1981–85 National Development Plan and in meeting recurrent government expenditures. Recurrent federal expenditures consist mainly of transfers to state governments, defense, general administration, health care, and education.

The increase in government revenues following the rise in oil prices in the early 1970s led to unprecedented expansion in education, in prestige projects, and in the misappropriation of public funds, resulting in a heavy public external debt and federally owed domestic debt. The servicing of the external debt since 1984, at a time when oil prices have remained low, has been a major strain on the management of the economy. In 1985 and 1986, for example, the country spent as much as 40 percent of annual export earnings to service foreign debts, leaving little for economic development.

The Central Bank, established in 1959, has branches in all the state capitals and provides guidelines to all commercial and merchant banks in the country. In 1976 all foreign banks were compelled to sell 60 percent of their shares to Nigerians.

Trade. The direction of domestic trade in staple foods is largely north and south between different ecological zones but also to the major urban centres. The southern states supply plantains, cassava, and fruit to the northern states, which in turn supply beans, onions, and vegetables to the southern states. Yams and potatoes from the middle belt are traded in the southern and the far northern cities. Livestock traded in the south comes mostly from the north, which relies on the forest belt in the south for kola nuts. There is a sizable trade in foodstuffs from the eastern states and Bendel to the urbanized Yoruba cultural area of southwestern Nigeria. Women play a dominant role in the marketing of foodstuffs and manufactured goods. There are few department stores, and most of the food items and manufactured goods are sold in open market stalls and small neighbourhood shops.

There is very little trade between Nigeria and other African countries. The main markets for Nigerian exports—consisting mostly of crude oil, cocoa, palm kernel, timber, and tin ingots—are the United States, Great Britain, Japan, France, and West Germany. Trade with the Soviet Union increased during the civil war and has continued because of Soviet involvement in the construction of the Ajaokuta steel mill. The main imports are consumer goods, nondurable goods, industrial raw materials, industrial and construction machinery, and passenger cars, most of which come from Japan and western Europe.

Before 1954, when consumer goods dominated the import list, Nigeria had a favourable balance of trade. The growth of imports in machinery and equipment following the granting of internal autonomy by Britain, and low commodity export prices, combined to create a trade deficit from 1955 to 1965. Thereafter Nigeria had a trade surplus based on revenue from crude oil exports. The structure of imports altered during the early 1970s, with capital goods and industrial raw materials accounting for about 70 percent of all imports by value. Reckless spending and fraud brought about a balance of payment deficit beginning in 1976. Improved earnings from oil and restrictions on imports the year before the return to civilian rule in 1979 reversed the unfavourable trend and resulted in a trade surplus in 1980. The resumption of projects in the new federal capital designate of Abuja contributed to a high import–export imbalance in 1982. By the end of 1983, when the military again intervened in Nigerian politics, the country had become virtually bankrupt, to the point where importers found it difficult to obtain letters of credit from foreign banks.

Transportation. The general pattern of transport is from north to south, running from the interior to the southern seaports. This pattern dates back to the colonial period, when raw materials were produced in the interior and shipped to Great Britain and other western European countries, which in turn supplied manufactured goods. The transport pattern also resulted from the fact that different agricultural products are grown in zones, or belts, running from east to west, so that the general direction of internal exchange trade is from north to south.

The north–south traffic pattern

Nigerian roads fall into three categories: trunk A roads, which are maintained by the federal government and which link Lagos with all the state capitals; trunk B roads, which are maintained by state governments and which connect provincial capitals and other large towns with the trunk A system; and other roads, which are maintained by local governments, carry local traffic, and act as feeders to the trunk-road systems. All trunk A and most trunk B roads are surfaced; almost all other roads are earth. The conversion of the Benin–Shagamu and the Port Harcourt–Enugu trunk A roads into four-lane divided highways was completed in 1981. The only expressway in the country, opened to traffic in 1978, runs between Lagos and Ibadan.

The traffic flow on the roads is greatest in the cacao belt of southwestern Nigeria, the peanut and cotton belt of the Kano–Katsina region, the Jos Plateau tin fields, and the palm belt of southeastern Nigeria. These areas are served by a dense network of all-weather roads. The relatively unproductive and sparsely settled areas of the middle belt, of the Cross River region, and of the Chad Basin have tenuous road links that carry only a few trucks a day.

Roads are now the most important means of transportation, but formerly the railroad was dominant. On several occasions Nigeria's railroads have proved incapable of transporting large cargoes of peanuts and cotton from the north. The slow speed of trains (which is attributed largely to the narrow gauge of the railroads), along with the unhygienic conditions of the coaches and poor quality service, has resulted in lack of patronage of the passenger services. By the 1980s almost all cargo and passenger transport was by road and air, and movement by rail had declined sharply. The railroad system has two single-track

The railroad system

main lines: the eastern line from Port Harcourt to Maiduguri and the western line from Lagos to Kano. Branch lines connect the western main line to Kaura Namoda, to Nguru, and to Baro on the Niger. Since 1960 tracks have been relaid with heavier rail to permit heavier axle loads and higher speeds, improvements in signaling have speeded rail movements, and steam engines have been replaced by diesel locomotives.

In spite of the downturn in the Nigerian economy in 1983 and the devaluation of the naira by more than 80 percent in 1986, the government in 1989 resuscitated the plan to extend existing rail lines. There is also an urgent need to convert the existing single-track to double-track railroads so as to permit a considerable increase in rail haulage. This would reduce the volume of heavy trailer traffic, which has caused extensive damage to all roads.

The creeks and rivers were the first means of communication in 20th-century Nigeria. The most important waterways are the Niger and Benue rivers, which still carry substantial quantities of goods. The Cross River is used to ship exports to Calabar, but, like other rivers in Nigeria, it is not navigable during the dry season. Passenger and cargo boats operate on the lagoons and on the many creeks along the Nigerian coast from Lagos to the Cross River.

The ports

Lagos and Port Harcourt, administered by the Nigerian Ports Authority since its establishment in 1954, are the main international seaports. Chronic congestion at these two ports was largely responsible for the authority's takeover in 1970 of the port installations and administration of the smaller ports of Warri, Sapele, Koko, and Calabar. The expansion of the Lagos port complex in 1978 and the modernization and expansion of facilities in the smaller ports resulted in excess capacity by 1980.

Almost all the state capitals are served by air transport supplied by Nigerian Airways. There are smaller airfields in some provincial cities and in the oil-producing areas of the Niger Delta and the Cross River estuary. Lagos and Kano handle most of the transcontinental traffic. Other international airports include those at Port Harcourt, Calabar, Maiduguri, and Sokoto.

Administration and social conditions. *Government.* Under the constitution of 1979 Nigeria adopted a system of government in which the president, who is directly elected, exercises power as the chief executive and head of state. Also provided is a National Assembly, which consists of the Senate and the House of Representatives. After a military coup on Dec. 31, 1983, parts of the constitution were modified or suspended, and the National Assembly was replaced by a Supreme Military Council. This was replaced by the Armed Forces Ruling Council in 1985. The military administration has promised to return Nigeria to civilian rule in 1992.

There are three tiers of government: the central, or federal, government, the state governments, and the local governments. Until 1987 the third tier was merely an appendage of the state governments, which usurped the functions of local government. This situation altered in 1988 when the federal government decided to pay federal grants directly to each local government and no longer through the various state governments. This put local governments for the first time in a position to function effectively.

Judicial system

The Nigerian legal and judicial system is subject to regional variations, particularly between the Muslim north and the Christian south. There are three codes of law: customary law, Nigerian statute law, and English law. Customary laws are administered by native, or customary, courts, usually presided over by persons with no formal legal education. In the case of the Alkali, or Muslim, Courts, however, the judges (*alkalis*) undergo formal training in Muslim law. Customary courts are of different grades, with Grade A courts serving as courts of appeal for cases from Grade B, C, and D courts. In addition, under the 1979 constitution, any state could establish a special Customary Court of Appeal or, in the case of Muslim communities, a Sharī'ah Court of Appeal.

Nigerian statute law includes much of the British colonial legislation, most of which has been revised. State legislatures may pass laws on matters not included in the Exclusive Legislative List, which includes such subjects as defense, external affairs, and mining—all of which can be legislated upon only by the federal government. Federal law prevails whenever federal legislation conflicts with state legislation. In addition to Nigerian statutes, English law is used in the magistrates' and all higher courts.

Each state has a High Court, which is presided over by a chief judge. The Federal Revenue Court was established in 1973 to handle matters relating to taxes, corporations, banking, and copyrights; it has a status comparable to that of a state High Court. Appeals from state High Courts and the Federal Revenue Court go to the Federal Court of Appeal, which was created in 1976. The Supreme Court, under the chairmanship of the chief justice of Nigeria, is the highest court.

The Nigeria Police Force, established by the federal constitution, is headed by the inspector general of police, who is appointed by the president. The general inefficiency of the force is attributable in part to the low level of education and the low morale of police recruits, who are poorly housed and very poorly paid, and to the lack of modern equipment. Corruption is widespread.

Education. Both the federal and state governments are responsible for education at all levels. Local governments may, however, establish and administer primary and secondary schools. Most secondary schools, trade centres, technical institutes, teacher-training colleges, and colleges of education and of technology are controlled by the state governments, and there are a number of state universities. There is at least one federal college of technology and a college of education in each state. The involvement of the federal government in secondary education is, however, largely restricted to the so-called unity schools, admission to which is based on a quota to ensure representation from all states.

The growth of free education

Before 1976, when the Universal Primary Education (UPE) scheme began, primary education was free only in Lagos, Oyo, Ogun, Ondo, Bendel, Sokoto, and Niger states. The estimated primary-school population in 1976 of 8.3 million grew to more than 11 million by 1981, after the UPE scheme had become compulsory. Secondary education has always been free to indigenes of most of the northern states, but in practice only a small proportion of the children receive secondary education—largely the children (usually the sons) of the traditional feudal ruling class and the educated elite. In the southern states free education at the secondary level began in 1979 in Lagos, Ogun, Oyo, Ondo, and Bendel states.

Until 1950 most schools were controlled by Christian missionary bodies. There were a few government schools, although government policy was to give grants to mission schools rather than to expand its own. After independence many secondary schools were established by local governments, Muslim missionary groups, and prominent citizens. A shortage of teachers after the civil war of 1967–70 led to the recruitment of hundreds of Indian and Pakistani teachers, especially in the educationally disadvantaged Muslim areas in the north. Most of these Asian teachers left Nigeria between 1986 and 1988, after the devaluation of Nigerian currency and the refusal of some state governments to renew their contracts in view of mounting unemployment of local university graduates.

In 1966 there were five universities, two of which (at Lagos and Ibadan) were operated by the federal government. The federal military administration took control of the state universities at Ife, Benin, Nsukka, and Zaria in 1975, and, when civilian rule was restored in 1979, there were 13 federal universities. Like the older universities, each of the universities established in 1975 at Kano, Sokoto, Maiduguri, Jos, Ilorin, Calabar, and Port Harcourt has a medical school. The Federal Universities of Technology at Bauchi, Makurdi, Owerri, Akure, Minna, and Yola were the first to be established after the return to civilian rule. In 1987 Makurdi and Abeokuta were constituted into agricultural universities, and a federal university was established at Abuja. A number of state governments also established new universities after 1980; these include the Rivers State University of Science and Technology at Port Harcourt and the state universities of Imo and Cross River. Attempts by individuals and private organizations,

including the Christian churches, to establish universities have been frustrated by the federal Ministry of Education. The university student population has continued to increase. A shortage of qualified teaching staff as well as inadequate financing and the slow pace of physical development of the campuses are major problems.

Health and welfare. The concentration of people in the cities has created enormous sanitary problems, particularly sewage disposal, water shortages, and poor drainage. Large heaps of domestic waste spill across narrow streets, causing traffic delays, while the dumping of garbage along streambeds constitutes a major health hazard and has contributed to the floods that often plague Ibadan, Lagos, and other cities during the rainy season. Many people die from malaria, waterborne diseases, cerebrospinal meningitis, and other preventable diseases.

Health conditions are particularly poor in the shantytown suburbs of Greater Lagos and other large cities, where domestic water supplies are obtained from wells that are often polluted by seepages from pit latrines. Rural communities also suffer from inadequate or impure water supplies. Some villagers have to walk as far as six miles to the nearest water point—usually a stream. Because people wash clothes, bathe, and fish (sometimes using fish poison) in the same streams, the water drawn by people in villages farther downstream is often polluted. During the rainy season, wayside pits containing rainwater, which are often dug very close to residential areas, are the main source of domestic water supply. Cattle are often watered in the shallower pools, contributing to the high incidence of intestinal diseases and guinea worm in many rural areas.

Medical and health services are the responsibility of the state governments, which maintain hospitals in the large cities and towns. There are specialized hospitals in most of the state capitals, each of which also has a university teaching hospital financed by the federal Ministry of Health. There are many private hospitals, clinics, and maternity centres. Medical services are inadequate, even in the five western states where a free health service scheme was introduced in 1979. Many hospitals do not have enough medical personnel, and drugs are scarce. Rural areas are the most deprived of services.

There is no health insurance scheme or social welfare system for all Nigerians. Most commercial firms and factories provide free medical services for their employees and, in some cases, for their immediate families as well. Civil servants are entitled to free medical care in government-financed hospitals. Most elderly Nigerians and the unemployed depend on the extended family, which serves as the traditional social welfare system.

Housing

Overcrowding in the cities has led to the spread of slums and the emergence of shantytown suburbs in most of the larger urban centres. Since there are only a few housing corporations, most houses are built by individuals, who, unable to obtain loans from either commercial or government mortgage banks, must rely on their savings for constructing houses. House rents have steadily increased in the major urban centres as people have migrated into the cities in search of jobs. A federal housing program, begun in 1980, provides for the construction of low-cost housing for low- and middle-income workers in the state capitals, local government headquarters, and other large towns.

Cultural life. Nigeria has a rich and varied cultural heritage, deriving from the different ethnic elements, as well as Arabic and western European cultural influences. Secret societies, such as Ekpo and Ekpe among the peoples of the southeast, were formerly used as instruments of government, while other institutions were associated with matrimony. According to the Fulani custom of *sharo* (test of young manhood), rival suitors underwent the ordeal of caning as a means of eliminating the less persistent grooms-to-be, while in Ibibio territory, girls were confined for several years in bride-fattening rooms before they were handed over to their husbands. These and other customs were discouraged by colonial administrators and missionaries. Some of the more adaptable cultural institutions have been revived since independence; these include, for example, the Ekpo and Ekong societies for young boys in parts of the southeast.

The Institutes of African Studies at the Universities of Ibadan and Ife have done much to reawaken interest in traditional folk dancing and poetry, as have the School of Fine Arts and the School of Drama at Zaria and Ibadan. With the establishment of radio and television stations in all state capitals, programs featuring traditional music and dance, folk operas, and storytelling are available in some 25 languages. Because writing became common only after 1900 (except in the Muslim north), and because few educated Nigerians showed any interest in folk traditions and local culture until the late 1960s, much of the country's culture is believed to have perished. Many ancient folk songs have been revived by popular singers who use modern musical instruments to produce sounds that villagers can hardly identify with the songs they inherited from their ancestors.

With a few exceptions, such as at Oron and Umuahia, national museums are restricted to large cities, usually state capitals. In Jos, the national museum and the zoo are a major tourist attraction. Movie theatres, showing mostly Indian and American films, are popular among the urban middle- and low-income groups.

For statistical data on the land and people of Nigeria, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (R.K.U.)

HISTORY

The peoples of Nigeria may be classified broadly into three linguistic groups: (1) the Niger–Congo; (2) the Saharan, principally Kanuri; and (3) the Chadic branch of Afro-Asiatic languages, notably Hausa, Margi, and related languages. The huge Niger-Congo group may be further subdivided into the Kwa subgroup, which is the largest and includes such languages as Yoruba, Edo, Ijo, Igbo, Igala, Idoma, Nupe, and Gwari; the West Atlantic subgroup, notably Fulani (Fulfulde); the Benue-Congo subgroup, previously called Semi-Bantu, which includes Tiv, Jukun, and several languages of the Cross River basin such as Efik, Ibibio, Anang, Ekoi; and Adamawa. While some peoples (such as the Fulani and the Tiv) are relatively recent immigrants, it is estimated on the basis of glottochronology that the bulk of Nigerian languages, specifically the Kwa subgroup, must have been spoken in roughly the same locations as they are now for some 4,000 years.

Early Nigerian cultures. *The Nok culture.* The oldest skeleton found by archaeologists in the Nigerian area, at Iwo Eleru, near Akure, in the southwest, has been dated to about 9000 BC. There are isolated collections of ancient tools and artifacts of different periods of the Stone Age, but the oldest recognizable evidence of an organized society belongs to the Nok culture (*c.* 500 BC–*c.* AD 200).

A series of accidental finds of fine terra-cotta figurines by tin miners on the Jos Plateau has revealed an ancient culture named after the village of Nok, where some of the most representative heads were found. It was a Neolithic culture that made the transition to the Iron Age. The people raised crops and cattle, and they seem to have paid particular attention to personal adornment, especially the hair. Distinctive features of Nok art include naturalism; stylized treatment of the mouth and eyes; relative proportions of the human head, body, and feet; distortions of the human facial features; and the treatment of animal forms. The spread of Nok-type figures in a wide area south of the Jos Plateau, covering southern Kaduna state southeastward to Katsina Ala, south of the Benue River, suggests a well-established culture that left traces still identifiable in the lives of the Numan and other peoples of the area today. Many of the distinctive features of Nok art can also be traced in later developments of Nigerian art at Igbo Ukwu, Ife, Esie, Benin City, and other places.

Features of Nok art

Igbo Ukwu. The Igbo Ukwu bronzes, which have been dated to about AD 900, reveal not only a high artistic tradition but also a well-structured society with wide-ranging economic relationships. Of particular interest is the source of the copper and lead, which may have been Tadmekka in the Sahara, for making the bronzes and of the coloured glass beads, which may have come from Venice. It is believed that the bronzes were part of the furniture of the burial chamber of a high personage, a priest-king, proba-

bly a forerunner of the Eze Nri, the king of Nri, a highly ritualistic monarchy surviving in northern Igbo territory. Nri may have been influenced by the Igala and seems in turn to have exercised considerable influence in earlier times not only on the Igbo but also on the Igala and other peoples of the Niger–Benue confluence.

Kingdoms and empires of precolonial Nigeria. *Kanem-Bornu.* While there is no continuous evidence to link the peoples of the Jos Plateau with the Nok culture, or the Eze Nri of today with Igbo Ukwu, the history of Bornu (Borno) antedates the 9th century, when Arabic writers in North Africa first noted the kingdom of Kanem, east of Lake Chad. The lake was then much larger, and its basin attracted populations and encouraged exchange. A pastoral group, ancestors of the Kanuri, was establishing a centralized state over those referred to collectively as the Bulala, or the Sao. Initially, the trading links were with Egypt and the Nile valley. There is some evidence of contact with the Christian kingdoms of Nubia before they were overrun by Islām, but it was Islām that gained a foothold in the ruling family of Kanem in the 11th century. From Kanem the rulers tried to dominate the areas south and west of the lake as well. By the 12th century, they were compelled by the attacks of the Sao to move the capital to the region west of Lake Chad and gradually lost control of most of the original Kanem.

For a long time, Bornu was the dominant power in the central Sudan, including much of Hausaland. The Bayajidda legend, concerning the mythical Middle Eastern ancestor of the Hausa, seems to suggest that the rise of a centralized political system in Hausaland was influenced from Bornu. Though the rulers of Bornu embraced Islām, the structure of the monarchy remained traditional, with the queen mother and other female officials exercising considerable power. The selection of the monarch, the coronation rites, and other bases of the royal authority were dictated by pre-Islāmic beliefs. The princes and other members of the royal family were granted fiefs and posted away from the capital to govern frontier zones, while people of slave origin were preferred for the royal guard and palace officials.

Hausaland. The Hausa occupy the northern plains beyond the Jos Plateau, west of Bornu, a crossroads open not only to Bornu but also to the states of the western Sudan, especially Mali and Songhai to the west; the trans-Saharan routes to the Maghrib; and various trade routes to the forest areas, through Nupe and Gwariland to Borgu, Old Oyo, and Benin. Perhaps because of this strategic location, the Hausa developed a number of centralized states—Daura, Katsina, Kano, Zaria, Gobir, and, later, Kebbi being the best-known—each with a walled city, a market centre, and a monarchical system of government.

Introduction of Islām

Islām, which was introduced from the Mali empire in the 14th century, strengthened both the monarchical system and the commercial contacts. But it remained predominantly an urban religion until the beginning of the 19th century, and, even within the walled cities, some pre-Islāmic rites continued to feature in the ceremonies that sustained monarchical authority. There was considerable rivalry among the different states, not only about the area of agricultural land that each controlled but even more about the control of trade and trade routes. This lack of unity was both a cause and a result of periodic conquests from powerful neighbours, especially Bornu, Songhai, and, especially in the 17th and 18th centuries, Jukun.

Yorubaland and Benin. In the forested areas west of the Niger and south of Hausaland, it would appear that Ife, which flourished between the 11th and 15th centuries, emerged as the first major power. There some of the characteristic features of Yoruba culture emerged: a monarchical system based on city-states and nucleated villages; a pantheon of gods in which a few enjoyed widespread recognition, but with several variations locally; and Ifa divination with its corpus of sacred chants. Ife is best known for its potsherd pavements and the great artistry of its terra-cottas and bronzes, especially the naturalism of many of its bronze figures. The influence of Ife on surrounding states is attested from the fact that the monarchies in all the Yoruba states claim descent from Ife as a way of establishing legitimacy, sometimes borrowing regalia from Ife to use in coronation rites and sometimes sending remains of deceased rulers to Ife for burial.

By the 15th century, monarchies in Benin and Old Oyo that derived from Ife were becoming stronger militarily and in the amount of land and peoples they controlled. Old Oyo was located beyond the forests. After more than a century of struggle with Borgu and Nupe, it established itself strategically as the emporium for exchanging goods from the north—rock salt, copper, textiles, leather goods, and horses—with products of the forests—kola nuts, indigo, parrots, and cowries. By the 17th century, it had built up a cavalry force with which it dominated people in western Yorubaland and in the dry gap to the coast.

When the Portuguese arrived in the kingdom of Benin in the 15th century, they found a monarchy with a complex structure of chiefs and palace officials presiding over a kingdom that was expanding in all directions. In time, Benin dominated not only the Edo-speaking peoples to the north and south but also the area eastward to the Niger and, along the lagoon, to Lagos and beyond. It also exerted considerable influence on eastern Yorubaland and maintained trading connections with Old Oyo. Benin art, beginning to flourish in the 15th century, was to maintain its vigour until the colonial period.

Igboland and the Delta states. Many Nigerian peoples did not develop centralized monarchical states. Of these, the Igbo are probably the most remarkable because of the size of their territory and the density of population. The characteristic decentralized pattern of life, in spite of the traditions of Nri, seems deliberate; monarchical institutions in such outlying parts as Asaba, Onitsha, and Aboh were due probably to the influence of the Igala and Benin. Igbo lineages were organized in self-contained villages or federations of village communities, with societies of elders and age-grade associations sharing various governmental functions. The same was true of the Ijo of the Niger Delta and peoples of the Cross River area, where secret societies also played a prominent role in administration. By the 18th century the needs of the overseas trade began to encourage the emergence of monarchical structures.

Initial Portuguese contacts were focused on Benin and Warri. By the 17th and 18th centuries, in the heyday of the slave trade, the delta city-states had become the principal outlets of the trade. Various coastal communities organized to take advantage of the trade as middlemen, so that they would not become its victims. Similarly, the Igbo, like the Edo and the Yoruba kingdoms, organized to get supplies of slaves to the coast. The deleterious effect of the trade on the society and the economy was felt everywhere, but in terms of loss of population, the ones who suffered most appear to have been the noncentralized peoples of the middle belt.

The Sokoto jihad. At the beginning of the 19th century, Islām was well established at all the major centres of the Hausa states and Bornu. The *etsu* (ruler) of Nupe had accepted Islām, and a few teachers and itinerant preachers were also known in parts of the Old Oyo empire. A group of Muslim intellectuals, mostly Fulani, led by Usman dan Fodio, were unhappy that in all these places the rulers allowed Islām to continue to be mixed with aspects of traditional religion and that nowhere was Islāmic law, the Sharī'ah, observed in full. After 20 years of writing, teaching, and preaching in Gobir and surrounding states, Shaykh Usman withdrew his followers to Gudu, where they formally proclaimed him *amīr al-mu'minīn* (commander of the faithful), pledged their loyalty, and prepared for war. In 1804 he called on his followers and all lovers of true Islām to rise up and overthrow the unjust rulers. He appealed to the masses of slaves and to the pastoral Fulani as oppressed people to join the revolt.

Usman dan Fodio

It indicates the degree of communication that existed at the time among different peoples in what was to become Nigeria that the call to jihad made in Gudu in the northwest corner had repercussions throughout the country. The amount of interaction along trade routes and rivers draining the northern plains to the Niger–Benue valley, through the delta, and across the coastal lagoons should not be underestimated. The call to jihad was answered

not only in the different Hausa states, such as Kano, Katsina, and Zaria, but also in Bornu, Bauchi, Gombe, and Adamawa, and eventually in Nupe, Ilorin, and other places where there were pockets of Fulani scholars.

Thus was created the caliphate, with its seat at the newly established town of Sokoto. Each emirate enjoyed autonomy but pledged loyalty to the *amīr al-mu'minīn* and made contributions for the upkeep of Sokoto. Disputes within or between emirates were referred to Sokoto for settlement by officials who traveled as often as possible to oversee developments. Usman himself retired in 1811 to concentrate on the intellectual direction of the movement, which followed the teachings of the Qādirī brotherhood and strict adherence to the Mālikī code of laws. His brother Abdullahi and his son Muhammad Bello carried on the jihad and laid the basis of administration. When Usman died in 1817, Muhammad Bello succeeded him as *amīr al-mu'minīn,* while Abdullahi as emir of Gwandu was given charge of the western emirates, notably Nupe and Ilorin. In this way, all the Hausa states, parts of Bornu, Nupe, Ilorin, and Fulani outposts in Bauchi and Adamawa were drawn into a single politico-religious system. The rulers of Bornu invited Shaykh Muhammad al-Amin al-Kānemī, a distinguished scholar and statesman who disagreed with the Fulani view that jihad was permitted against backsliding Muslims, to lead their forces in the defense of Bornu against the forces of the Fulani jihad. In the process Islām was revived in Bornu, and the old Seyfawa dynasty was eventually replaced by that of Shaykh Muhammad al-Kānemī.

The collapse of Old Oyo. Fulani intrusion into Ilorin was largely the result, rather than the cause, of the collapse of the Old Oyo empire. Deep-seated conflicts arose between the *alafin,* or ruler, and his chiefs, both provincial rulers and lineage chiefs and councillors at the capital. In spite of the external threat from the Fulani and others, the conflicts could not be resolved. Fulani ascendancy at Ilorin cut off the supply of horses to Old Oyo and made the capital untenable. Large groups of Oyo people had to migrate southward to establish a new capital at present-day Oyo and at other centres such as Ibadan and Ijaye. This pressure in turn pushed the Egba farther south to seek a new home at Abeokuta. The collapse of the Old Oyo empire unleashed a major redistribution of the Yoruba people and a series of Yoruba wars that lasted until 1886.

The coming of the British. The Sokoto jihad and the Yoruba wars stimulated the slave trade just when the British were actively trying to stop it. Slaves were traded for European goods, especially guns and gunpowder. The British encouraged the trade in palm oil in the Delta states, ostensibly to replace the trade in slaves. They later discovered that the demand for palm oil was in fact stimulating the internal slave trade, because slaves played the major role in collecting palm fruits, manufacturing palm oil, and transporting it to the coast by canoe or by human porterage. The palm oil trade was linked to the Sokoto jihad and the Yoruba wars because many warriors recognized the importance of slaves, not only as soldiers and producers of food to feed soldiers but also as producers of palm oil to trade for European Dane guns and other goods.

Growth of palm oil trade

Many of the slaves exported in the 1820s and '30s were intercepted by the ships of the Royal Navy, emancipated, and deposited in Sierra Leone under missionary tutelage. Some of these began to migrate back from Sierra Leone in search of home and trade. They invited missionaries to follow them and, in the 1840s, made themselves available as agents for gaining entry not only for missionaries but also for British trade into such places as Lagos, Abeokuta, Calabar, and various locations on the Niger, such as Lokoja, Onitsha, Brass, and Bonny. In 1841, the British tried to settle some of them on a model farm in Lokoja, but the plan was aborted because of the high rate of mortality of European officials. It was also partly to protect them that the British shelled Lagos in 1851, expelled Kosoko, the reigning *oba,* and restored his uncle Akitoye, who appeared more willing to join in a campaign to abolish the slave trade. In 1861, to protect Akitoye's son and successor, foil Kosoko's bid to return, and secure a base for further activities, the British annexed Lagos.

The British were as yet unwilling to run the expense of maintaining an administration in Nigeria. To reduce expenses, Lagos was administered from Freetown in Sierra Leone, along with the Gold Coast forts, and later from Accra (in present-day Ghana). It was not until 1886 that Lagos became a separate colony. A consul was maintained at Fernando Po to oversee the lucrative trade in palm oil in the region called the Oil Rivers. Missionaries were active: Presbyterians in Calabar, the Church Missionary Society (CMS), Methodists, and Baptists in Lagos, Abeokuta, Ibadan Oyo, and Ogbomoso. The CMS took the initiative to pioneer trade on the Niger by encouraging a private businessman, Macgregor Laird, to run a monthly steamboat to western Africa so as to provide transportation to missionary agents and Sierra Leonean traders going up the Niger. In this way Bishop Samuel Ajayi Crowther of the CMS was able to establish mission stations at Onitsha, Lokoja, and Eggan, and later at Brass and Bonny.

By the 1870s, the Niger trade was becoming profitable, and a few French companies took notice. French Roman Catholic missionaries, established in Ouidah (Whydah), arrived in Lagos and considered missionary work on the Niger. The British responded to such evidence of rivalry. By the 1880s, they showed every sign of their determination to keep out all possible rivals including not only the French but also the Sierra Leonean pioneers of the Niger trade. At the Berlin West African Conference of 1884–85, the British defended the right of free navigation on the Niger; at the same time George Taubman Goldie bought out all French rivals and created a company, to which the British granted a charter, to control trade on the Niger and administer the immense territories of the Sokoto caliphate and Bornu. In addition, two other protectorates were declared, one over the Oil Rivers and the other over the hinterland of Lagos, to establish a claim that these areas were also British "spheres of interest."

Thereafter two developments should be noted. The first was the attempt to establish to Nigerian peoples that the British were no longer, as hitherto, negotiating for rights as traders and missionaries but rather were taking over control as rulers and overlords. Rulers that seemed to stand in their way, such as Jaja of Opobo and Nana of Itshekiri, were deposed and deported. While taking over Yorubaland in the guise of peacemakers between the warring parties, they decided to make an example of the Ijebu in 1892 by mounting an expeditionary force to demand that trade routes be kept open. A major force was sent to conquer Benin City in 1897. In that year the Royal Niger Company, having no force to dispose of, employed Frederick Lugard to assemble an army and assert the authority of the company in Bida and Ilorin.

The second development was that the boundaries of the two protectorates and the territories of the Royal Niger Company were difficult to define. There was rivalry between the Royal Niger Company and the Lagos Protectorate over the boundary between the emirate of Ilorin and the empire of Ibadan, just as there was tension between the Lagos and Oil Rivers protectorates over the boundary to be drawn between Benin and Yorubaland. The latter problem was eased in 1894 when both protectorates were merged into the Niger Coast Protectorate. The abrogation of the charter of the Royal Niger Company on Jan. 1, 1900, in return for wide mineral concessions eased the other problem. All the territories were now under British control, and the search for an identity had begun, first as Northern and Southern Nigeria, then with eventual amalgamation. All these were acts of the British, but they do not justify the claim that Nigeria was a creation of the British: it was the logic of the precolonial experience of the Nigerian peoples that dictated the wisdom of administering all the territories as one country. (J.F.A.A.)

The 20th century. After the British government assumed direct control of the Royal Niger Company's territories, the northern areas were renamed the Protectorate of Northern Nigeria, and the land in the Niger Delta and along the lower reaches of the river was added to the Niger Coast Protectorate, which was renamed Southern Nigeria. Lagos remained the capital of the south, with Zungeru the new capital of the north. On Jan. 1, 1914, following

Southern Nigeria and Northern Nigeria

the recommendations of Sir Frederick Lugard, the two protectorates were amalgamated to form the Colony and Protectorate of Nigeria under a single governor-general resident in Lagos. Between 1919 and 1954, the title reverted to that of governor.

Indirect rule. As the first High Commissioner of Northern Nigeria, Lugard had been instrumental in subjugating the Fulani emirs. Some were deposed, some defeated in battle, and others collaborated. It was not until 1903 that the conquest of the emirates was complete. The mud-walled city of Kano was captured on February 3 and, after a vigorous skirmish at Kotorkwashi, the sultan's capital, Sokoto, fell on March 15. Lugard's proclamation following the appointment of a new sultan set out the principles of the administrative system subsequently institutionalized as "indirect rule." Essentially, local government was to be left in the hands of the traditional chiefs, subject to the guidance of European officers. Emphasis was placed on utilizing native institutions and on minimum interference with local customs. Although it had built-in contradictions, over the years the Nigerian system developed into a sophisticated form of local government, especially in the emirates, and under the title of "native administration" became the hallmark of British colonial rule in Africa. In Nigeria the major uprisings at Satiru in 1906, Abeokuta in 1918, and Aba in 1929, all involving grave loss of life, were to raise questions about its value as a model.

Further dislocation accompanied the outbreak of World War I. Locally this involved the immediate invasion of the German-held Kamerun by Nigerian forces, followed by a costly campaign that lasted until 1916. Later, Nigerian troops were sent to East Africa. (During World War II they again served in East Africa, as well as in Burma.) In 1922 Kamerun was divided under a League of Nations mandate between France and Britain. The latter administered its area within the government of Nigeria. After 1946, the mandated areas were redesignated as a UN Trust Territory.

Constitutional changes. Structurally, the constitutional amalgamation (it was not an administrative unification) of 1914 remained unchanged until the accession of the mandated areas in 1922 and the division of the Southern provinces into Western and Eastern in 1939. In 1954 the groups of provinces were renamed the Northern, Western, and Eastern regions as part of the reconstruction of Nigeria into a federal state.

In 1914 a legislative council was established, but its jurisdiction did not extend beyond the colony. A larger council was set up in 1922, with elected members (three from Lagos, one from Calabar). Its powers were limited and the Northern provinces remained outside the jurisdiction of the Nigerian Legislative Council. A radical change occurred in the quasi-federal Richards constitution of 1946. Each geographic group of provinces had its own house of assembly, with a majority of nonofficial (though not yet all elected) members. Two houses of chiefs were created, and there was a central legislative council in Lagos.

These changes failed to satisfy the groundswell of public demand for self-government. In particular, Nigerians resented the lack of discussion over the Richards constitution. A nationwide consultative process ensued, down to village level, culminating in a constitutional conference at Ibadan in 1950. The 1951 Macpherson constitution provided for a central house of representatives, but friction between the central and regional legislatures, related to the question of where supreme party authority lay, soon caused a breakdown.

The Lyttelton constitution

In 1954 Nigeria introduced the Lyttelton constitution, its third in eight years. This time, at the Nigerians' urging, it was a fully federal one, comprising the three geographic regions, the Southern Cameroons, and the Federal Territory of Lagos. Each region had a governor, premier, cabinet, legislature, and civil service, with the significantly weaker federal government represented in Lagos by a governor-general, bureaucracy, house of representatives, and senate. The post of prime minister was created in 1957 and was filled by Sir Abubakar Tafawa Balewa.

Internal self-government was granted to the Western and Eastern regions in 1957 but, at its request, not to the Northern region until 1959. Northerners were fearful that their region might lose its claim to an equal share in the operation and opportunities of the federal government if it were not given time to catch up with the educationally advanced south. Among the problems calling for attention before the British would grant full independence was that of the minorities' fear of discrimination by a future government based on majority ethnic groups. After the Willink Commission reported on this issue in 1958, independence was granted on Oct. 1, 1960, under a federal constitution, with an elected prime minister and a ceremonial head of state. Following a UN-supervised referendum, the northern part of the Trust Territory of the Cameroons joined the Northern Region on June 1, 1961, while on October 1 the Southern Cameroons united with Cameroun to form the Federal Republic of Cameroon. In August 1963 the Mid-West region was created by dividing the Western Region, and on October 1 Nigeria became a republic. Nnamdi Azikiwe was the first president.

Independent Nigeria. After a honeymoon period, Nigeria's long-standing regional stresses of ethnic competitiveness, educational inequality, and economic imbalance again came to the fore in the controversial census of 1962–63. In the west the government fell apart in 1962. At the centre, a boycott over the federal election of 1964 brought the country to the brink of breakdown. The point of no return was reached on Jan. 15, 1966, when, after the collapse of order in the west following the fraudulent election of October 1965, a group of army officers attempted to overthrow the federal government. Prime Minister Balewa and two of the regional premiers were murdered. A military administration was set up under Major General Johnson Aguiyi-Ironsi, but his plan to abolish the regions and impose a unitary government met with anti-Igbo riots in the north. He was assassinated on July 29, 1966, and Lieutenant Colonel (later General) Yakubu Gowon came to power.

Gowon's attempt to hold a conference to settle the constitutional future of Nigeria was abandoned after the ethnic massacres of October 1966. A last-hope effort to save the country was made in January 1967 when the Eastern delegation, led by Lieutenant Colonel (later General) Odumegwu Ojukwu, agreed to meet the others on the neutral ground of Aburi, Ghana, but the situation deteriorated after differences developed over the interpretation of the accord. On May 27 the Eastern Region's consultative assembly authorized Ojukwu to establish a sovereign republic. At the same time the federal military government promulgated a decree dividing the four regions into 12 states, including six in the north and three in the east.

The civil war. On May 30, 1967, Ojukwu declared the secession of the three Eastern states under the name of the Republic of Biafra. The federal government interpreted this as an act of rebellion. Fighting broke out on July 6 and within weeks Gowon's so-called "police action" escalated into a full-scale war. In August Biafran troops crossed the Niger, seized Benin City, and were well on their way to Lagos before they were checked at Ore. Then federal troops entered Enugu and penetrated the Igbo heartland. The next two years were marked by stiff resistance in the shrinking Biafran enclave and by heavy casualties among civilians as well as in both armies, all set within what threatened to be a military stalemate. Peacemaking attempts by the Organization of African Unity remained ineffective. Biafra began to earn recognition by African states and to secure aid for its starving population from international organizations. Then suddenly the conflict was over. Ojukwu fled to Côte d'Ivoire on Jan. 11, 1970, and on January 15 a Biafran deputation formally surrendered in Lagos.

Formation of Biafra

General Gowon announced a program of reconstruction and reintegration. In 1974 he postponed the 1976 target date for a return to civil rule but was overthrown on July 29, 1975, by Brigadier Murtala Ramat Mohammed, who was assassinated on Feb. 13, 1976. It was left to Mohammed's successor, Lieutenant General Olusegun Obasanjo, to implement the revised program.

Shehu Usman Aliyu Shagari was installed as president of the Second Republic on Oct. 1, 1979. His regime ran

into severe trouble, especially after the collapse of oil prices, and, although he was reelected in mid-1983, he was overthrown by the military on Dec. 31, 1983. Major General Mohammed Buhari's authoritarian Supreme Military Council was overturned by another coup (Nigeria's sixth in 20 years) when Major General Ibrahim Babangida became head of state on Aug. 27, 1985. Babangida's Armed Forces Ruling Council (AFRC) concentrated on economic reforms. After postponing the date of transfer to a civil government from 1990 to 1992, the AFRC began a transitional program toward a Third Republic.

For later developments in the history of Nigeria, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (A.H.M.K.-G.)

Sierra Leone

The Republic of Sierra Leone is a sovereign country in western Africa, bordered on the north and east by Guinea, on the south by Liberia, and on the west by the Atlantic Ocean. One of the smaller African countries, it has an area of 27,699 square miles (71,740 square kilometres) and is larger than Togo and The Gambia but smaller than Benin. Its population is large in relation to the nation's size by African standards. The national capital of Freetown commands one of the largest natural harbours in the world.

PHYSICAL AND HUMAN GEOGRAPHY

The country owes its name to the 15th-century Portuguese explorer Pedro de Sintra, the first European to sight and map Freetown harbour. The original Portuguese name of Serra Lyoa (Lion Mountains) referred to the range of hills that surrounds the harbour.

Although most of the nation's population is engaged in subsistence agriculture, Sierra Leone is also a mining centre. Its land yields important quantities of diamonds, gold, bauxite, and rutile (titanium dioxide). Increasing urbanization has resulted in the gradual depopulation of the rural areas and the growth of a jobless population in the cities. Migrants have also swelled the population of the mining areas.

The land. *Relief.* The country can be divided into four distinct physical regions. The coastal swamp region extends along the Atlantic for about 200 miles (320 kilometres). It is a flat, low-lying, and frequently flooded plain that is between 20 and 40 miles wide and is composed mainly of sands and clays. Its numerous creeks and estuaries contain mangrove swamps. Parallel ridges, often separated by silting lagoons, are common and sometimes form the actual coast. The Sierra Leone Peninsula, which is the site of Freetown, is a region of thickly wooded mountains that run parallel to the sea for about 25 miles. The Peninsula Mountains rise from the coastal swamps and reach 2,913 feet (888 metres) at Picket Hill.

Inland from the coastal plain is the interior plains region. In the north it comprises featureless grasslands (savanna) that are known as "Bolilands" (*Boli* being a Temne word for those lands which are flooded in the rainy season and dry and hard in the dry season, and on which only grass can grow). In the south the plains comprise rolling wooded country where isolated hills rise abruptly to more than 700 feet. The interior contains a variety of landforms ranging from savanna-covered low plains to rocky scarp and hill country. The plateau region, encompassing roughly the eastern half of the country, is composed mainly of granite with a thick laterite (iron-bearing) crust; to the west it is bounded by a narrow outcrop of mineral-bearing metamorphic rocks known as the Kambui Schists. Rising above the plateau are a number of mountain masses; in the northeast the Loma Mountains are crowned by Mount Loma Mansa (Mount Bintimani) at 6,390 feet, and the Tingi Hills rise to 6,079 feet at Sankanbiriwa Peak.

Drainage and soils. The country's drainage pattern is dense. Numerous rivers rise in the well-watered Fouta Djallon highlands of Guinea and flow in a general northeast to southwest direction across Sierra Leone. Their middle courses are interrupted by rapids that restrict navigability to only a short distance inland. River levels show considerable seasonal fluctuations.

The drainage system has nine major rivers and a series of minor coastal creeks and tidal streams. From north to south, the principal rivers are the Great Scarcies, Little Scarcies, Rokel (known in its lower course as the Sierra Leone River), Gbangbaia, Jong, Sewa, Wanje, Moa, and Mano. The Great Scarcies and Moa form portions of the border with Guinea, while the Mano forms much of the country's frontier with Liberia. The major rivers

In most areas, the dominant soils are of the weathered and leached lateritic (iron-bearing) type. Red to yellow-brown in colour, they contain oxides of iron and aluminum and are acid. Kaolin (china) clays are important in some areas, and when cultivated a light, readily workable, free-draining soil results, whose productivity depends largely on the nutrients provided from the vegetation previously cleared and burned. In the coastal plains lateritic soils developed on sandy deposits are agriculturally poor, but those derived from basic igneous rocks are somewhat better. Swamp soils occur over large areas on the coastal plains where drainage is a problem. In coastal and estuarine areas where mangrove is the natural vegetation, productive soils can be acquired by clearance, but careful water control is sometimes needed to prevent toxicity. At the foot of the main escarpment, on the Sula Mountain plateau, and elsewhere an iron-rich laterite crust forms a surface that is intractable for agricultural production.

Climate. The climate is tropical and is characterized by the alternation of the rainy and dry seasons. Conditions are generally hot and humid. Mean monthly temperatures range from 77° F (25° C) to 83° F (28° C) in low-lying coastal areas; inland the range may be from 73° F (23° C) to 82° F (28° C). In the northeast, where extremes of temperature are greater, mean daily minimums fall to 56° F (13° C) in January, and mean daily maximums rise to 90° F (32° C) in March. During the rainy season, from May to October, humid air masses from the Atlantic dominate the weather. The sky is cloudy, the winds are southwesterly, sunshine is minimal, and rainfall is almost a daily occurrence, especially during July and August. The rainy season usually begins with a series of violent line squalls marked by thunder and lightning. Precipitation is greater on the coast than inland; as much as 200 inches (5,080 millimetres) of rain falls annually on the Peninsula Mountains, while the northeast receives about 80 inches a year.

The dry season, from November to April, is characterized by the dry harmattan that blows from the Sahara. The rainy season tends to have cooler daily maximum temperatures than the dry season by about 10° F (6° C). The relative humidity, however, may be as high as 90 percent for considerable periods, particularly during the wettest months, from July to September.

Plant and animal life. The distribution of plants and animals has been influenced by such factors as relief and soil types and, perhaps more importantly, by farming methods. Remnants of the extensive original forest cover remain in the Gola Forest reserve in the southeastern hill country near the Liberian border. Secondary forest is now dominant, and valuable timber species, such as *Khaya* (African mahogany) and African teak, that were common in the original forests are now rare. The secondary forest is characterized by other tree species, such as the fire-resistant palm tree, a valuable source of palm oil and kernels.

The prevalence of savanna vegetation increases to the north as rainfall decreases. The savannas owe their present extent and character largely to the erosion produced by farming, grazing, and the use of fire. There are some small areas of climax savanna (a closed area of broad-leaved, low-growing trees) and tall tussocky grasses, but these are only remnants of a much wider spread. Other savannas are derived from forest and are characterized by fire-resistant savanna trees with tall grasses. Tracts of tall-grass savanna also occur. Remnants of mangrove swamps constitute the main coastal vegetation community, especially in the saline tidal areas of river estuaries. Piassava, a kind of raffia palm, is common in the swamps of the south. Coastal mangrove swamps

Large game animals, such as elephants, leopards, lions, hyenas, and buffalo, are rarely seen. Chimpanzees and various species of monkeys are common in the forest

zones, while tiger cats, porcupines, antelope, and bushpigs are more generally distributed. There is a wide variety of insects, including the malaria-carrying mosquito and the tsetse fly. Hippopotamuses, crocodiles, manatees, and alligators occupy the rivers. The coastal waters, estuaries, and rivers, such as the Sierra Leone and Sherbro, also contain a wide variety of fish, such as tuna, barracuda, and mackerel, as well as lobsters and sharks. Bird life includes parrots, owls, kingfishers, green pigeons, African magpies, vultures, and many other species.

Settlement patterns. Villages of about 35 buildings and 300 inhabitants dominate the rural landscape. Modernization is slowly altering the traditional pattern of rural settlement; the old circular village form, with a tight cluster of houses, is rapidly yielding to the linear village along a road or the regular gridiron pattern with adequate spacing between houses. Economic activity still centres largely around rice farming. The extended family provides farm labour for both rice farming and cash crop production. Fishing is becoming increasingly important. The raising and herding of cattle is largely confined to the north. The small shopkeeper is typical of the villages, as are the tailor and carpenter. Traditional crafts, such as metalworking, cloth dyeing and weaving, and woodworking, are rapidly disappearing with the increased importation of cheap manufactured goods.

Urban development

Except for Freetown, the development of large towns occurred only after World War II. Because of the incipient nature of urbanism, functional specialization is rudimentary, the most prominent feature being the daily market of petty traders, the majority of whom are women.

Freetown is a long, narrow city. Central Freetown—the administrative and commercial hub of the city—houses government buildings and embassies, the law courts, hotels, and the Roman Catholic and Anglican cathedrals. Eastern Freetown is mainly residential, with retail trading concentrated along major roads; there are also a few impressive mosques in the area. Farther east is the port area. Western Freetown is functionally similar to the east; it also contains the nation's main stadium, the central prison, and administrative offices. On Mount Aureol, overlooking the city, is Fourah Bay College, black Africa's first institution of higher learning.

Bo, located in the southeast, is the second largest town. An early administrative and educational centre, it has shown rapid expansion and has engulfed several surrounding villages. Commerce and retail trade are located in the centre of town near the important daily market. Its role as an educational centre has increased steadily, and Bo now contains many schools, including the Bo Government Secondary School, founded in the early 20th century.

Other important towns include Kenema, east of Bo, which has grown as a result of diamond mining, and, in the north, Makeni, a provincial capital and major commercial centre. Mining has also been important to Koidu, Sefadu, Yengema, and Jaiama in the east and Lunsar in the north. Port Loko, Kabala, Bonthe, Moyamba, Kailahun, Kambia, Pujehun, and Magburaka are administrative centres with retail trading and produce marketing.

The people. There are about 18 ethnic groups that exhibit similar cultural features, such as secret societies, chieftaincy, patrilineal descent, and farming methods. The Mende, found in the east and south, and the Temne in the north form the two largest groups. Other major groups include the Limba, Kuranko, Susu, Yalunka, and Loko, in the north; the Kono and Kissi, in the east; and the Sherbro, in the southwest. Minor groups include the coastal Bullom, Vai, and Krim and the Fulani and Malinke (Mandingo), who are immigrants from Guinea concentrated in the north and east. The Creoles—descendants of free blacks who colonized the coast in the 19th century—are found mainly in the Western Area and Freetown. Ethnic complexity is further enhanced by the presence of Lebanese and Indian traders in urban centres.

Linguistic patterns

Krio, a language derived from English and a variety of African languages, is the mother tongue of the Creoles and the country's lingua franca. Among the indigenous languages, the Mande group is the most widespread; it includes the Mende, Kuranko, Kono, Yalunka, Susu, Vai, and Malinke languages. The Mel group, which is similar to the Bantu languages of Central and East Africa, includes the Temne, Krim, Kissi, Bullom, Sherbro, and Limba languages. English, the official language, is used in administration, education, and commerce. Arabic is used among Lebanese traders and adherents of Islām.

About half the population practices a variety of animist religions. More than one-third of the people are Muslims, and the rest are Christians.

The economy. Private capital dominates mining concerns, commerce, and banking. European, Lebanese, and Indian interests are predominant, and participation by Sierra Leoneans is small. The public sector features the Sierra Leone Produce Marketing Board (SLPMB), which has a monopoly on cash crops, and other public corporations, such as that of road transport, which is entirely owned by the government. Various inefficient parastatals were privatized in the 1980s. Government revenue is derived from direct and indirect taxes. In addition to import and export taxes, the government can also rely on company, excise, income, and mining taxes for revenue. The government's revenue from trade has been undermined by the growth of smuggling of diamonds and agricultural produce.

There were growing economic difficulties in the 1980s, including a heavy external debt burden, escalating costs of food and fuel imports, and erratic mineral export production. Substantial devaluations of the national currency, the leone, occurred, and a series of International Monetary Fund-supported economic stabilization programs were initiated to address these problems.

Mineral resources

Resources. Mineral resources are fairly well distributed and include diamonds, chromite, and reserves of rutile that are among the world's largest. There are iron ore reserves, but these are no longer commercially mined. Other minerals include bauxite, columbite (a black mineral of iron and columbium), pyrochlore, gold, platinum, and monazite. Major concentrations of those minerals have been found in the southern plateau region.

Forest covers more than one-fourth of the country, the most important area of which is the Gola Forest Reserve, a tract of primary tropical rain forest, near the Liberian border. Biological resources include large herds of tsetse-resistant Ndama cattle. The coastal waters provide rich fishing grounds for shad, herring, snapper, tuna, shrimp, and lobster. The hydroelectric power potential of Sierra Leone's deeply incised river valleys is appreciable.

Agriculture and forestry. Agriculture is carried out largely by traditional methods. More than three-fifths of the population engages in production for the domestic and export markets. Rice, the main food crop, is widely cultivated on swampland and upland farms. Swamp-rice cultivation is concentrated in the lower reaches of river basins, of which the Scarcies is the most important. Efforts are being made to reduce upland-rice farming with its attendant soil erosion in favour of swampland farming with its superior yields. Other food crops include millet, peanuts (groundnuts), cassava, sweet potatoes, and oil palms. Vegetable gardening is important around the major urban centres, where markets are available to farmers. In the 1970s, the government attempted to improve agricultural productivity through integrated development projects funded by the World Bank. However, the country has become a net food importer, especially of rice.

The major cash crops are palm kernels, cocoa, coffee, piassava, and ginger. Production is carried out entirely by small-scale farmers. Timber (which includes *Guarea cendrata,* a cedar-scented, pink, mahogany-type wood, and the *Lopjara plata* variety *procera,* or red ironwood) is produced for the domestic market; there are major sawmills located in the Gola Forest; the main furniture factory is in Kenema.

The mining industry

Industry. Mining is the most important industry in terms of employment in the modern sector and in exports and is second only to agriculture in its contribution to the national economy. Diamonds are mined by the National Diamond Mining Company (Diminco), by a few private companies, and by vast numbers of private prospectors. Mining methods range from mechanical grab lines with

Preparing piassava fibre for shipment, Bonthe, Sierra Leone.
J. Allan Cash Photolibrary

washing and separator plants to crude hand digging and panning. About half of the diamonds are gem stones found in river gravels, especially along the Sewa-Bafi river systems. Official exports of diamonds have declined dramatically since the 1960s due to extensive smuggling and the depletion of reserves. There have been several changes in the ownership of Diminco, and investment has been sought for deep-mining of diamonds as the alluvial sources are depleted.

The privately owned Sierra Leone Development Company (SLDC) mined iron ore at Marampa from 1935 to 1975. In 1981 the government reopened the mine at Marampa, but soon encountered financial difficulties and suspended operations in 1985. During 1963 the Sierra Leone Ore and Metal Company (Sieromco) opened open-cast bauxite mines at Mokanji Hills. The ore is shipped to Europe for reduction and refining into aluminum. Rutile is found in the southwest; in 1965 rutile production was begun, and production and prospecting activities were in progress in the early 1980s.

Industrialization is restricted largely to import substitution. Manufacturing is concentrated in Freetown, and production is mainly of consumer goods, such as cigarettes, sugar, alcoholic beverages, soap, tires, textiles, mineral fuels, and lubricants. Although factories are small and employ fewer than 1,000 workers each, their role in economic diversification is important. In the provinces industries are concerned exclusively with the processing of agricultural and forest produce, such as rice, timber, and palm oil. Traditional industries, such as fish curing and leatherwork, continue.

Electricity is generated primarily by thermal plants, which are supplemented by a few small hydroelectric installations.

Financial services

Finance and trade. The Bank of Sierra Leone is the nation's central bank; it issues currency, maintains external reserves, and acts as banker and financial adviser to the government. The National Development Bank is charged with providing finances to investors within the country. The Sierra Leone Commercial Bank provides credit and technical assistance to farmers. Post Office savings banks are found in all main towns, and there are also various kinds of thrift and credit societies.

Foreign trade has expanded substantially since independence, although its character still reflects the colonial nature of the economy. An excessive reliance is placed upon a few primary products, most of which go to the United Kingdom, the United States, and western Europe. Minerals and agricultural products account for the bulk of exports. Imports, however, have become more diversified; they include machinery, textiles, vehicles, fuel, and food products.

Transportation. A government railway was completed in 1916 as a means of opening the country to commerce and ensuring effective British occupancy. The railway was phased out by 1975, however.

A road network, originally developed as a feeder system to the railway, has become the principal transport carrier. The network is dominated by a series of highways radiating from Freetown to inland urban centres. The government launched a long-term program in the late 1980s to modernize the road system to meet the needs of rapidly expanding traffic. In addition, road links with Conakry, Guinea, and Monrovia, Liberia, were to be improved.

Inland waterways carry a considerable volume of mineral ores, piassava, and food products. Launches and sailing boats are important, especially on the southern route to Bonthe and the northern route to the Great and Little Scarcies. Freetown is the country's principal port. Its facilities handle all imports and agricultural exports. Specialized ports include Niti, which handles all bauxite and rutile exports, and Bonthe, which exports agricultural products.

Port facilities

The international airport of Lungi is situated on the north bank of the Sierra Leone River opposite Freetown. It can accommodate commercial jets and a large annual volume of traffic. Domestic air transport is limited.

Administration and social conditions. *Government.* The constitution of 1971 made Sierra Leone a republic within the Commonwealth. Adoption of the constitution of 1978 created a one-party republic based on the All-People's Congress. The head of state, or executive president, is elected by delegates of the All-People's Congress for a seven-year term. The House of Representatives is composed of a speaker and 127 members, 105 of whom are elected for five years by universal suffrage. Ten members are appointed by the president and 12 representatives are paramount chiefs representing the 12 administrative districts of the three provinces. The chiefs are hereditary rulers whose local powers have been largely superseded by those of officials of the central and local government. Their influence remains important, however, particularly in matters of traditional culture and justice.

Local government

The country is divided into four administrative units—the Western Area, which was the former Crown Colony of Sierra Leone, and three provinces (the Northern, Eastern, and Southern provinces), which were the former Protectorate of Sierra Leone. The Western Area includes the capital of Freetown. Northern Province is divided into five districts, Southern Province into four districts, and Eastern Province into three districts.

The districts are subdivided into some 148 chiefdoms, which are controlled by paramount chiefs and chiefdom councillors. The chiefdoms are further divided into sections and villages.

In addition, there are district councils, which in some cases override the chiefdom administrations. The councils deal largely with local matters and are under the indirect control of the central government. Town councils also have been established in the larger provincial towns of Bo, Kenema, Makeni, Koidu, and Bonthe.

Sierra Leone is a relatively politically stable and peaceful society. Compared to its neighbours, the levels of political violence have been low—though by no means absent—and there have been no major political upheavals since the coups d'état of the 1960s. There is a vigorous and at times critical press.

The laws of Sierra Leone follow the pattern of British law. Until 1971 the framework of the courts was equally similar, and the final court of appeal was the Privy Council in London. Since the adoption of a republican constitution, however, the highest court is the Supreme Court, headed by the chief justice.

There are local courts that take account of indigenous laws and customs, Magistrates' Courts administering the English-based code, a High Court, and a Court of Appeal. There are presiding officers in the local, magistrate, and juvenile courts who are not qualified lawyers but who are citizens of wide experience. The attorney general is also the minister of justice.

Education. Education in Sierra Leone is offered in private and government-sponsored schools; it is not compulsory. There are primary schools for children from ages five to 12, secondary schools that also offer a seven-year pro-

gram, technical institutes, and several vocational schools, trade centres, and teacher-training colleges in the country. The University of Sierra Leone consists of Fourah Bay College, founded in 1827, and Njala University College, founded in 1963.

Health and welfare. Most health and welfare services are provided by the central government. There are also a few hospitals belonging to religious societies and mining companies and private doctors. Every district in the interior has at least one hospital. The major hospitals with specialist facilities are in Freetown and Bo.

The Public Health Section controls port and airport sanitation, schemes for the control and eradication of malaria and other infectious or endemic diseases, and the sanitation of Freetown. In other areas sanitation is under the control of district health authorities and town councils supervised by doctors who are appointed by the central government.

There is a fascinating variety of housing in the interior districts, depending on the availability of materials. Roofs are made of grass in the savanna region and of bamboo in the forest areas. Walls may be circular or rectangular and constructed of dried mud bricks, palm fronds, or, more generally, lattice pole work filled with mud and coated with clay or chalk. There is usually a veranda or shaded porch. In most villages and towns along the roads, houses are roofed with corrugated zinc and the walls constructed of cement. In Freetown and Bonthe, some houses remain that were built of wood or laterite stone in a Brazilian or Victorian Colonial style and roofed with slate.

Cultural life. The most outstanding feature of the country's cultural life is its dancing. The Sierra Leone Dance Troupe is internationally known. The different communities of the nation have their own styles of costume and dance. In addition, certain closed societies, such as the Wunde, the Sande (Bundu), and the Gola, have characteristic ceremonial dances. A wide range of agility, gracefulness, and rhythm is displayed; in addition, there are elements of symbolism in most of the dances. Drums, wooden xylophones (called balaphones), and various stringed instruments provide the musical background.

The Poro society for men and the Sande society for girls play an educational role in village culture.

The Vai script has the distinction of being one of the few indigenous scripts in Africa. Some of the local languages are written in European script, but a few, especially in the Muslim areas in the north, have been transcribed into Arabic.

Handicrafts. The carving of various wooden masks in human and animal figures for the dances is especially advanced in the southern region. The Sande mask worn on the head of the chief dancer during the ceremony attending the reappearance of the female initiates from their period of seclusion is perhaps the most well-known carved figure in Sierra Leonean art. It is a symmetrically stylized black head of an African woman with an elaborate plaited pyramidal coiffure adorned with various figures and with a facial expression of grave dignity and beauty.

Ivory figures are characteristic of the Sherbro, Bullom, and Temne peoples of the coastal and northern regions. Fine examples of these figures, which were bought or commissioned by Portuguese traders during the 16th century, are still extant. There are also steatite human figures, sometimes distorted, called *nomoli,* or, in wooden form, *pomtan* (singular, *pombo*), which certainly date earlier than the 16th century and were used probably for ancestor worship or fertility rites. At present, they are used for ceremonies to ensure abundance of crops.

Containers or rattles are carved from gourds and are decorated with intricate geometric patterns that are burned into them.

The weaving and dyeing of cloth

The weaving of blue or brown cloth of thick texture with linear designs is carried out in the southern and eastern regions by the Mende and the Kono. The cloth is made into coats for men or wrapped around as a lower garment by women and is also used as a bedspread. In the north, among the Temne, imported cotton or satin is dyed with indigo, the red juice of the kola nut, or imported dyes into beautiful patterns by tie-dyeing. In the west, baskets are made with dyed raffia, and patterned slippers are fashioned from dyed wool.

Painting and literature. There is an active school of modern artists who are trained in Europe and the United States and whose paintings have been exhibited locally and abroad. Olayinka Burney Nicol, Hassan Bangura, John Vandi, Koso Thomas, and Gladys Metzger are among the best-known artists of Sierra Leone.

School texts, information bulletins, and collections of folktales are produced in indigenous languages such as Mende and Temne. There has been a literary tradition in Freetown since the 19th century. One of the most prolific writers was James Africanus Beale Horton, who wrote books and pamphlets on politics, science, and medicine while he was a medical officer in the British army between 1857 and 1871.

There were also 19th-century works on exploration by such Sierra Leone Africans as Samuel Crowther, a bishop of the Anglican Protestant faith, and another clergyman, John Cristopher Taylor. Sierra Leone is represented in most anthologies of African- and English-language poetry and short stories. In addition, the modern novels and short stories of Sarif Easmon, William Conton, and Eldred Jones give a vivid picture of modern life in the country.

The Sierra Leone National Museum in Freetown contains historical, ethnographic, and archaeological collections. Fourah Bay College and Njala University College both have libraries; the former houses the public archives.

For statistical data on the land and people of Sierra Leone, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (D.S.H.W.N./S.M.S.)

HISTORY

Early history. Archaeological findings show that Sierra Leone has been inhabited for thousands of years. Traditional historiography has customarily presented it as peopled by successive waves of invaders; but the language pattern suggests that the coastal Bulom (Sherbro), Temne, and Limba have been in continuous settled occupation for a long time, with subsequent sporadic immigration from inland by Mande-speaking peoples, including Vai, Loko, and Mende. They organized themselves in small political units—independent kingdoms or chiefdoms—whose rulers' powers were checked by councils. Secret societies, notably the Poro society, also exercised political power as well as instructing initiates in the customs of the country.

Muslim traders brought Islām, which became firmly established in the north and subsequently spread through the rest of the country.

Portuguese voyagers gave the name Serra Lyoa (Lion Mountains), later corrupted to Sierra Leone, to the mountainous peninsula at the mouth of the Rokel River where, from the 15th century onward, European traders congregated near the site of modern Freetown under the protection of African rulers, who welcomed them for the commercial opportunities they provided, exchanging imported manufactured goods for ivory and slaves to be employed across the Atlantic.

A group of freed slaves arrived in Sierra Leone from England in 1787 to form a settlement. It failed but was revived by a commercial company, the Sierra Leone Company, sponsored by English opponents of the slave trade. Black settlers who had liberated themselves from American slavery were brought over from Nova Scotia and built a new settlement, named Freetown. "Maroons," free blacks from Jamaica, were also brought in. These settlers were English-speaking, and many were literate and Christian.

British rule

After the British Parliament made the slave trade illegal in 1807, the British government took over the settlement (Jan. 1, 1808) as a naval base against the slave trade and as a centre to which slaves, captured in transit across the Atlantic, could be brought and freed. Between 1807 and 1864, when the last slave ship case was adjudicated in the Freetown courts, the British Navy brought in more than 50,000 "recaptives." Drawn from all over western Africa, these heterogeneous people lacked any common language or culture. The government therefore introduced a deliberate policy of turning them into a homogeneous Christian community. Protestant missionaries, along with the black

pastors of Freetown churches, worked with such success that within a generation the policy was virtually fulfilled. The (Anglican) Church Missionary Society founded an institution to train teachers and missionaries, Fourah Bay College, which was affiliated to the University of Durham, England, in 1876. The Society also opened boys' and girls' secondary schools.

The recaptives and their children, known as Creoles (today usually rendered Krios), prospered as traders, and some entered the professions, qualifying in Britain as doctors and lawyers. Thus they formed an educated West African elite. Notable examples include James Africanus Beale Horton, who qualified as a doctor and served as an officer in the British army, publishing books on medical and political subjects, and Sir Samuel Lewis, a distinguished barrister. Many Creoles sought employment opportunities in other parts of West Africa. At their suggestion, Anglican missions were founded in what is now Nigeria, where one of them, Samuel Adjai Crowther, became a bishop.

Colony and protectorate. During the 19th century the area surrounding the coastal settlements was drawn increasingly into the British economic sphere. There was a market in Britain for shipbuilding timber, and most of the accessible forest trees in the coastal country were felled, altering the environment irrevocably. There was also a European market for vegetable oils, and unprocessed palm produce and peanuts were supplied in return for imported manufactures. Rulers fought wars to gain control of the trading centres and built up larger territories for themselves.

The colonial government made treaties of friendship with neighbouring rulers and gradually acquired jurisdiction over the coastline. At the period of the European partition of Africa frontiers were delimited with the neighbouring French and Liberian governments, and a British protectorate was proclaimed in 1896 over the area within the frontier lines, though the original colony retained its status. To raise revenue to pay for administration of the protectorate, a hut tax was imposed. The ruling chiefs, who had not been consulted about the protectorate, objected, and a revolt broke out in 1898 under Bai Bureh. It was suppressed by the end of the year. There were no further large-scale armed risings against the British.

In the protectorate the chiefs ruled under the supervision of British district commissioners. Innovation was discouraged, and little was done to extend education. In the colony many Creoles had held senior official posts in the 19th century and looked forward to governing themselves ultimately. After the protectorate was assumed, however, they were gradually removed from office, and the colony and protectorate were governed by British administrators.

African rule

Independence. After World War II the British government gave in to nationalist demands in Sierra Leone, as elsewhere in West Africa. Democratic institutions were hurriedly constituted. The small Creole minority hoped to entrench their rights politically, but the 1951 constitution gave control to the majority. The government elected under it was led by Milton (later Sir Milton) Margai, leader of the Sierra Leone People's Party (SLPP), a predominantly protectorate party.

During the 1950s parliamentary institutions on the British pattern were introduced in stages. The last stage was reached on April 27, 1961, when Sierra Leone became an independent state within the Commonwealth.

The first years of independence were prosperous. Mineral resources (iron ore and diamonds) brought in substantial revenue, much of which was used for development, particularly education. Njala University College was founded in the early 1960s and amalgamated in 1967 with Fourah Bay College as the University of Sierra Leone.

Sir Milton Margai died in 1964 and was succeeded by his brother, Sir Albert Margai, who tried unsuccessfully to create a one-party state. The opposition All-Peoples' Congress (APC), led by Siaka Stevens, won the 1967 general election. But the army intervened and set up a military government under Lieutenant Colonel Andrew Juxon-Smith. After a year the privates and noncommissioned officers mutinied, imprisoned their officers, and restored parliamentary rule under Stevens and the APC.

The subsequent years were stormy, the government regularly imposing states of emergency and executing its political opponents. In 1971 Sierra Leone became a republic with Stevens as executive president. Meanwhile, the economy deteriorated; the supply of iron ore was exhausted and most of the diamonds were smuggled, thus depriving the government of revenue. Stevens' style of government encouraged his supporters to enrich themselves at public expense. Public dissatisfaction grew, led by student protests. Stevens' answer was to introduce one-party rule in 1978.

In 1985 Stevens retired, having chosen the head of the army, Joseph Saidu Momoh, as his successor. Momoh's installation was hailed as the start of a new era. Yet widespread corruption continued. The economy further deteriorated, and the hopes that had greeted his accession to power were disappointed.

For later developments in the history of Sierra Leone, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (C.Fy.)

Togo

The Republic of Togo (République Togolaise), a sovereign western African nation, consists of that part of the former German colony of Togoland that was made French-mandated territory after World War I. Situated on the Gulf of Guinea, it has a total area of 21,925 square miles (56,785 square kilometres). From its 32-mile (51-kilometre) coastline, Togo extends northward for about 320 miles between Ghana to the west and Benin to the east to its boundary with Burkina Faso in the north. Lomé, the capital, is the largest city and port.

PHYSICAL AND HUMAN GEOGRAPHY

The land. *Relief, drainage, and soils.* Togo consists of six geographic regions. The low-lying, sandy beaches of the narrow coastal region are backed by tidal flats and shallow lagoons, the largest of which is Lake Togo. Beyond the coast lies the Ouatchi Plateau, which stretches about 20 miles inland at an altitude of some 200 to 300 feet (60 to 90 metres). This is the region of the so-called *terre de barre,* a lateritic (reddish, leached, iron-bearing) soil.

Togo's six geographic regions

Northeast of the plateau is a tableland, the highest altitudes reaching 1,300 to 1,500 feet. This region is drained by the Mono River and its tributaries, including the Ogou, and other smaller rivers. West and southwest of the tableland the terrain gradually rises toward the Togo Mountains, which run across central Togo from the south-southwest to the north-northeast. Part of a chain that begins in the Atakora Mountains of Benin, the range ends in the Akwapim Hills of Ghana. Mount Baumann (Agou), which rises to about 3,235 feet (986 metres), is the highest mountain in Togo. Beyond the Togo Mountains to the north lies the Oti River sandstone plateau. This is a savanna region drained by the Oti River, one of the main tributaries of the Volta. To the far northwest is a higher region of granite and gneiss; the cliffs of Dapaong (Dapango) are located in this region.

Climate. Togo has a tropical climate. In the south the rainy seasons occur from mid-April through June and from mid-September through October. The narrow coastal zone, which receives about 35 inches (890 millimetres) of rain annually, is the driest region. The region of Kpalimé (Palimé), about 65 miles inland, receives the highest amount of rain—about 70 inches annually. The north has only one rainy season with an average rainfall of about 45 inches, mostly falling from June to the end of September; during the rest of the year the warm, dry harmattan (a dust-laden wind) predominates. Mean annual temperatures vary from 79° F (26° C) along the coast and in the mountains to 82° F (28° C) on the northern plateau. Daily minimum temperatures of about 68° F (20° C) are recorded in the mountains in August. Daily maxima of about 100° F (38° C) occur in the north during March and April at the end of the long dry season.

Plant and animal life. Savanna-type vegetation is predominant in Togo. On the southern plateaus large trees,

including the baobab, are common, but they are rare in the north. The southwestern highland regions are covered with tropical forests, also found along the river valleys. The coastal zone is dotted with mangrove and reed swamps.

Wild animals are not found in great numbers, especially in the southern and central regions. A few lions, leopards, and elephants can be seen in the north. Monkeys, snakes, and lizards are numerous in many areas, and crocodiles and hippopotamuses abound in the rivers. In the Keran Forest Reserve near Sansanné-Mango in the north, there are wild herds of buffalo, asses, warthogs, antelope, and deer. Numerous species of birds and insects are found in the country. Fish caught off the coast include mackerel, bass, seabream, red snapper, triggerfish, dorado, ray, and sole, while crustaceans include shrimp and lobster.

Settlement patterns. The majority of Togo's population live in small villages scattered throughout the rural areas. A common sight along the coast is the rectangular houses built either of clay and timber or of coconut or palm branches and topped by double-eaved thatched roofs. Scattered throughout the coconut plantations, they are not far from the beaches. Inland in the south, thatched rectangular huts made of adobe are clustered around big trees and surrounded by earthen walls or fences made of palm branches. In the north, the traditional adobe or stone huts are circular and are topped by conical roofs or thatched turrets. They are usually gathered in units corresponding to family groups; often enclosed by earthen walls, they are sometimes interlinked. Distinctive of the northern Kara region is the high density of villages that stretch along the highway or climb up the slopes of the many hills.

Rural settlement

Lomé, the largest urban centre, is spread along the coast. At its centre, there is a mixture of old and new commercial and administrative buildings. The traditional housing unit is the big, walled compound composed of a group of isolated rooms, each opening onto a courtyard.

Aného (Anécho), another coastal town, was once the country's leading European trade centre but is now declining. Other main towns include Tsévié and Tabligbo in the lowland plateau; Kpalimé, Atakpamé, Sokodé, Bassar (Bassari), and Kara (Lama-Kara) at the base of the Togo Mountains; and Sansanné-Mango (Mango) and Dapaong in the far north.

The people. The population of Togo comprises about 30 ethnic groups, many of whom are immigrants from other parts of western Africa. The groups indigenous to Togo live in the north and southwest. The northern groups include the following Gur-speaking Voltaic peoples: the Gurma; the Natemba, Dye, Bu-Bankam, Bu-Kombong, and Konkomba; the Tamberma; the Basari; the Moba; the Naudemba (Losso); the Kabre and Logba; and the Namba (Lamba); a small number of West Atlantic-speaking Fulani; and the Kebu (Akebu). In the southwest the indigenous Kwa peoples also belonging to the central Togo group are the Akposo, the Adele, and the Ahlo.

The immigrants came from east, west, and north. The Ewe, who emigrated from Nigeria between the 14th and 16th century, form the major ethnic group. There are also some scattered Yoruba, mainly Ana. Groups who emigrated from present-day Ghana and Côte d'Ivoire since the 17th century include the Ane (or Mina), the Ga-Adangme, the Kpelle and the Anyana, the Chakossi, and the Dagomba. The northern groups of the Tem (Kotokoli and Temba), Gurma, and Mossi came from the north, mainly from areas in Burkina Faso.

The Ewe immigrants

Most of the nation's non-Africans live in Lomé. Mainly French, they include a few mulattoes of Brazilian, German, and French ancestry. Brazilians, or Portuguese of Brazilian birth, constituted the original trading settlement in Togo, and today Brazilian mulattoes are closely associated with economic and political development.

Although Christianity has profoundly marked the country, more than half the population still adhere to traditional animistic beliefs and in the south participate in *voudou* (voodoo) cults. The main Protestant (Calvinistic) church has been governed for a long time by Togolese moderators. Since independence, the Roman Catholic church in Togo has been headed by a Togolese archbishop. There is also a growing Islāmic population.

The economy. To encourage private investment, the Investment Code of 1965 guaranteed foreign investors the right of freely transferring abroad all investment capital and income. The code also provided for tax benefits for priority enterprises. The trend in the 1970s of direct state involvement in the economy changed in the early 1980s to a pattern of offering incentives for foreign investment and privatization of state enterprises.

Indirect taxes, almost entirely on imports and exports, account for most of the government's ordinary budget revenues. Direct taxes consist of an income tax, a progressive tax on all profits, taxes on wages paid by employers, a tax on rental values and land, and head taxes.

Unlike other former French territories, Togo has not extended preferential trade treatment to France and subsequently to the European Economic Community. This open-door, nondiscriminatory trade policy—together with the expanded production of phosphate and tropical produce—has contributed to the development of the economy.

The open-door trade policy

Resources. Phosphate is the major mineral resource and by far the country's leading export item. The deposits at Hahoetoé and Kpogamé, directly northeast of Lomé, are mined by the government's Togolese Office of Phosphates. Togo is one of the world's largest phosphate producers. Marble is quarried by Sotoma (Société Togolaise de Marbres et de Matériaux), a mixed-economy company with shares held by the Togolese government and an Italian firm. Togo's considerable limestone reserves, also mined near Lomé, are utilized primarily for cement production.

Other mineral resources with commercial potential include iron ore, bauxite, uranium, chromite (an oxide of iron and chromium), gold, diamonds, rutile (titanium dioxide), manganese oxide, and kaolin (china clay). While the iron ore reserves are large, the metal content is only slightly more than 50 percent. The bauxite has a low mineral content.

Agriculture, forestry, and fishing. The variety of soils and climates enables Togo to grow a wide range of products. Export crops include cocoa beans, coffee, shea nuts, cotton, and palm kernels; staple crops are corn (maize), cassava, rice, yams, sorghum, millet, and peanuts (groundnuts).

Cattle, sheep, and pigs are raised in the plateau region and the north. Fishing is carried out on the coast and in the well-stocked inland rivers and ponds. Most of the catch is consumed locally. Forests, which cover about one-fourth of Togo's total area, are a source of tropical hardwoods and other products.

A government agency, the Office of Agricultural Products of Togo, has a monopoly on the foreign sale of Togolese products. Export sales are made by local firms in Paris and London, acting as agents of the Office.

Industry. Mining and quarrying dominate industry in Togo. Manufacturing in the past centred on the processing of agricultural commodities and the import substitution of consumer goods (textiles, footwear, beverages, and tires). In the late 1970s and early 1980s, however, major investments in heavy industrial schemes included a cement plant, a petroleum refinery, a steelworks, and a phosphoric acid plant, but some of these have since closed down.

Trade. Imports include machinery, transport equipment, food, construction materials, pharmaceuticals, and paper products. Low customs duties have encouraged significant smuggling of imported consumer goods to neighbouring countries with higher tariffs, especially Ghana. Besides phosphate and agricultural products, some refined petroleum and cement are exported. Togo's main trading partners are France, The Netherlands, and West Germany.

Transportation. The three main road systems are the scenic coastal road between Ghana and Benin; the road from Lomé north to Burkina Faso; and roads serving the cocoa- and coffee-producing area of Kpalimé, Badou, and Atakpamé.

The government-owned national railway consists of four lines, all of which emanate from Lomé. One line connects Kpalimé with the capital; other lines run to Aného, Tabligbo, and Blitta.

Lomé is Togo's principal port. Its artificial harbour was inaugurated in 1968. A second port at Kpémé, about 22

miles northeast of Lomé, is used exclusively to handle phosphate shipments.

The international airport at Lomé links Togo with European and other African countries. A second international airport at Niamtougou in the north opened in the early 1980s. There are local airports in Atakpamé, Sokodé, Sansanné-Mango, and Dapaong.

Administration and social conditions. *Government.* The coup d'état of 1967 abolished the constitution of 1963 and dissolved the National Assembly. A new constitution approved in 1979 created one-party rule with power vested in the president, who is elected by universal adult suffrage for a seven-year term and is eligible for reelection. The president nominates the ministers and can dissolve the National Assembly. Members of the National Assembly are elected for five-year terms.

Local government

The country is divided into five *régions*—Maritime, Plateaux, Centrale, Kara, and Savanes—for the purposes of economic planning. The five *régions* are subdivided into 21 *préfectures,* each of which is headed by a district chief assisted by a district council. Seven *communes* have been established—for the cities of Aného, Atakpamé, Bassar, Lomé, Kpalimé, Sokodé, and Tsévié, respectively.

The administrative apparatus is complemented by traditional authorities, which include tribal kings or chiefs, village chiefs, and heads of family groups. These traditional authorities play a role in the judicial system, dealing with certain questions of customary law. The judicial system, headed by a Supreme Court, consists of a number of law courts in which civil, commercial, administrative, and criminal cases are heard.

Education. Education is modeled after the French system. Togolese teachers, who have replaced French personnel to a large extent, are expected to adapt the system to the Togolese context. Primary and secondary education is provided by public or parochial schools.

The University of Benin at Lomé (founded in 1970) has schools of humanities and science, and a university institute of technology. A school of architecture and town planning, also at Lomé, was founded in 1975 by the African and Mauritian Common Organization (OCAM).

Traditional culture

Cultural life. Like other African peoples, the Togolese have a strong oral tradition. Little has been done, however, to promote vernacular literature. Before independence there were a few Togolese writers using French. Since independence, regional (especially Ewe) literature emerged with the works of several novelists and playwrights. Founded in 1967, the African Ballet of Togo has aimed at popularizing the finest traditional dances of the various ethnic groups.

For statistical data on the land and people of Togo, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL. (M.K.P./S.De.)

HISTORY

Until 1884 Togoland was an indeterminate buffer zone between the warring states of Asante and Dahomey. The only port was Petit Popo (Anécho, or Aného). Throughout the 18th century the Togo part of the Slave Coast was held by the Danes.

German occupation. German missionaries arrived in Ewe territory in 1847, and German traders were soon established at Anécho. In 1884 Gustav Nachtigal, sent by the German government, induced a number of coastal chiefs to accept German protection. The protectorate was recognized in 1885, and its coastal frontiers with Dahomey and the Gold Coast were defined by treaties with France and Great Britain. German military expeditions (1888–97) met with little resistance, securing a hinterland the boundaries of which also were determined by treaties with France (1897) and Great Britain (1899).

Lomé, at the western end of the coast, was selected as the colonial capital in 1897, a modern town was laid out, and in 1904 a jetty was built. Three railways were constructed to open up the interior. Exploitation was confined to the coastal and central areas and was exclusively agricultural. Plantations were established both by the government and by private German corporations, but crop development was left mainly to the Togolese, assisted by agriculturists trained at a college in Nuatja (Notsé). Upwardly-mobile Ewe were recruited into what was supposed to be Germany's *Musterkolonie* (model colony). Trade was chiefly in palm products, rubber, cotton, and cocoa. German administration was efficient but marred by its harsh treatment of Africans and use of forced labour.

On Aug. 7, 1914, at the outset of World War I, British and French colonial troops from the Gold Coast and Dahomey invaded Togoland and on August 26 secured the unconditional surrender of the Germans. Thereafter the western part of the colony was administered by Britain, the eastern part by France. By an Anglo-French agreement of July 10, 1919, France secured the railway system and the whole coastline. After Germany renounced its sovereignty in the Treaty of Versailles, the League of Nations in 1922 issued mandates to Britain and France for the administration of their spheres.

League of Nations mandate. The northern part of the British-mandated territory was administered with the Northern Territories of the Gold Coast, the southern part with the Gold Coast Colony. Although the British administration built roads connecting its sphere with the road system of the Gold Coast, the bulk of the territory's external trade passed over the railways of French Togo.

French Togo was administered by a commissioner assisted by a consultative executive council. When British Togo was attached to the Gold Coast, French Togo was formed into a distinct unit until 1934, when a kind of economic union was established with Dahomey; this was replaced in 1936 by a qualified integration with French West Africa that lasted 10 years. Agricultural development was pursued, and a planned settlement of the interior by the Kabre and other peoples was carried out. Peanut growing was introduced in the northern areas, and energetic action was taken against sleeping sickness.

After World War II French Togo sent a deputy to the French National Assembly, a counselor to the Assembly of the French Union, and two senators to the Council of the Republic. A representative assembly was concerned with internal affairs.

United Nations trusteeship. In 1946 the British and French governments placed their spheres of Togoland under UN trusteeship. After 1947 the Ewe people in southern Togoland represented to the Trusteeship Council that either their territories or the whole of Togoland should be brought under a common administration. These proposals were difficult to implement because Ewe also inhabited the southeastern part of the Gold Coast Colony and because not all the people of southern Togoland were Ewe. The British colony was also rapidly advancing toward self-government, and the incorporation of the northern part of the British sphere with the Northern Territories of the Gold Coast had reunited the Dagomba and Mamprusi kingdoms, both of which had been cut in two by the pre-1914 boundary. Following a plebiscite held under UN auspices on May 9, 1956, the British trust territory of Togoland was on December 13 incorporated into the Gold Coast (although in the southern districts of Ho and Kpandu the Ewe vote showed a two-to-one majority in favour of continued British trusteeship). The Gold Coast and Togoland together were renamed Ghana and achieved independence in 1957.

Independence. French Togoland became an autonomous republic within the French Union on Aug. 30, 1956. This status was confirmed (despite Ewe opposition) by a plebiscite held in October under French auspices. Nicolas Grunitzky was appointed premier. Following UN representations, elections in April 1958 favoured complete independence and rejected Grunitzky's Togolese Progress Party in favour of Sylvanus Olympio's Togolese National Unity Party. Togo became independent on April 27, 1960.

After the 1961 elections, which established a presidential form of government, Olympio became the first president. He maintained economic cooperation with France. Togo became a member of the Organization of African Unity (OAU) in 1963 and in 1965 subscribed to the renewed Joint African and Malagasy Organization, which provided for economic, political, and social cooperation among French-speaking African states. (H.J.D./S.De.)

Togo's relations with Ghana

Ghanaian pressure for the integration of Togo with Ghana was resisted by the Togolese and led to strained relations between the two republics, including a trade embargo imposed by Ghana. Olympio's increasingly harsh rule and policy of fiscal austerity came to an end on Jan. 13, 1963. Having rejected petitions to integrate into the national army Togolese noncommissioned officers recently demobilized from France's colonial armies, he was shot at the gates of the U.S. Embassy (while seeking sanctuary) by Sergeant Étienne Gnassingbé Eyadema (later called Gnassingbé Eyadema). Grunitzky was invited to return from exile and assume the presidency, and he was confirmed in office in subsequent elections that also created a new constitution and legislature. Most of the noncommissioned officers were integrated into an expanded army—many, including Eyadema, as commissioned officers.

Cabinet infighting, aggravated in the south by Ewe feelings that with Olympio's assassination they had lost power to Grunitzky's largely pro-northern administration, led to chronic instability. On Jan. 13, 1967, Eyadema, then a lieutenant colonel and chief of staff, once again seized power and dissolved all political parties. Though relying primarily on the support of his kinsmen in the north and the largely northern-staffed army, Eyadema's rule was stabilized by a number of other factors. Phosphate exports dramatically alleviated the economic picture, allowing the regime to satisfy regional and ethnic interests and to begin the first serious effort at transforming the countryside. In 1974 the phosphate industry was nationalized, generating increased state revenues. Meticulous ethnic balancing of the Cabinet and an open-door economic policy further attracted support from prospering traders (and smugglers into Ghana), and by 1972 Eyadema felt secure enough to seek popular legitimation via a presidential plebiscite. On Dec. 30, 1979, the first legislative elections since 1967 were held under a new constitution that formally civilianized Togo under one-party rule headed by President Eyadema and the Rally of the Togolese People. Legislative elections were held again in 1985, and Eyadema was elected to a second seven-year term the following year.

The regime has not been without its opponents, however. Most of these have been Ewe from the south (including the self-exiled sons of Olympio) rebelling against the northerner Eyadema and the cult of personality that has progressively surrounded him. The opposition has sponsored conspiracies to topple Eyadema and a number of bombings in Lomé.

The 1980s also saw an economic downturn that began to undermine the regime's patronage base and, by extension, its stability. Falling global market prices for phosphates led to sharply lower state revenues; at the same time, growing corruption and massive expenditures on the bloated civil service and inefficient public enterprises strained the fiscal resources of the state. By 1982 Togo was forced to limit government spending. In that year the costly parastatal sector began to be dismantled or privatized, and in the next few years the country's heavy national debt was often rescheduled.

For later developments in the history of Togo, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL. (S.De.)

BIBLIOGRAPHY

Physical and human geography: Profiles of the western African countries are found in *Africa Contemporary Record* (annual); *West Africa Annual;* and DONALD GEORGE MORRISON, ROBERT CAMERON MITCHELL, and JOHN N. PADEN, *Black Africa: A Comparative Handbook,* 2nd ed. (1989). The physical environment is explored in REUBEN K. UDO, *A Comprehensive Geography of West Africa* (1978), and *The Human Geography of Tropical Africa* (1982); R.J. HARRISON CHURCH, *West Africa,* 8th ed. (1980), an environmental study; J.B. WRIGHT (ed.), *Geology and Mineral Resources of West Africa* (1985); BRIAN HOPKINS, *Forest and Savanna,* 2nd ed. (1974), on West African tropical ecology; and F. WHITE, *Vegetation of Africa* (1983). RICHARD W. FRANKE and BARBARA H. CHASIN, *Seeds of Famine* (1980), is a study of the Sahelian famine of 1969–74. JOHN C. CALDWELL (ed.), *Population Growth and Socio-economic Change in West Africa* (1975), surveys demographic conditions by region. P.H. ADY (ed.), *Africa* (1965), is useful.

Migration and urbanization are examined in SAMIR AMIN (ed.), *Modern Migrations in Western Africa* (1974); JOSEF GUGLER and WILLIAM G. FLANAGAN, *Urbanization and Social Change in West Africa* (1978); and K.C. ZACHARIAH and JULIEN CONDÉ, *Migration in West Africa* (1981).

Economic history and development are analyzed by A.G. HOPKINS, *An Economic History of West Africa* (1973); DOUGLAS RIMMER, *The Economies of West Africa* (1984); and CLAUDE MEILLASSOUX (ed.), *The Development of Indigenous Trade and Markets in West Africa* (1971). See also J.O.C. ONYEMELUKWE and M.O. FILANI, *Economic Geography of West Africa* (1983). Works on agriculture include WILLIAM ALLAN, *The African Husbandman* (1965, reprinted 1977); ROBERT H. BATES, *Markets and States in Tropical Africa* (1981); KEITH HART, *The Political Economy of West African Agriculture* (1982); JONATHAN BARKER (ed.), *The Politics of Agriculture in Tropical Africa* (1984); and PAUL RICHARDS, *Indigenous Agricultural Revolution* (1985).

Traditional cultures: Overviews of African culture are provided by GEORGE MURDOCK, *Africa: Its Peoples and Their Culture History* (1959), still recommended; and JOCELYN MURRAY (ed.), *Cultural Atlas of Africa* (1981). Survey articles on the major ethnic groups may be found in GEORGES BALANDIER and JACQUES MAQUET (eds.), *Dictionary of Black African Civilization* (1974; originally published in French, 1968).

(*Western Sudan cultures*): A summary of Wolof and Serer ethnography is given by DAVID P. GAMBLE, *The Wolof of Senegambia, Together with Notes on the Lebu and the Serer* (1957, reprinted 1967). PAUL PÉLISSIER, *Les Paysans du Sénégal* (1966), describes Wolof, Serer, Malinke, and Diola systems of agriculture. On the Mossi, see PETER B. HAMMOND, *Yatenga: Technology in the Culture of a West African Kingdom* (1966); and SUZANNE LALLEMAND, *Une Famille mossi* (1977). Sources for the Fulani include DERRICK J. STENNING, *Savannah Nomads* (1959); and MARGUERITE DUPIRE, *Peuls nomades: étude descriptive des Wodaabe du Sahel nigérien* (1962). Two studies of ethnic culture in urban areas are ENID SCHILDKROUT, *People of the Zongo* (1978); and JOHN N. PADEN, *Religion and Political Culture in Kano* (1973), a study of the Hausa and Fulani in Kano, Nigeria. PATRICK R. MCNAUGHTON, *The Mande Blacksmiths* (1988), describes artistic, social, and spiritual realms of Mande smiths. Other works include BARBARA J. CALLAWAY, *Muslim Hausa Women in Nigeria* (1987); and CAROL BECKWITH and MARION VAN OFFELEN, *Nomads of Niger* (1983), a pictorial study.

(*Guinea Coast cultures*): Ethnographic studies include WILLIAM TORDOFF, *Ashanti Under the Prempehs, 1888–1935* (1965); IVOR WILKS, *Asante in the Nineteenth Century* (1975); DENNIS M. WARREN, *The Akan of Ghana,* rev. ed. (1986); LEO SPITZER, *The Creoles of Sierra Leone: Responses to Colonialism, 1870–1945* (1974); J.S. EADES, *The Yoruba Today* (1980); ROSEMARY L. HARRIS, *The Political Organization of the Mbembe, Nigeria* (1965); MICHAEL JACKSON, *The Kuranko* (1977); J.D.Y. PEEL, *Ijeshas and Nigerians: The Incorporation of a Yoruba Kingdom, 1890s–1970s* (1983); MICHEL VERDON, *The Abutia Ewe of West Africa* (1983); DARYLL FORDE, *Yako Studies* (1964); and VICTOR C. UCHENDU, *The Igbo of Southeast Nigeria* (1965).

(M.B.T./D.P.Ga./R.L.Ha.)

History: Much original work is available only in articles in such journals as *Journal of African History* (3/yr.); *The International Journal of African Historical Studies* (quarterly); *History in Africa* (annual); *African Affairs* (quarterly); and *The Journal of Modern African Studies* (quarterly). The standard work is J.F. ADE AJAYI and MICHAEL CROWDER (eds.), *History of West Africa,* 2nd ed., 2 vol. (1976–87). In addition there are the western African chapters with comprehensive bibliographies and bibliographical essays in J.D. FAGE and ROLAND OLIVER (eds.), *The Cambridge History of Africa,* 8 vol. (1975–86); and there are some stimulating contributions in UNESCO INTERNATIONAL SCIENTIFIC COMMITTEE FOR THE DRAFTING OF A GENERAL HISTORY OF AFRICA, *General History of Africa* (1981–), but its bibliographies are less reliable. Useful atlases include J.D. FAGE, *An Atlas of African History,* 2nd ed. (1978); and J.F. ADE AJAYI and MICHAEL CROWDER (eds.), *Historical Atlas of Africa* (1985). A short handbook is J.D. FAGE, *A History of West Africa,* 4th ed. (1969); while ANTHONY ATMORE and GILLIAN STACEY, *Black Kingdoms, Black Peoples* (1979), is a well-illustrated introduction to West African history, peoples, and cultures.

For French-speaking territories, JEAN SURET-CANALE, *French Colonialism in Tropical Africa, 1900–1945* (1971; originally published in French, 1964), contains useful material; and JOHN D. HARGREAVES, *West Africa, the Former French States* (1967), is an excellent short study. RONALD H. CHILCOTE, *Portuguese Africa* (1967), is a useful general work. RAYMOND MAUNY, *Tableau géographique de l'Ouest africain au Moyen Âge* (1961), contains much useful information up to about AD 1400, though it should be supplemented by later research, such as NEHEMIAH LEVTZION, *Ancient Ghana and Mali* (1973, reprinted 1980). Other useful works for the pre-jihad history of the western and central Sudan are the classic works by E.W. BOVILL, *The Golden*

Trade of the Moors, 2nd ed. rev. by ROBIN HALLETT (1968); and CHARLES MONTEIL, *Les Empires du Mali* (1930, reissued 1968).

Three works on the early history of Guinea may be recommended: J.D. FAGE, "Upper and Lower Guinea," ch. 6 in *The Cambridge History of Africa*, vol. 3, pp. 463–518 (1977); FRANK WILLETT, *Ife in the History of West African Sculpture* (1967); and ROBERT SMITH, *Kingdoms of the Yoruba*, 3rd ed. (1988). The coming of the European traders to the Guinea Coast is treated in JOHN W. BLAKE, *West Africa: Quest for God and Gold, 1454–1578,* 2nd ed. rev. and enlarged (1977). PHILIP D. CURTIN, *The Atlantic Slave Trade* (1969), provided an innovative analysis; but the basic data have been reworked in a useful general history, PAUL E. LOVEJOY, *Transformations in Slavery* (1983).

The subsequent history of Guinea has occasioned many important monographs, such as JOHN VOGT, *Portuguese Rule on the Gold Coast, 1469–1682* (1979); WALTER RODNEY, *A History of the Upper Guinea Coast, 1545–1800* (1970, reprinted 1982); A.F.C. RYDER, *Benin and the Europeans, 1485–1897* (1969); RAY A. KEA, *Settlements, Trade, and Polities in the Seventeenth-Century Gold Coast* (1982); PHILIP D. CURTIN, *Economic Change in Precolonial Africa* (1975); DAVID NORTHRUP, *Trade Without Rulers: Pre-colonial Economic Development in Southeastern Nigeria* (1978); and ROBIN LAW, *The Oyo Empire, c. 1600–c. 1836* (1977). PETER B. CLARKE, *West Africa and Islam* (1982), is a useful introductory work, covering the 8th to the 20th century. For the great outburst of Islām in western Africa in the 18th and 19th centuries, see MURRAY LAST, *The Sokoto Caliphate* (1967, reprinted 1977); and DAVID ROBINSON, *The Holy War of Umar Tal: The Western Sudan in the Mid-nineteenth Century* (1985).

The growth of European influence is covered in works such as ROBIN HALLET, *The Penetration of Africa: European Exploration in North and West Africa to 1815* (1965); A. ADU BOAHEN, *Britain, the Sahara, and the Western Sudan, 1788–1861* (1964); PHILIP D. CURTIN, *The Image of Africa: British Ideas and Action, 1780–1850* (1964, reissued 1973); K. ONWUKA DIKE, *Trade and Politics in the Niger Delta, 1830–1885* (1956, reprinted 1981); BERNARD SCHNAPPER, *La Politique et le commerce français dans le Golfe de Guinée, de 1838 à 1871* (1961); and J.F. ADE AJAYI, *Christian Missions in Nigeria, 1841–1891* (1965). For the European partition of western Africa, the standard works are JOHN D. HARGREAVES, *Prelude to the Partition of West Africa* (1963), and *West Africa Partitioned,* 2 vol. (1974–85); but see also HENRI BRUNSCHWIG, *French Colonialism, 1871–1914* (1966; originally published in French, 1960); and A.S. KANYA-FORSTNER, *The Conquest of the Western Sudan* (1969).

For the colonial period, two contemporary studies are invaluable: LORD HAILEY (WILLIAM MALCOLM HAILEY), *An African Survey,* rev. ed. (1957, reprinted 1968); and S. HERBERT FRANKEL, *Capital Investment in Africa* (1938, reprinted 1969). For indirect rule, MARGERY PERHAM, *Native Administration in Nigeria* (1937, reprinted 1962), is a classic. A general synthesis is provided by MICHAEL CROWDER, *West Africa Under Colonial Rule* (1968); see also MICHAEL CROWDER and OBARO IKIME (eds.), *West African Chiefs: Their Changing Status Under Colonial Rule* (1970); and MICHAEL CROWDER (ed.), *West African Resistance: The Military Response to Colonial Occupation,* new ed. (1978). The chapters on western Africa in L.H. GANN and PETER DUIGNAN (eds.), *Colonialism in Africa, 1870–1960,* vol. 1–2 (1969–70), are often extremely valuable. There is much useful material in two collections: PROSSER GIFFORD and WM. ROGER LOUIS (eds.), *Britain and Germany in Africa* (1967), and *France and Britain in Africa* (1971).

Works dealing with the transition to independence include JOHN D. HARGREAVES, *The End of Colonial Rule in West Africa* (1979), and *Decolonization in Africa* (1988); KEN POST, *The New States of West Africa,* rev. ed. (1968); RUTH SCHACHTER MORGENTHAU, *Political Parties in French-speaking West Africa* (1964); PROSSER GIFFORD and WM. ROGER LOUIS (eds.), *The Transfer of Power in Africa: Decolonization, 1940–1960* (1982); and EDWARD MORTIMER, *France and the Africans, 1944–1960* (1969).

General works on the period since independence include overviews in vol. 8 of *The Cambridge History of Africa* (1986). Treatments of politics include JOHN DUNN (ed.), *West African States: Failure and Promise* (1978); and PATRICK CHABAL (ed.), *Political Domination in Africa* (1986). For the Francophone states, a good recent survey is PATRICK MANNING, *Francophone Sub-Saharan Africa, 1880–1985* (1988). No comparable survey exists for the Anglophone states, but see PETER DUIGNAN and ROBERT H. JACKSON (eds.), *Politics and Government in African States, 1960–1985* (1986). An introduction to the Lusophone states is BASIL DAVIDSON, *No Fist Is Big Enough to Hide the Sky* (1981), on Guinea and Cape Verde. For recent events, the weekly *West Africa* is indispensable. (J.D.F./T.C.McC.)

Western Sudan nations. *Burkina Faso:* General works include YVES PÉRON and VICTOIRE ZALACAIN (eds.), *Atlas de la Haute-Volta* (1975); GINETTE PALLIER, *Géographie générale de la Haute-Volta* (1978); NORBERT NIKIÉMA, *La Situation linguistique en Haute-Volta* (1980); and CLAUDETTE SAVONNET-GUYOT, *Etat et sociétés au Burkina* (1986). No general history exists, but DANIEL MILES MCFARLAND, *Historical Dictionary of Upper Volta (Haute Volta)* (1978), provides useful information. Precolonial and conquest history and ethnography are treated in JEAN CAPRON, *Communautés villageoises bwa: Mali, Haute-Volta* (1973); MICHEL IZARD, *Introduction à l'histoire des royaumes mossi,* 2 vol. (1970); PAUL IRWIN, *Liptako Speaks* (1981); and ANNE-MARIE DUPERRAY, *Les Gourounsi de Haute-Volta: conquête et colonisation, 1896–1933* (1984). The colonial and independence periods are covered by JEAN AUDOUIN and RAYMOND DENIEL, *L'Islam en Haute-Volta à l'époque coloniale* (1978); SALFO ALBERT BALIMA, *Genése de la Haute-Volta* (1970); SIDIKI COULIBALY, JOEL GREGORY, and VICTOR PICHÉ, *Les Migrations voltaïques,* vol. 1 (1980); and ELLIOT P. SKINNER, *The Mossi of the Upper Volta* (1964). FRANÇOIS D. BASSOLET, *Évolution de la Haute-Volta, de 1898 au 3 janvier 1966* (1968), is a subjective account of the postindependence era. (M.Ec.)

Cape Verde: ANTÓNIO COSTA, *Cabo Verde,* 2 vol. (1980–81), is an overview of the islands' geography and natural history. COLM FOY, *Cape Verde* (1988), provides information on the postindependence nation. DEIRDRE MEINTEL, *Race, Culture, and Portuguese Colonialism in Cabo Verde* (1984), analyzes social structures and the role of race. ANTÓNIO CARREIRA, *The People of the Cape Verde Islands,* trans. from Portuguese (1982), is a history of emigration. MANUEL FERREIRA, *A aventura crioula,* 3rd ed. rev. (1985), studies cultural life and has an extensive bibliography. CHARLES VERLINDEN, *António de Noli e a colonização das ilhas de Cabo Verde* (1963), discusses the period of discovery and settlement; and T. BENTLEY DUNCAN, *Atlantic Islands* (1972), includes a general history of 17th-century commerce. RICHARD LOBBAN and MARILYN HALTER, *Historical Dictionary of the Republic of Cape Verde,* 2nd ed. (1988), is a useful reference. JOSEPH M. MCCARTHY, *Guinea-Bissau and Cape Verde Islands* (1977), provides a comprehensive bibliography. (C.S.Sh.)

Chad: HAROLD D. NELSON *et al., Area Handbook for Chad* (1972), is still a useful introduction. See also PIERRE HUGOT, *Le Tchad* (1965); MARIO AZEVEDO (ed.), *Cameroon and Chad in Historical and Contemporary Perspectives* (1989); and JEAN CABOT and CHRISTIAN BOUQUET, *Atlas pratique du Tchad* (1972). Ethnographic studies include JEAN CHAPELLE, *Nomades noirs du Sahara* (1958, reissued 1982), the authoritative study on the Teda; ALBERT LE ROUVREUR, *Sahéliens et sahariens du Tchad* (1962), information about the northern and eastern populations; and JEAN CHAPELLE, *Le Peuple tchadien* (1980). Politics and government are treated in GEORGES DIGUIMBAYE and ROBERT LANGUE, *L'Essor du Tchad* (1969); and MICHAEL P. KELLEY, *A State in Disarray: Conditions of Chad's Survival* (1986), discussing dependency. SAMUEL DECALO, *Historical Dictionary of Chad,* 2nd ed. (1987), has an extensive bibliography. Early history is discussed in JEAN PAUL LEBEUF, *Archéologie tchadienne: les Sao du Cameroun et du Tchad* (1962); and JEAN PAUL LEBEUF and A. MASSON DETOURBET, *La Civilisation du Tchad* (1950). JACQUES LE CORNEC, *Histoire politique du Tchad, de 1900 à 1962* (1963), provides information on the colonial era and the first years of independence. ROBERT BUIJTENHUIJS, *Le Frolinat et les révoltes populaires du Tchad, 1965–76* (1978), and *Le Frolinat et les guerres civiles du Tchad (1977–1984)* (1987), are the best sources of information on Chad's civil wars. The boundary disputes are discussed in BERNARD LANNE, *Tchad-Libye,* 2nd ed. (1986). (A.T.G.)

The Gambia: OYEYEMI HAFFNER, *A New Geography of Senegambia* (1981); and HUBERT DESCHAMPS, *Le Sénégal et la Gambie,* 3rd ed. updated (1975), are useful introductions. Historical information may be found in J.M. GRAY, *A History of the Gambia* (1940, reissued 1966); HARRY A. GAILEY, *A History of The Gambia* (1964), and *Historical Dictionary of The Gambia,* 2nd ed. (1987); CHARLOTTE A. QUINN, *Mandingo Kingdoms of the Senegambia* (1972); and S.A. BAKARR, *The Gambia Yesterday, 1447–1979* (1980), a historical chronology. For bibliographic information, see DAVID P. GAMBLE (comp.), *The Gambia* (1988). (H.A.G.)

Mali: PHILIPPE DECRAENE, *Le Mali* (1980); and MAMADOU TRAORÉ and YVES MONNIER (eds.), *Atlas du Mali* (1980), provide overviews. Ethnographic studies include JEAN GALLAIS, *Hommes du Sahel: espaces-temps et pouvoirs: le delta intérieur du Niger, 1960–1980* (1984); DANIELLE JONCKERS, *La Société minyanka du Mali* (1987); and BOKAR N'DIAYÉ, *Groupes ethniques au Mali* (1970). MOUSSA COLA CISSÉ *et al., Le Mali, le paysan et l'état* (1981); and EMIL SCHREYGER, *L'Office du Niger au Mali, 1932 à 1982* (1984), focus on agriculture. GUÉDEL NDIAYE, *L'Échec de la Fédération du Mali* (1980), is a political study. Historical works include PASCAL JAMES IMPERATO, *Historical Dictionary of Mali,* 2nd ed. (1986); ATTILIO GAUDIO, *Le Mali* (1988); AMIDU MAGASA, *Papa-commandant a jeté un grand filet devant nous: les exploités des rives du Niger, 1900–*

1962 (1978), on forced labour and colonization; WILLIAM J. FOLTZ, *From French West Africa to the Mali Federation* (1965), a political history; and CHEICK OUMAR DIARRAH, *Le Mali de Modibo Keïta* (1986). (K.M.B./P.J.I.)

Mauritania: Introductions include ALFRED G. GERTEINY, *Mauritania* (1967); BRIAN DEAN CURRAN and JOANN SCHROCK, *Area Handbook for Mauritania* (1972); CHARLES TOUPET and JEAN-ROBERT PITTE, *La Mauritanie* (1977); and CHARLES TOUPET and GEORGES LACLAVÈRE, *Atlas de la République Islamique de Mauritanie* (1977). CHARLES TOUPET, *La Sédentarisation des nomades en Mauritanie centrale sahélienne* (1977), studies geographic and social transformations. The economy is treated in RICHARD M. WESTEBBE, *The Economy of Mauritania* (1971); and JÉRÔME PUJOS, *Croissance économique et impulsion extérieure: étude sur l'économie mauritanienne* (1964). ALFRED G. GERTEINY, *Historical Dictionary of Mauritania* (1981), is a good reference. Specific periods are examined in C.C. STEWART and E.K. STEWART, *Islam and Social Order in Mauritania: A Case Study from the Nineteenth Century* (1973); FRANCIS DE CHASSEY, *Mauritanie, 1900–1975* (1978, reissued 1984); and GENEVIÈVE M. DÉSIRÉ-VUILLEMIN, *Contribution à l'histoire de la Mauritanie de 1900 à 1934* (1962). (A.G.G.)

Niger: AROUNA HAMIDOU SIDIKOU and PHILIPPE CLAUDE CHAMARD, *Géographie du Niger* (1976); and PIERRE DONAINT and FRANÇOIS LANCRENON, *Le Niger,* 3rd ed. rev. and updated (1984), are general works. Ethnographic studies include CATHERINE BAROIN, *Anarchie et cohésion sociale chez les Toubo* (1985); ANGELO MALIKI BONFIGLIOLI, *Dudal: histoire de famille et histoire de troupeau chez un groupe de Wodaabe du Niger* (1989); PETER FUCHS, *Das Brot der Wüste: Sozio-ökonomie der Sahara-Kanuri von Fachi* (1983); HENRI GUILLAUME, *Les Nomades interrompus: introduction à l'étude du canton twareg de l'Imanan* (1974); JOHANNES NICOLAISEN, *Ecology and Culture of the Pastoral Tuareg* (1963); GUY NICOLAS, *Dynamique sociale et appréhension du monde au sein d'une société hausa* (1975); JEAN-PIERRE OLIVIER DE SARDAN, *Système des relations économiques et sociales chez les Wogo, Niger* (1969); MARC HENRI PIAULT, *Histoire Mawri: introduction à l'étude des processus constitutifs d'un état* (1970); and MAÏKOREMA ZAKARI, *Contribution à l'histoire des populations du sud-est nigérien: le cas du Mangari (XVIe–XIXe s.)* (1985). Two city studies are SUZANNE BERNUS, *Niamey: population et habitat* (1962); and EDMOND BERNUS and SUZANNE BERNUS, *Du Sel et des dattes: introduction a l'étude de la communauté d'In Gall et de Tegidda-n-tesamt* (1972). Historical treatments include EDMOND SÉRÉ DE RIVIÈRES, *Histoire du Niger* (1965); FINN FUGLESTAD, *A History of Niger, 1850–1960* (1983); STEPHEN BAIER, *An Economic History of Central Niger* (1980); and SAMUEL DECALO, *Historical Dictionary of Niger,* 2nd ed. (1987). (F.F.)

Senegal: Overviews are found in HAROLD D. NELSON *et al., Area Handbook for Senegal,* 2nd ed. (1974); HUBERT DESCHAMPS, *Le Sénégal et la Gambie,* 3rd ed. updated (1975); and PHILIPPE DECRAENE, *Le Sénégal* (1985). See also *Atlas national du Sénégal* (1977); and PAUL PÉLISSIER (ed.), *Atlas du Sénégal* (1980). Sociological works include GILLES BLANCHET, *Elites et changements en Afrique et au Sénégal* (1983); and ABDOULAYE-BARA DIOP, *La Famille wolof* (1985). DONAL B. CRUISE O'BRIEN, *The Mourides of Senegal* (1971); CHRISTIAN COULON, *Le Marabout et le Prince: Islam et pouvoir au Sénégal* (1981); and MORIBA MAGASSOUBA, *L'Islam au Sénégal* (1985), discuss the Islāmic heritage and present religiopolitical conflicts. Politics and government are studied by MICHAEL CROWDER, *Senegal: A Study of French Assimilation Policy,* rev. ed. (1967); and MAR FALL, *Sénégal, l'état Abdou Diouf* (1986). Historical works include ERIC MAKÉDONSKY, *Le Sénégal: la Sénégambie,* 2 vol. (1987); PHILIP D. CURTIN, *Economic Change in Precolonial Africa* (1975); and SHELDON GELLAR, *Senegal* (1982), on the period since independence. G. WESLEY JOHNSON, JR., *The Emergence of Black Politics in Senegal: The Struggle for Power in the Four Communes, 1900–1920* (1971), is supplemented by KENNETH ROBINSON, "Senegal: The Elections to the Territorial Assembly, March 1957," in W.J.M. MACKENZIE and KENNETH ROBINSON, *Five Elections in Africa* (1960), pp. 281–390. LUCIE GALLISTEL COLVIN, *Historical Dictionary of Senegal* (1981), is a useful reference. (Ca.C./J.D.H.)

Guinea Coast nations. *Benin:* Ethnographic studies include JACQUES LOMBARD, *Structures de type "féodal" en Afrique noire* (1965); WILLIAM J. ARGYLE, *The Fon of Dahomey* (1966); and PAUL MERCIER, *Tradition, changement, histoire: les "Somba" du Dahomey septentrional* (1968). PATRICK MANNING, *Slavery, Colonialism, and Economic Growth in Dahomey, 1640–1960* (1982), is a detailed study. ROBERT CORNEVIN, *La République Populaire du Bénin: des origines dahoméennes à nos jours* (1981), is an indispensable reference. For a treatment of the traditional Dahomey kingdom, see MELVILLE J. HERSKOVITS, *Dahomey: An Ancient West African Kingdom,* 2 vol. (1938, reprinted 1967). DOV RONEN, *Dahomey* (1975), is the only general history of the independent country available in English. I.A. AKINJOGBIN, *Dahomey and Its Neighbours, 1708–1818* (1967), is the principal published study of the precolonial kingdom of Dahomey. C.W. NEWBURY, *The Western Slave Coast and Its Rulers* (1961, reprinted 1983), is a study of the development of European commerce and imperialism in the 19th century. SAMUEL DECALO, *Historical Dictionary of Benin,* 2nd ed. (1987), has an extensive bibliography. (D.Ro./R.La.)

Cameroon: Descriptions of the physical and cultural environment are given in AARON S. NEBA, *Modern Geography of the Republic of Cameroon,* 2nd ed. (1987); and J.A. NGWA, *A New Geography of Cameroon,* new ed. (1978). MICHAEL G. SCHATZBERG and I. WILLIAM ZARTMAN (eds.), *The Political Economy of Cameroon* (1986), is a collection of essays. VICTOR JULIUS NGOH, *Cameroon, 1884–1985* (1987); and VICTOR T. LE VINE and ROGER P. NYE, *Historical Dictionary of Cameroon* (1974), provide introductions. DAVID E. GARDINER, *Cameroon: United Nations Challenge to French Policy* (1963); and VICTOR T. LE VINE, *The Cameroons, from Mandate to Independence* (1964, reprinted 1977), are the best accounts up to the time of reunification. The history of the country since 1961 is examined in WILLARD R. JOHNSON, *The Cameroon Federation* (1970); NEVILLE RUBIN, *Cameroun: An African Federation* (1971); VICTOR T. LE VINE, *The Cameroon Federal Republic* (1971); and MARIO AZEVEDO (ed.), *Cameroon and Chad in Historical and Contemporary Perspectives* (1989). MARK W. DELANCEY and PETER J. SCHRAEDER (comps.), *Cameroon* (1986), provides bibliographic information. (G.Be./Ed.)

Côte d'Ivoire: T.D. ROBERTS *et al., Area Handbook for Ivory Coast,* 2nd ed. (1973), provides a general overview. RAYMOND BORREMANS, *Le Grand Dictionnaire encyclopédique de la Côte d'Ivoire* (1986–), is more current. See also PIERRE VENNETIER (ed.), *Atlas de la Côte d'Ivoire,* 2nd ed. rev. and updated by PIERRE VENNETIER and GENEVIÈVE DAVERAT (1983). Ethnographic studies include ENID SCHILDKROUT (ed.), *The Golden Stool: Studies of the Asante Center and Periphery* (1987); and IVOR WILKS, *Wa and the Wala* (1989). LAURENT GBAGBO, *La Côte-d'Ivoire* (1982), is a history of economic and social developments in the 20 years before independence. For further discussions of economic history and policies, see BASTIAAN A. DEN TUINDER, *Ivory Coast* (1978); GEORGES LORY, *Introduction à l'économie ivoirienne* (1981); I. WILLIAM ZARTMAN and CHRISTOPHER DELGADO, *The Political Economy of Ivory Coast* (1984); and JEAN-PIERRE FOIRRY and DENIS REQUIER-DESJARDINS, *Planification et politique économique en Côte d'Ivoire, 1960–1985* (1986). ARISTIDE R. ZOLBERG, *One-Party Government in the Ivory Coast,* rev. ed. (1969), is a study of political history. Sociological works include ARTHUR CONTE, *Côte d'Ivoire, ou, les racines de la sagesse* (1981); ABDOU TOURÉ, *La Civilisation quotidienne en Côte-d'Ivoire* (1981); and Y.-A. FAURÉ and J.-F. MEDARD (comps.), *État et bourgeoisie en Côte-d'Ivoire* (1982). ROBERT J. MUNDT, *Historical Dictionary of the Ivory Coast (Cote d'Ivoire)* (1987), a valuable reference work, has an extensive bibliography. JEAN-NOËL LOUCOU, *Histoire de la Côte d'Ivoire,* vol. 1, *La Formation des peuples* (1984), presents the origin and development of the country's major ethnic groups. F.J. AMON D'ABY, *La Côte d'Ivoire dans la cité africaine* (1951), is a pioneering and still valuable survey. (J.Co./N.E.L.)

Equatorial Guinea: MAX LINIGER-GOUMAZ, *Small Is Not Always Beautiful* (1988; originally published in French, 1986), *La Guinée Équatoriale* (1979), and *Historical Dictionary of Equatorial Guinea,* 2nd ed. (1988), are authoritative. RENÉ PÉLISSIER, *Études hispano-guinéennes* (1969), is the only detailed study of the preindependence era. See also SUZANNE CRONJÉ, *Equatorial Guinea, the Forgotten Dictatorship* (1976); AKINJIDE OSUNTOKUN, *Equatorial Guinea–Nigerian Relations* (1978); and ALEJANDRO ARTUCIO, *The Trial of Macías in Equatorial Guinea* (1979). Further bibliographic information is found in MAX LINIGER-GOUMAZ, *Guinea Ecuatorial* (1974), and periodic supplements. (R.J.H.–C.)

Ghana: E.A. BOATENG, *A Geography of Ghana,* 2nd ed. (1966), is an account of the systematic and regional geography. JACK GOODY, *Changing Social Structure in Ghana* (1975), is an ethnological examination. Political development is studied in DAVID KIMBLE, *A Political History of Ghana: The Rise of Gold Coast Nationalism, 1850–1929* (1963, reprinted 1971); F.M. BOURRET, *Ghana: The Road to Independence, 1919–1957,* rev. ed. (1960); DENNIS AUSTIN, *Politics in Ghana, 1946–1960* (1964); and DAVID E. APTER, *Ghana in Transition,* 2nd rev. ed. (1972). J.D. FAGE, *Ghana* (1959, reprinted 1983); and DANIEL MILES MCFARLAND, *Historical Dictionary of Ghana* (1985), provide introductions. Ghana's historical importance in trading is related by K. YEBOA DAAKU, *Trade and Politics on the Gold Coast, 1600–1720* (1970); J.K. FYNN, *Asante and Its Neighbours, 1700–1807* (1971); and EDWARD REYNOLDS, *Trade and Economic Change on the Gold Coast, 1807–1874* (1974). (E.A.B./J.D.F.)

Guinea: JEAN SURET-CANALES, *La République de Guinée* (1970), now dated, is the best single source. CLAUDE RIVI-

ÈRE, *Guinea: The Mobilization of a People* (1977), although also dated, is a sound work. HAROLD D. NELSON *et al., Area Handbook for Guinea,* 2nd ed. (1975), is still valuable for background information. 'LADIPO ADAMOLEKUN, *Sékou Touré's Guinea* (1976), is an excellent work on the political system under the country's first president. THOMAS E. O'TOOLE, *Historical Dictionary of Guinea (Republic of Guinea/Conakry),* 2nd ed. (1987), contains an extensive bibliography. (T.E.O'T.)

Guinea-Bissau: For contemporary settlement and production patterns, ANNE-MARIE HOCHET, *Paysanneries en attente* (1983), is unequaled. CARLOS LOPES, *Guinea-Bissau* (1987), is a political study. Works on the war for national independence include LUÍS CABRAL, *Crónica da libertação* (1984); LARS RUDEBECK, *Guinea-Bissau: A Study of Political Mobilization* (1974); and STEPHANIE URDANG, *Fighting Two Colonialisms: Women in Guinea-Bissau* (1979). RICHARD LOBBAN and JOSHUA FORREST, *Historical Dictionary of the Republic of Guinea-Bissau,* 2nd ed. (1988), is a useful introduction. A comprehensive historical and contemporary survey through 1986 is ROSEMARY E. GALLI and JOCELYN JONES, *Guinea-Bissau* (1987). A. TEIXEIRA DA MOTA, *Guiné Portuguesa,* 2 vol. (1954), is the standard work on the colonial period, with English and French summaries at the end of both volumes. Other histories include JOÃO BARRETO, *História da Guiné, 1418–1918* (1938); and WALTER RODNEY, *A History of the Upper Guinea Coast, 1545–1800* (1970, reprinted 1982). See also JOSEPH M. MCCARTHY, *Guinea-Bissau and Cape Verde Islands* (1977), a bibliography. (R.E.Ga./D.Bi.)

Liberia: An overview is found in HAROLD D. NELSON (ed.), *Liberia,* 3rd ed. (1985). Physical features are described in WILLI SCHULZE, *A New Geography of Liberia* (1973). Afro-American 19th-century settlement patterns are analyzed by TOM W. SHICK, *Behold the Promised Land* (1980). Economic and political studies include CHRISTOPHER CLAPHAM, *Liberia and Sierra Leone* (1976); and R.W. CLOWER *et al., Growth Without Development* (1966). D. ELWOOD DUNN and SVEND E. HOLSOE, *Historical Dictionary of Liberia* (1984), provides an introduction. J. GUS LIEBENOW, *Liberia: The Evolution of Privilege* (1969), is supplemented by his *Liberia: The Quest for Democracy* (1987). D. ELWOOD DUNN and S. BYRON TARR, *Liberia* (1988), traces the political development of modern Liberia. (A.B.J./S.E.H.)

Nigeria: An overview is found in HAROLD D. NELSON (ed.), *Nigeria,* 4th ed. (1982). REUBEN K. UDO, *Geographical Regions of Nigeria* (1970); K.M. BARBOUR *et al., Nigeria in Maps* (1982); and K.M. BUCHANAN and J.C. PUGH, *Land and People in Nigeria* (1955, reissued 1966), are standard geographic texts. AKIN L. MABOGUNJE, *Urbanization in Nigeria* (1968), contains detailed studies of Lagos and Ibadan. W.T.W. MORGAN, *Nigeria* (1983), is a good presentation of the physical and cultural environment. An account of the economy up to World War II can be found in DARYLL FORDE and RICHENDA SCOTT, *The Native Economies of Nigeria* (1964); and GERALD K. HELLEINER, *Peasant Agriculture, Government, and Economic Growth in Nigeria* (1966). SUSAN M. MARTIN, *Palm Oil and Protest: An Economic History of the Ngwa Region, South-eastern Nigeria, 1800–1980* (1988), takes the story up to the oil boom. This last period is best read about in SCOTT R. PEARSON, *Petroleum and the Nigerian Economy* (1970); and SAYRE P. SCHATZ, *Nigerian Capitalism* (1977); and, in its wider dimension of political economy, in GAVIN WILLIAMS (ed.), *Nigeria* (1976); and I. WILLIAM ZARTMAN (ed.), *The Political Economy of Nigeria* (1983). On politics, JAMES S. COLEMAN, *Nigeria: Background to Nationalism* (1958); and RICHARD L. SKLAR, *Nigerian Political Parties* (1963), remain indispensable; along with K.W.J. POST, *The Nigerian Federal Election of 1959* (1963); JOHN P. MACKINTOSH (ed.), *Nigerian Government and Politics: Prelude to the Revolution* (1966); and LARRY DIAMOND, *Class, Ethnicity, and Democracy in Nigeria: The Failure of the First Republic* (1988). Constitutional development can be followed in B.O. NWABUEZE, *A Constitutional History of Nigeria* (1982). Studies of public administration include BILLY DUDLEY, *An Introduction to Nigerian Government and Politics* (1982); and 'LADIPO ADAMOLEKUN, *Politics and Administration in Nigeria* (1986). EKPO EYO, *Two Thousand Years, Nigerian Art* (1977), traces the development of an important part of Nigeria's culture.

The best general history is MICHAEL CROWDER, *The Story of Nigeria,* 4th ed. rev. (1978); it is supplemented by OBARO IKIME (ed.), *Groundwork of Nigerian History* (1980). Regional and special studies include R.A. ADELEYE, *Power and Diplomacy in Northern Nigeria, 1804–1906: The Sokoto Caliphate and Its Enemies* (1971); E.A. AYANDELE, *The Missionary Impact on Modern Nigeria, 1842–1914* (1966); A.E. AFIGBO, *The Warrant Chiefs: Indirect Rule in Southeastern Nigeria, 1891–1929* (1972); and G.O. OLUSANYA, *The Second World War and Politics in Nigeria, 1939–1953* (1973). The Biafran War is covered in JOHN DE ST. JORRE, *The Nigerian Civil War* (1972). On the Second Republic, RICHARD A. JOSEPH, *Democracy and Prebendal Politics in Nigeria* (1987); TOYIN FALOLA and JULIUS IHONVBERE, *The Rise & Fall of Nigeria's Second Republic, 1979–84* (1985); ANTHONY KIRK-GREENE and DOUGLAS RIMMER, *Nigeria Since 1970* (1981); and WILLIAM D. GRAF, *The Nigerian State* (1988), are informative. See also A. OYEWOLE, *Historical Dictionary of Nigeria* (1987).

(R.K.U./J.F.A.A./A.H.M.K.-G.)

Sierra Leone: D.R.G. GWYNNE-JONES *et al., A New Geography of Sierra Leone* (1978), is an introduction; see also JOHN I. CLARKE (ed.), *Sierra Leone in Maps,* 2nd ed. (1969). KENNETH LITTLE, *The Mende of Sierra Leone,* rev. ed. (1967); MICHAEL JACKSON, *The Kuranko* (1977); W.T. HARRIS and HARRY SAWYERR, *The Springs of Mende Belief and Conduct* (1968); and R.H. FINNEGAN, *Survey of the Limba People of Northern Sierra Leone* (1965), are accounts of the most populous communities. AKINTOLA WYSE, *The Krio of Sierra Leone* (1989), contains an account of the Krio of the western area. On the economy, see RALPH GERALD SAYLOR, *The Economic System of Sierra Leone* (1967). Accounts of modern political developments include GERSHON COLLIER, *Sierra Leone* (1970); WALTER BARROWS, *Grassroots Politics in an African State: Integration and Development in Sierra Leone* (1976); CHRISTOPHER CLAPHAM, *Liberia and Sierra Leone* (1976); THOMAS S. COX, *Civil-Military Relations in Sierra Leone* (1976); ARTHUR ABRAHAM, *Mende Government and Politics Under Colonial Rule* (1978), on the period 1890–1937; JOHN R. CARTWRIGHT, *Political Leadership in Sierra Leone* (1978); and GUSTAV H.K. DEVENEAUX, *Power Politics in Sierra Leone* (1982). CHRISTOPHER FYFE, *A History of Sierra Leone* (1962), is a major historical work; it is complemented by CYRIL P. FORAY, *Historical Dictionary of Sierra Leone* (1977). Other historical studies include C. MAGBAILY FYLE, *The History of Sierra Leone* (1981); ADAM JONES, *From Slave to Palm Kernels: A History of the Galinhas Country (West Africa), 1730–1890* (1983); D.A. TURAY and ARTHUR ABRAHAM, *The Sierra Leone Army: A Century of History* (1987); and MURRAY LAST and PAUL RICHARDS (eds.), *Sierra Leone, 1787–1987: Two Centuries of Intellectual Life* (1987). (D.S.H.W.N./C.Fy.)

Togo: ROBERT CORNEVIN, *Histoire du Togo,* 3rd ed. rev. and expanded (1969), covers geography, prehistory, ethnology, and history. SAMUEL DECALO, *Historical Dictionary of Togo,* 2nd ed. (1987), is a useful reference with an extensive bibliography. Ethnographic studies include JEAN-CLAUDE FROELICH, PIERRE ALEXANDRE, and ROBERT CORNEVIN, *Les Populations du Nord-Togo* (1963); RAYMOND VERDIER, *Le Pays kabiyé* (1982); and FRANÇOIS DE MEDEIROS (ed.), *Peuples du Golfe du Bénin* (1984). The political evolution of the country since independence is presented in SAMUEL DECALO, "The Benevolent General: Military Rule in Togo," in his *Coups and Army Rule in Africa* (1976), pp. 87–121; and in the issue titled "Togo Authentique," *Politique Africaine,* 27 (September–October 1987). ARTHUR J. KNOLL, *Togo Under Imperial Germany, 1884–1914* (1978), is a study of the German colonial era. COMI M. TOULABOR, *Le Togo sous Eyadéma* (1986); and ANDOCH NUTÉPÉ BONIN, *Le Togo du sergent en général* (1983), discuss the period of military rule. (S.De.)

Work and Employment

The story of work is a story of humanity's trials and triumphs—from the ordeal of hard work, sometimes under conditions of slavery with obedience mandated by the whip, to the development of tools and machines, which take the burdens off of human backs and even human minds. These advances in technology, which are still taking place, extend the reach of the hand, expand muscle power, enlarge the senses, and multiply the capacities of the mind. This story of work is still unfolding, with great changes taking place throughout the world and in a more accelerated fashion than ever before.

But work involves more than the use of tools and techniques. The form and nature of the work process help determine the character of a civilization, but, in turn, a society's economic, political, and cultural characteristics shape the form and nature of the work process as well as the role and status of the worker within the society.

Work is essential in providing the basic physical needs of food, clothing, and shelter, and differing explanations have been given at different times for its existence and purpose in human survival. Thus, in Chinese civilization, work became part of the Taoist flow of nature to which a person must adapt as part of the natural world. However, in the Judeo-Christian religious tradition (and in pagan religions as well), it was regarded as a punishment sent by God (or by the gods or spirits) to punish human beings for some deviation from the wishes or rules of the divine.

The human spirit, however, is too resilient and optimistic to face an eternal and damning process of hard physical labour, as most work was during most of human history for most people, so more benign explanations of the meaning and purpose of work came into use. For example, in western Christendom the Benedictine monks enunciated the rule that "to work is to pray," to fulfill one's duty to God and thus achieve salvation. This notion of work bringing spiritual rewards, in addition to physical survival, was carried further during the 17th century by the Puritans, whose work ethic led them to regard the accumulation of material wealth through labour as a sign of God's favour as well as of the individual's religious fervour. This attitude still appears in the American expression, "You are what you do," implying that people define themselves by the nature of their work.

With the onset of the Industrial Revolution and the development of powered machinery during the 18th and 19th centuries, much onerous physical effort was gradually removed from work in factories and fields. Work was still regarded, however, as something separate from pleasure, and the dichotomy between work and play persists even in today's highly industrialized society.

Most recently, the development of automated work devices and processes and the introduction of the computer into the service trades, especially in offices, has led futurists to speak of a "postindustrial society." This vision has not prevailed, however, with the spread of industrial production to developing nations. This globalization of production has meant that economic and political questions of working-class and managerial relationships have altered on an international front, affecting political relationships on a global scale. Furthermore, new demands have been placed on educational systems in the developing nations as they attempt to train their workers for industrial production. Similarly, new demands have been placed on the educational systems of the developed nations as the old assembly-line routine of specialized work has been taken over by "smart" machines.

In brief, the world of work and of the varying worker–manager relationship is in a state of flux as a result of changes in the technological, cultural, political, and economic environments. The study of changes in the organization of work resulting from such alterations in the past can perhaps lead to a better understanding of the present problems—now on a worldwide scale—resulting from ongoing technical, political, and economic changes. Hence, this article employs both historical and current perspectives in order to provide a basis for understanding work in today's world and possible changes in the future.

The article begins with a history of the organization of work from the Stone Age to the modern age of computerized automation. The scientific study of the behaviour of workers in organizations of employment is then covered in the following section, on industrial and organizational relations. A section on labour economics presents the economic models that have been used to explain the prices placed upon labour and the products of labour. Finally, the history of organized labour and the structures of labour unions are traced.

A discussion of labour law is included in the article BUSINESS LAW. The design and management of workplaces and work processes are covered in INDUSTRIAL ENGINEERING AND PRODUCTION MANAGEMENT. The principles underlying labour economics are explained in ECONOMIC THEORY. For discussions of the inventions that have shaped the nature of work, see TECHNOLOGY, THE HISTORY OF; and TOOLS. For coverage of the economic activities under which people labour, see FARMING AND AGRICULTURAL TECHNOLOGY and industrial articles such as INDUSTRIES, MANUFACTURING.

For coverage of related topics in the *Macropædia* and *Micropædia,* see the *Propædia,* sections 533 and 712, and the *Index.*

The article is divided into the following sections:

History of the organization of work 917
- Organization of work in preindustrial times 917
  - Prehistory
  - The ancient world
  - The medieval world
  - From the 16th to the 18th century
- Organization of work in the industrial age 923
  - The coming of mass production
  - Sophistication of mass production
  - Automation

Industrial and organizational relations 928
- Worker, manager, and society 928
  - Conceptions of the worker
  - Conceptions of the manager
  - Responsibility to society
- The changing work force 932
  - Interests, values, and expectations
  - Work careers
- Organizational design 935
  - Two models of work structure
  - Change in organizations
- Union–management relations 936
  - Union organizing
  - Collective bargaining
- The workplace in different cultures 937
  - The United States
  - Japan
  - West Germany
  - Common challenges

Labour economics 939
- Quantity and quality of the labour force 940
  - Activity rates
  - Quality of labour
- Deployment of the labour force 940
- Fixing rates of pay 941
  - Trade unions and bargaining areas
  - Effects of collective bargaining
  - Theory of bargaining
  - Methods of payment
  - Public regulation of rates of pay

The structure of pay 944
Occupational pay theories
The supply price of labour
Productivity and the price of labour
Movement of the general level of pay 946
Structure of the labour force: employment and unemployment 947
The three sectors 947
The effect of technology on employment patterns 947
Measuring unemployment 947
Underemployment
Rates of unemployment
Government efforts to increase employment 948
Migrant labour 949
Characteristics of migrant labour 949
Association with industrialization
The migrant flow
Economic and social conditions of migrant workers 949
Living conditions and wages
Difficulties of regulation
Patterns of movement 950
The United States
Western Europe
Africa
Organized labour: trade unionism 951
Origins and background of the trade union movement 951
Union development in industrialized nations 951
Great Britain
United States
Western Europe
Communist nations
Japan
Australia and New Zealand
Development of trade unions in emerging nations 957
The Indian subcontinent
Latin America
Africa
International organizations 958
Bibliography 958

HISTORY OF THE ORGANIZATION OF WORK

Organization of work in preindustrial times

PREHISTORY

Organization of work may have begun before the evolution of *Homo sapiens*. Along with tools, a more complex brain structure, and language communication, division of labour may have been responsible for starting the human conquest of nature and differentiating human beings from other animal species.

In these early stages of human development, work was confined to simple tasks involving the most basic of human needs: food. There could be no widespread geographic division of labour, because the population was sparse and isolated in small groups. Prehistoric people were largely food gatherers and hunters. The intermittent availability of food allowed little surplus for exchange, and there were few contacts with groups in different places that might have specialized in obtaining different foods. Nevertheless, there was some opportunity to augment the food supply by organizing the work of foraging and hunting and, later, agriculture.

Age, sex, and class. The most obvious division of labour arose from differences in age and sex. The oldest people in the tribe lacked strength and agility to hunt or forage far afield and so performed more sedentary tasks. The very youngest members of the tribe were similarly employed and were taught simple food gathering. There was also a sexual division of labour based upon the fact that women bore and nursed children and could not as easily participate in hunting. Women, however, were required to work as hard as men.

Early division of labour

In the earliest human groupings there was no division of labour based upon class. The precarious character of food gathering made it necessary for the whole group to contribute; there could be no leisure class or even a class of full-time specialists producing articles not directly related to the food supply. There were, however, part-time specialists. When men learned to make flint tools and weapons, some individuals became so skillful that they could produce a surplus to be exchanged for food.

Communal organization. Although there were no full-time specialists and no true division of labour in prehistory, work nevertheless often required organization. Capture of game and fish required varying degrees of cooperation among members of the group. Communal activity of this type had important social implications. Food had to be equitably shared, and a leader was needed to organize and direct the group. Because the basic social group was the family tribe, kin relationships—from the tribal chief down—formed the basis for the "managerial hierarchy." Bones of large animals killed by hunters have been found in sites of the Upper Paleolithic period (about 40,000 BC to about 10,000 BC), indicating a high degree of organization in hunting at this early stage of human development. Shortly thereafter men began using dogs as assistants in hunting.

Pottery. A more complex organization of work came with the development of pottery. While some sort of clay adequate for making passable pottery can be found nearly everywhere, the best potter's clay is not universally distributed. Thus, people in some locations could make pottery products for exchange. Skilled workmanship and specialized tools aided production, perhaps further encouraging specialization. There is no conclusive evidence that the earliest potters spent their full time at that task or that pottery making was carried on by women in its primitive stages, before introduction of the potter's wheel. There is reason to believe, however, that in prehistoric times some organization of the work existed. The gathering of the clay and firing materials may have been the work of the men, while women may have fashioned and decorated the pots.

Textiles. The same type of specialization is also implied in the making of textiles. Early protective garments were derived from animal skins. The development of agriculture cut into the supply of available skins and required a substitute material for clothing. To make textiles, yarn had to be spun; the earliest apparatus for this work consisted of a spindle and a distaff (a forked stick holding the unspun fibres). The figurative usage of the term distaff to refer to a female possibly indicates sexual division of labour in the early spinning process, with the man operating the spindle and the woman holding the distaff.

Agriculture. Primitive agriculture, too, may have dictated a division of work along sexual lines, with the fields entrusted to the women while the men hunted. In some agricultural tasks, however, such as clearing land, the men doubtless helped. Possibly there was little full-time specialization; in its beginnings, agriculture was carried on part-time. Yet even in its earliest stages agriculture was significant to the organization of work, for it provided a slight surplus that could be used to support the first real specialists: makers of metal tools and weapons.

Metallurgy. Although the origins of metallurgy are as yet unclear, employment of copper tools and weapons on a substantial scale suggests a new organization of work in which some persons devoted their full time to mining, smelting, and forging. Although deposits of flint for stone tools and weapons were fairly widely and evenly distributed, copper ores were not. Some of the earlier copper artifacts and remains of early copper mines have been found in areas where climate and topography would render agriculture difficult if not impossible. Geography thus dictated that the earliest miners and metalworkers could not be part-time agriculturalists but had to specialize. Besides, the techniques of prospecting, mining, smelting, casting, and forging were probably so demanding of physical strength and mental concentration as to preclude the metallurgist from farming or hunting activities.

Specialists in metalworking

Because copper ores are generally located in mountainous regions, the metal had to be transported to its users. Thus, metalworkers and their families had to be supported by the surplus foodstuffs of farmers; only when agriculture had developed beyond the subsistence level could such specialists survive. Not surprisingly, metallurgy developed first near the irrigation-farming valleys of the great river systems of the Nile, Tigris–Euphrates, and Indus, all of which provided a high yield of foodstuffs per acre. If metalworkers pursued their occupations full-time, their existence also implies other craft specialties. The combination of agricultural surpluses with copper and bronze tools provided the basis for development of the great irrigation civilizations of the Middle East. There the organization of work developed along lines that remained unchanged for the next 5,000 years, until the beginnings of mechanization and industrialization in the 18th century.

THE ANCIENT WORLD

The first general theory advanced to explain the development of ancient civilizations with systematic organization of work on a large scale, the emergence of social classes, and widespread specialization was elaborated in the United States by the historian and political scientist Karl Wittfogel in his seminal book *Oriental Despotism* (1957). Wittfogel believed that the development of irrigation works in such areas as Mesopotamia and Egypt led to the use of mass labour, to an organizational hierarchy for coordinating and directing its activities, and to government control for ensuring proper distribution of the water. Though tribal societies had had some form of government, this was usually personal in nature, exercised by a patriarch over a tribal group related by various degrees of kinship. Now, for the first time, an impersonal government as a distinct and permanent institution was established.

Effects of irrigation

Irrigation increased the food supply, allowing larger numbers of people to agglomerate into towns and cities. Because farmers were vulnerable to attack, armies were needed, with the implication of an officer class. Town specialization of labour brought the emergence of potters, weavers, metalworkers, scribes, lawyers, and physicians, while the new surpluses also created the basis for commerce. The more complex economy required records, so writing, of which the first examples come from the bookkeeping records of the storehouses in ancient Mesopotamia, was born.

Wittfogel's theory has been modified by scholars who point to the emergence of urban civilizations in some areas without the presence of large-scale irrigation works. In their view, several factors, including geographic features, natural-resource distribution, climate, kinds of crops and animals raised, and relations with neighbouring peoples, entered into the response to the environment. These scholars might be said to apply a "systems" approach to the interpretation of the origins of organized societies.

Social classes. In any case, by the time written history began, distinct social and economic classes were in existence, charged with identifiable functions. At the apex of the social pyramid stood the ruler and the nobles, probably grown out of a warrior group that had subjugated their neighbours. Closely aligned with them were the priests; often, as in Mesopotamia and Egypt, the ruler was worshiped as a divinity. Possessing knowledge of writing and mathematics, the priests served as government officials, organizing and directing the economy and overseeing clerks and scribes. Occupying a rung in the hierarchy between the noble–priest class and the peasants were the traders and merchants, who distributed and exchanged goods produced by others. A sizable group of artisans and craftsmen, producing specialized goods, belonged economically to the lower classes. At the bottom of the social scale were the slaves, originating as war captives or ruined debtors. The social structure in classical Greece and Rome continued to follow these lines. For relatively short periods of time republican or democratic governments did away with the ruling group, substituting a class of free landholders and providing a citizen army of warriors, but the basic economic organization remained unchanged.

Certain characteristics of the ancient organization of work emerged from the rigid social stratification described above. Chief among these was the hereditary nature of occupations and status. At certain times and places—in the later Roman Empire, for example—heredity of occupation was enforced by law, but tradition was usually sufficient to maintain the system. The social structure remained remarkably stable, a tendency fortified by the organizations of workers engaged in the same occupation, either voluntary or with the sanction of the law, that were the prototypes of the medieval guilds.

Heredity of occupation

Agriculture. *The family farm.* The basic agricultural work unit in the ancient world was the family. Even in certain regions where the state owned the land, farms were allocated by family. Furthermore, when large farming estates were formed during the Roman Empire, the structure of rural society was little affected, because the owners commonly left cultivation of their land to peasants who became their tenants.

Work within the family farm unit was divided along sexual lines: the men bore chief responsibility for such seasonal tasks as plowing, sowing, tilling, and harvesting, while the women prepared food and made clothing. Sons and daughters were apprenticed to their parents, and, if slaves were available, their work was similarly divided.

Technology and the seasonal nature of agriculture also influenced work organization. The usual draft team in antiquity, a pair of oxen, required two operators: a driver for the team and a guide for the plow. During planting and harvesting, the entire family performed fieldwork.

Estates. In the large estates, or latifundia, of the Roman Empire, the organization of work was quite complex, and a hierarchy of supervisors came into existence. The Greek historian Xenophon (5th–4th century BC) and the Roman statesman Marcus Porcius Cato (3rd–2nd century BC) wrote handbooks for management of such estates. Cato also outlined the work organization for a medium-sized farm. For an estate of 150 acres (60 hectares) with olive trees he recommended an overseer, a housekeeper, five farmhands, three carders, a donkey driver, a swineherd, and a shepherd. To these 13 on the permanent labour force, extra hands would be added for the harvest period.

On the larger latifundia that developed from about the 2nd century BC, the owner was usually nonresident, often because he had many scattered estates. Direction of the affairs of each was left in the hands of a bailiff under whose command slaves, numbering in the hundreds or even in the thousands, were divided into gangs charged with different responsibilities.

Crop specialization. Ancient agricultural work was also characterized by specialization in crops, Greece and Italy concentrating on the vine and the olive, with cereals being cultivated on the richer soils of Sicily, North Africa, and Asia. Wine and oil required craftsmen to produce amphorae for storage and conveyance, as well as tradesmen and small sailing vessels for transport.

Crafts. Economic growth, sophistication of taste, and enlarged markets ultimately brought mass production of a sort, with large workshops dedicated to production of a single item. These workshops, however, never achieved the size of even a small modern factory; craft production throughout antiquity remained distinctly modest in scale; a building in which a dozen persons worked was considered a large factory, though a few workshops were larger.

The earliest specialized craftsmen were probably itinerant, gravitating to wherever their services were in demand. As the market grew, craftsmen became sedentary, and traders carried their products to market. Economic growth also led to a multiplication of specialized crafts, to the organization of guildlike groups, and to a geographic division of labour, with members of one craft located in a special quarter of a city or in one area of a country. In the pottery industry, specialization was carried even further, with shaping, firing, and decoration sometimes done in separate establishments and with workshops specializing in cooking pots, jars, goblets, and funerary urns.

Specialization of crafts

As ancient workshops grew in size, slaves were introduced. The chief examples of large-scale production by slaves were in mining and metallurgy, in which working conditions were especially harsh but in which work or-

ganization was refined. In the silver mines at Laurium, in ancient Greece, the master miner commanded three gangs of labourers. The strongest handled picks at the ore face, weaker men or boys carried ore from the mine, and women and old men sifted the ore-bearing rock. The miners worked 10-hour shifts (followed by 10 hours of rest) in dark and narrow passages with smoky lamps that made the air almost unbreathable. On the surface, the master smelter supervised the workshops, in which the strongest men worked the mortar and the weakest the hand mill. Metallurgical working of the ore was carried out by small units, because the small leather bellows limited the size of the furnace. Metallurgy thus remained essentially a handicraft.

After weapons and tools, the chief use of metal was for ornamentation. The metalworker was more artisan, or even artist, than industrial worker, and in the trade there were patternmakers, smelters, turners, metal chasers, gilders, and specialized goldsmiths and silversmiths.

Large-scale building. The monumental public-works projects of the ancient world demonstrate a remarkable degree of organization in the absence of power and machinery. The Great Pyramid of Giza, built about 2500 BC, before the Egyptians knew the pulley or had wheeled vehicles, covers 13 acres (5.3 hectares) and contains the staggering total of 2,300,000 colossal blocks of granite and limestone weighing an average of 5,000 pounds (2,300 kilograms). There exists no complete historical or archaeological record of the exact methods of quarrying, transportation, and construction of the pyramids, and what evidence remains is often contradictory. Obviously, however, the need to organize the work on a systematic and rational basis was superbly met. It is estimated that some 100,000 workers were involved over 20 years in building the Great Pyramid, and the logistic problem alone, housing and feeding this large army of workers, required a high degree of administrative skill.

Building the Great Pyramid

The master builder, who planned and directed the erection of the pyramids and other great structures, occupied a high position in society. Ancestor of the modern architect and engineer, he was a trusted court noble and adviser to the pharaoh. He directed a host of subordinates, superintendents, and foremen, each with his scribes and recorders.

Although some slaves were employed in building the pyramids, most of the builders were peasants, drafted as a form of service tax (corvée) owed the state and employed when the Nile was flooding their fields. Workers were not regarded as expendable; overseers and foremen took pride in reporting on their safety and welfare. In a record of a quarrying expedition to the desert, the leader boasted that he had not lost a man or a mule. The labourers were organized into gangs: skilled workers cut granite for the columns, architraves, doorjambs, lintels, and casing blocks; masons and other craftsmen dressed, polished, and laid the blocks and probably erected ramps to drag the stones into place.

The Greeks and Romans used advanced organizational techniques in the building of monuments. The Roman road network, aqueducts, public buildings, public baths, harbours, docks, and lighthouses demanded exceptional skill in organizing materials and workmen, implying in turn a rational division of labour among craftsmen.

THE MEDIEVAL WORLD

The organization of work and division of labour may, in fact, be said to have reached a peak during the Roman Empire, but it inevitably declined as the empire disintegrated in the West. The social and political fragmentation and economic decay of the late empire reduced most of western Europe to small-scale, self-sufficient economic units. As this happened, the market for specialized production largely disappeared. As the new feudal society achieved stability, however, trade and town life began to revive; interregional commerce was stimulated, and specialized crafts reappeared to serve growing markets.

Important technological innovations in agriculture, power, transportation, metallurgy, and machines brought further areas of specialization into being. The emergence of the new burgher class, with rapidly growing wealth and breadth of enterprise, provided the basis for a more rational management of production, hastening the rise of industrialization.

Class structure. Social divisions, or class structure, in the medieval world reflected a division of labour. The noble class, or seigneurs, were fighters, protectors, important consumers, and, in a sense, organizers of work. Because they controlled the land, basic to production in this agrarian society, the nobles alone possessed the wealth to purchase the products of artisans, to buy goods brought from a distance, to acquire the weapons and armour made by metallurgists, and to construct castles and fortresses. The lords also decided, in accordance with prevailing custom, how the farm work should be organized.

Land-owning nobility

The clergy were both consumers and producers, responsible for the spiritual care of their parishioners. The monasteries were self-sufficient agrarian units that often produced a surplus for trade; indeed, the monks experimented in improving farming techniques and in producing special cheeses and wines for an outside market. Finally, the great churches required specialists in stained glass, bell founding, stonemasonry, wood carving, and other trades.

The bulk of the population were farmers of varying legal status. Most were serfs bound to the plots of ground their ancestors had tilled and owing certain specified services and dues to the lord of the manor, who extended protection in return. A few inhabitants of the manor were tenant farmers, or sharecroppers, renting the lord's land in return for payments of a share of the produce. Fewer still were free farm labourers who worked for wages. Slavery had all but disappeared. Because the manor was practically self-sufficient, peasants of whatever status performed a variety of tasks connected with their agricultural occupation.

Agricultural production. Four interrelated factors determined the work organization of medieval agriculture: the economic self-sufficiency of the manor, the development of mixed agriculture based on crop growing and stock raising, such technological improvements as the heavy wheeled plow and rigid horse collar, and the system of land tenure and division of holdings. Each peasant household produced nearly everything it needed and paid for the use of a feudal-monopoly mill or winepress, not in money but by a percentage of the crop being processed.

In antiquity, stock raising and crop production had been separate enterprises. In northwestern Europe during the Middle Ages, the two were combined. Animals were needed both for draft and for food, and, because the yield of the grainfields did not greatly exceed human requirements, stock was pastured on poor land or harvested fields. Thus a certain amount of land was reserved for pasturage, and some villager, usually an older member of the community, became herdsman.

Communal organization was made imperative by the land-tenure arrangements and by the way in which arable land was divided among villagers. In a typical manor at least half the cultivable soil was held by peasant cultivators. In order to assure an equitable apportionment, the land was divided into large fields, each of which was subdivided into narrow strips. Each peasant held strips in each field; the work of plowing, planting, and harvesting, therefore, had to be done in common and at the same time.

Communal crop farming

The wheeled plow, gradually introduced over several centuries, further reinforced communal work organization. Earlier plows had merely scratched the surface of the soil, necessitating cross-plowing, which in turn made square fields desirable. The new plow was equipped with a heavy knife (colter) to dig under the surface of the soil and a moldboard to turn over the sod, eliminating the need for cross-plowing and making strip fields possible. On the other hand, because the new plow required a team of eight oxen, more than any single peasant owned, plowing (and indeed all heavy work on the manor) was pooled. Of course, such a system allowed little room for individual initiative; everyone followed established routines, with the pace set by the slowest oxen.

Industrial production. *The craft guilds.* In contrast to the land-bound serfs, townspeople of the Middle Ages were free. Some engaged in commerce, including long-distance trade, and formed corporate bodies known as

Master, journeyman, and apprentice

merchant guilds. The majority, however, were small merchant-craftsmen, organized in craft guilds as masters (of highest accomplishment and status), journeymen (at a middle level), and apprentices (beginners). The medieval master was typically many things at once: a workman himself, the most skilled in his shop; a foreman, supervising journeymen and apprentices; an employer; a buyer of raw or semifinished materials; and a seller of finished products. Because medieval craftsmen employed simple hand tools, the skill of the workman rather than his equipment determined the quantity and quality of his output. Hence, there was a long period of learning for the apprentice and journeyman. At first the training was by "show and tell" and then "learning by doing" under the guidance of a more experienced workman. When a "masterpiece" could be produced that would receive the approval of the guild masters, the craftsman gained full admission into the guild.

The essence of craft-guild organization was regulation. By controlling conditions of entrance into a craft, guilds limited the labour supply. By defining wages, hours, tools, and techniques, they regulated both working conditions and the production process. Quality standards and prices were also set. Monopolistic in nature, the guilds, either singly or in combination, sought complete control over their own local markets and endeavoured to exclude outsiders. In order to attain and protect their monopoly, the guilds acquired a political voice and in some locations achieved the right to elect a number of their own members to the town council. In some towns, such as Liège, Utrecht, and Cologne, guilds achieved complete political control. The 32 craft guilds in Liège, for example, so dominated the town after 1384 that they named the town council and governors and required all important civic decisions to be approved by a majority vote of their membership.

The most prosperous period for the craft guilds was the 14th century. Specialties became so differentiated that there were more than 100 guilds in large towns. In northern Europe, for example, at the beginning of the period, carpenters built houses and made furniture. In time, furniture making became a new craft, that of joinery, and the joiners broke from the carpenters to establish their own guilds. The wood-carvers and turners (who specialized in furniture turned on a lathe) founded guilds also. Painting and gilding of furniture and wood carvings were done by members of the painters' guild.

Simultaneously, there was a countermovement toward amalgamation of different crafts, a tendency that reflected the growth of the market and the desire of enterprising masters to develop the trading function at the expense of the handicraft function. The typical medieval craft regulation had as its objective for each craftsman to gain an equal share in the trading function, a monopolistic division of the market. As craft differentiation reached a point at which a number of crafts participated in the production of the same or similar articles, such conditions could not be maintained. The steadily widening search for raw materials and steadily growing marketplace stimulated craftsmen at each end of the production chain to concentrate on trade, to assure themselves either of raw materials or of a market. In either case, masters were tempted to employ members of other crafts, and conflicts inevitably arose.

The same widening of the market led to differentiation of classes within a craft. The trading function acquired increasing importance and also came to be exercised by certain individuals within a single craft organization; those who remained craftsmen fell into a condition of dependence upon the traders. At the same time, merchant guilds, originally representatives of traders only, absorbed the craft guilds, over which they had gradually acquired economic control.

Internally, too, the craft guilds suffered a breakdown in structure. Because the masters sought to retain the profits of the growing market for themselves, they made it increasingly difficult for journeymen to enter their class, preferring instead to employ them as wageworkers. Apprentices similarly had little hope of rising to mastership. Thus the master–journeyman–apprentice relationship gave way to an employer–employee arrangement, with the master performing the functions of merchant while his employees did craftwork. Conditions for development of the early industrial system—the proto-factory—thus arose with the disintegration of the original craft-guild system. The excluded journeymen eventually became a class of free labourers who practiced their craft for wages outside town walls where guild regulations did not reach.

From craftsman to wage earner

The putting-out system. Certain industries that were small at the outset of the Middle Ages had by their close become large in scale, with accompanying changes in the organization of work. The most important of these was the wool-cloth industry.

The basic clothing material in western Europe until the beginning of modern times was wool. Linen and silk, though used since antiquity, were too costly to become raw materials for any large-scale industry until a later date, and cotton was grown only in small volume. The production of cloth from wool involved several time-consuming steps: cleaning and carding (straightening curled and knotted fibres sheared from the sheep), spinning the fibres into thread, weaving the thread into cloth, shearing off knots and roughness, and dyeing. All these processes could be carried on within a single peasant household, for they required only simple apparatus and rudimentary skills. Children carded the wool; women operated the spinning wheel; and men worked the loom shuttles.

The cloth produced by such crude tools and relatively unskilled workers was rough but serviceable. Those above the peasant class, however, wanted more comfortable and attractive clothes that could be produced only by skilled craftsmen. Thus a large and growing demand existed for better textiles, and the industry rapidly outgrew the peasant household economy. A new organization of work, called the putting-out system, was instituted, in which a merchant clothier bought raw wool, "put it out" to be carded, spun, and woven into cloth, and then carried the cloth through the finishing processes with the help of skilled craftsmen. Because the spinners and weavers remained peasants, they earned part of their living from the plots on which their cottages stood; agriculture and industry were carried on together as almost an integrated enterprise. The man could work in the field while his wife spun; in winter the man joined in industrial production. At harvesttime every hand was out in the fields, leaving spinning wheel and loom idle.

Early capitalist and worker

The putting-out system differed from peasant household production in that the merchant clothier, or entrepreneur, bought the raw wool and owned the product through its stages of preparation, although the cottage workers owned their own spinning wheels, looms, and other tools. Thus the peasant farmer came to work on materials that did not belong to him, and his position gradually changed to one of dependency. On the other hand, the work was performed at home (the cottage system, or domestic system) and at the worker's pace rather than in a factory under impersonal discipline. The merchant, furthermore, simply arranged the order and sequence of the various technical processes without exercising direct supervision over their actual performance. Nevertheless, the merchant clothier who began putting out cloth came to control the entire production process and represents a step toward the industrial capitalist of the 19th century.

Advances in technology. Change in the scale of commerce through the establishment of large national markets was not the sole cause of changes in the organization of work during the Middle Ages. Equally important was the transformation caused by advancing technology, especially in the application of wind power and waterpower—the beginning of a long development in the replacement of human labour by machine power. From the late 10th century, waterwheels, long used for grinding grain, were applied to many industrial processes: tanning; olive pressing; sawing; armour polishing; pulverizing; operating blast-furnace bellows, forge hammers, and grindstones; and crushing mash for beer. The first horizontal-axle windmill appeared in western Europe in 1185, and within a short time windmills spread from northern England to the Middle East.

A good example of the change in industrial organization

created by the application of waterpower to an industrial process is the fulling (*i.e.,* shrinking and thickening) of cloth, a process that was mechanized in the 13th century. Previously fulling had been accomplished by trampling the cloth or beating it with a fuller's bat, processes dating back to antiquity. The fulling mill invented during the Middle Ages was a twofold innovation: first, two wooden hammers replaced human feet; and second, the hammers were raised and dropped by the power of a water mill. A series of hammers could be set to work with only one man needed to keep the cloth moving properly in the trough, which was filled with water and fuller's earth. The mechanization of fulling also changed the organization of work indirectly by causing the cloth industry to relocate along streams, often away from the former urban textile centres, thus bringing a relaxation of old institutional constraints.

Perhaps the best example of specialization of labour in the Middle Ages is to be found in the large-scale metal-mining industry in central Europe, as described by the German scholar and man of science Georgius Agricola in *De re metallica* (1556), the leading textbook for miners and metallurgists for nearly two centuries. In addition to the *Bergmeister* ("master miner"), the chief mine administrator, there was a hierarchy of clerical and technical personnel and a series of craftsmen and mechanics specializing in different phases of the mining operation: miners, shovelers, windlass operators, carriers, sorters, washers, and smelters. The mines operated five days a week on a 24-hour basis, with the workday divided into three seven-hour shifts and the remaining three hours used for changing shifts. Animal power was used wherever possible, with teams of eight horses hitched in pairs to turn windlasses and raise buckets of ore or drain water from the mine. Agricola's illustrations show many types of pumps for mine drainage: crank-operated, treadmill-operated, and waterpower-operated. There were also suction pumps of varying degrees of complexity. All were operated by specialized mechanics.

The bellows for mine ventilation were also operated either by human and animal power or by waterpower. Other mining processes were less mechanized and were carried on much as they had been in antiquity. Ores brought to the surface were taken to a sorting table on which women, boys, and old men sorted the pieces by hand, putting the good ores into wooden tubs to be carried to the furnaces for smelting.

Monumental construction. The mechanization that was changing the organization of work in large-scale industries throughout the medieval period was little apparent in the construction of castles, cathedrals, and town walls. The technology involved in the lifting of weights, for instance, had made little progress during the Middle Ages, and, because the freemasons declined to handle large blocks of stone, the Romanesque and Gothic structures were built with smaller stone blocks, nevertheless achieving grandeur in scale. The organization of labour differed greatly from that employed in antiquity. These great monuments were the first in history to be built by free labour including, carpenters, glaziers, roofers, bell founders, and many other craftsmen, besides the stonemasons.

Freely contracted labour

The free and unfettered nature of medieval construction as compared with that of antiquity can be seen both in the works themselves, which reveal a high degree of individualism, and in the records. For a long time it was believed that medieval craftsmen, especially those engaged in the building of cathedrals, were humble, self-effacing artisans who laboured piously and anonymously for the glory of God and for their own salvation. The industry of scholars has dispelled this myth. Medieval builders often left their names or signatures upon their work, and surviving records show names, wages, and occasionally protests over wages. The artisans were by no means anonymous: historians have uncovered more than 25,000 names of those who worked on medieval churches.

Directing the guild craftsmen was the master mason, who, like his ancient counterpart, functioned in many capacities: architect, administrative official, building contractor, and technical supervisor. He designed the molds, or patterns, used to cut the stones for the intricate designs of doors, windows, arches, and vaults. He designed the building itself, usually copying its elements from earlier structures upon which he had worked as a master or during his apprenticeship. He sketched his plans out on parchment. As administrator he kept the accounts, hired and fired the workers, and was responsible for procurement of materials. As technical supervisor he was constantly on the job for spot decisions and planning. In the largest projects he was assisted by undermasters.

FROM THE 16TH TO THE 18TH CENTURY

The proliferation of industry during the early modern period—that is, from the 16th to the 18th century, immediately preceding the Industrial Revolution, properly so called—arose from four factors: (1) the growth of wealth, derived partly from the influx of precious metals from the New World and from growing commercial and banking activities; (2) the growth of demand, or markets; (3) the introduction of new products; and (4) the development of new technology. The result was an increase in the scale of manufacturing industry throughout Europe, accompanied by changes in the organization of work.

The growth in the size of the market was caused only partially by the geographic explorations of the preceding era and subsequent colonization. Most of the new demand for goods came about as a result of a growing population, especially toward the latter part of the 17th century and beginning of the 18th century, and a rise in the standard of living, especially among the class of burghers, or bourgeoisie—the town dwellers. The markets also grew in size with the replacement of the small medieval feudalities by larger political units—the royal kingdoms. When economic sway was exercised over a larger jurisdiction, it tended to eliminate many of the local restrictions on trade and commerce established by the previous smaller political units. Many new products—including spices from Asia and sugarcane from the New World—were also introduced into Europe, either directly, by the explorers, or indirectly, through expanded trade with distant points. The demand continued to increase along with the growing affluence and new manners of living in European society. Handicraft production no longer sufficed, and the guilds declined.

Beginnings of the factory system. The changes in industrial patterns that emerged were less in the introduction of new mechanical contrivances than in the growth in extent of the application of power, primarily water and wind, to old devices and, even more significant, in the organization of work to meet the needs for production on a larger scale—the beginnings of the factory system.

Production on a large scale

The proto-factory of the 17th and 18th centuries centred in certain industries, especially textiles. The old guild corporations broke down, and the cottage system began giving way to larger units of production. The organization of commerce changed rapidly. New instruments in the fields of banking, insurance, and export marketing helped to amass capital and make it available for investment in industrial enterprises. In this movement toward the capitalization of industry, some of the lesser masters were driven down into the wage-earning class, while greater masters became capitalist employers.

In Britain the development of commercial concentration—and hence of industrial scale—was mainly the work of large companies or corporate bodies, such as woolen manufacturers, ironmasters, and hatmakers. Government encouragement was given by means of special legislation, especially grants of monopolistic charters. In France, however, the practice of mercantilism, a government-directed policy aimed at increasing national wealth and power, meant that the government itself took an active part in developing industries that were state owned and operated, among them the famous Gobelins tapestry works and the manufactories for production of furniture, porcelain, and other luxury items.

Although the state-run factories in France represented at least two of the essentials of factory production—the gathering of large groups of workers in one place and the imposition of disciplinary rules—they had little effect upon the organization of work because they produced

small quantities of luxury goods and hence amounted to large handicraft operations. Furthermore, despite their size, the French Royal Manufactories did not possess the third prime element of a true factory system: mechanization. The great historical change in the organization of work came in 18th-century Britain with the onset of the Industrial Revolution, largely as the result of the new technology of power-driven machinery.

The Industrial Revolution. *Mechanization.* The new machines introduced in the 18th century compelled a rational organization of job functions that was quite different from the old handicraft tradition. Adam Smith in *The Wealth of Nations* (1776) gave the classical description of the new production system as exemplified by a pin factory: "One man draws out the wire; another straights it; a third cuts it; a fourth points it; a fifth grinds it at the top for receiving the head; to make the head requires two or three distinct operations; to put it on is a peculiar business; to whiten the pin is another; it is even a trade by itself to put them into the paper; and the important business of making a pin is in this manner divided into about 18 distinct operations." According to Smith, a single worker "could scarce, perhaps with his utmost industry, make one pin in a day, and certainly could not make 20." The new methods enabled a pin factory to turn out as many as 4,800 pins a day.

The great increase in productivity depended far more upon the rational organization of processes than upon individual skill. In the textile industry manual dexterity and alert response proved to be more valuable than experience, accounting for the use of inexpensive woman and child labour in the early mills. Some vestiges of the medieval guild apprenticeship, however, still remained in the early textile factories; the children were sometimes bound as apprentices for a period of at least seven years, usually until they were 21. In some areas the old cottage system of textile production was moved to the factory, with the entire family employed as a work team. In those cases the father would be employed for any heavy work while supervising his wife and children at the machines. Presumably the father also possessed the mechanical skills necessary to repair and maintain the machinery.

Division of labour in the workplace. Because machines could justify their high cost only if a heavy and continuous demand existed for their output, their presence led to a division of labour between the entrepreneur who owned them and his employees. The owner supervised his workers, compelling them to work at the pace of the machine.

Factory discipline

Even in enterprises that were not yet fully mechanized, the advantages of factory discipline were apparent at an early stage of the Industrial Revolution. Josiah Wedgwood designed his pottery works at Etruria in England "with a view to the strictest economy of labour." His plant was laid out so that the pots were first formed and then passed through the painting room, the kiln room, the account room, in which an inventory of production was made, and finally to storage. In potteries before this time, the workers could wander from one task to another; in Wedgwood's, the employees were assigned a particular post and worked at one task only. Out of 278 men, women, and children employed by Wedgwood in 1790, only five had no assigned post; the rest were specialists.

While the argument is sometimes made that the division of labour destroyed skill, the fact is that it might also have improved the quality of the finished product, for Wedgwood's pottery was superior to that of his competitors. It can be said that the division of labour does not so much destroy skill as limit it to a particular field of expression; and, within that particular task, it increases skills by continued repetition. It is interesting to note that Wedgwood's chief difficulty was not so much in training his workers as it was in introducing them to a novel form of discipline that ran contrary to centuries of independence. It was a constant test of Wedgwood's ingenuity to enforce six hours of punctual and constant attendance upon his workers, to get them to avoid waste, and to keep them from drinking on the job and taking unauthorized "holidays." Because he was a busy man involved in all the tasks of running an enterprise and could not stand over his workers and control their movements, he had to develop a hierarchy of supervisors and managers.

There can be little doubt that the condition of the workers, especially the women and children, in the early textile factories was miserable: 14 to 16 hours every day spent performing repetitive tasks in noisy, smelly, and unsanitary surroundings; and the workers' slum homes were equally unhealthy. It was at this period that the "social question" arose: why should poverty continue to exist in a nation that had the capacity to produce enormous quantities of goods? Answers to that question were to produce new social philosophies and political movements that have had major effects on society and politics ever since.

New industries. The introduction of steam-driven machinery at the time of the Industrial Revolution brought new industries into being or transformed older ones. This was especially true of the metalworking trades, where technological innovations made possible the replacement of wooden machinery with metal and the manufacture of such items as metal nails, glassware, and iron bearings. Expansion of manufacturing helped to stimulate the coal-mining industry. Coal was replacing wood as a fuel especially in England and northern France, where deforestation had made wood scarce. The pressure on fuel supplies came not only from domestic heating requirements and from the metallurgical trades but also from the brickmaking, brewing, dyeing, and glassmaking industries. In the coal mines the organization of labour remained much as it had been described by Agricola in the 16th century.

Another spur to the rise of new industries was the religious warfare of the 16th and 17th centuries. The forced movement of populations helped spread technical capabilities to new areas. For example, the Protestant Huguenots, expelled from France near the end of the 17th century, carried with them their special skills in metalworking and glassmaking when they migrated to England, Holland, Germany, and the American colonies.

Spread of technology

Urbanization. One of the greatest stimuli toward a more rational organization of work was the general demographic trend in Europe from the 17th to the 19th century. Population grew swiftly, moving from country to city. It is possible that only a few European cities—Paris and the great Italian commercial cities of Venice, Genoa, and Naples—had as many as 100,000 people at the beginning of the modern era. London may have had only about half that number. By the beginning of the 17th century, however, rapid growth had begun, and, by the end of that century, London probably had 500,000 inhabitants.

Colonization of the New World. *Worldwide division of labour.* Although exploration and colonization had originally been carried out in order to secure exotic and expensive spices, these products had little direct influence upon the organization of work in Europe; even the enormous trade in semitropical items such as sugar and coffee had little effect. However, wheat, wool, and meat from the temperate areas ultimately brought about an international division of labour, with the New World colonies furnishing agricultural produce to the manufacturing countries of Europe. In the 20th century such a division still exists in somewhat different form, with the underdeveloped nations of the tropics supplying agricultural and industrial raw materials to developed areas.

Slavery. In its effect on the organization of work, the most important result of the colonization of the New World and the demand for its products was the use of slave labour. Slavery was linked first with sugar production in the West Indies and later with cotton in southern North America.

Cultivation of sugarcane, especially its harvesting, requires heavy manual labour. Harvested cane must be sent to a mill for grinding within a few hours after cutting; this requirement necessitated establishment of a plantation system in which the workers would be housed close to the fields and the sugar mill. The natives of the West Indies were not numerous enough to perform the required work and were temperamentally unwilling to engage in such labour, even when harsh means were employed to force them to do it. The requirements of sugar planters thus introduced agricultural slavery into the Western Hemi-

The plantation

sphere. It began as early as 1518, when the Spanish government granted a license to import some 4,000 African slaves into the Spanish colonies. The plantation system and the consequent demand for African slaves spread during the next two centuries throughout the sugar-growing areas, including the British West Indies. Indeed, the British islands of the West Indies carried specialization in sugar so far that they found it most profitable to devote nearly all their land to the export crop of sugarcane and to import other foods.

In the temperate zone, where sugar production was not possible, slaves were little used except in tobacco-growing areas until near the end of the 18th century. The Puritan communities in New England engaged in small family farming, while the Southern colonies employed indentured servants (white labourers who agreed to work a number of years for some person who had paid their passage to the New World).

Eli Whitney's invention of the cotton gin in 1793 lowered the price of upland cotton and led to the use of that fibre as a staple for textile production. As a result, Negro slavery and the plantation system became fixtures in the American South. While slaves were employed chiefly as cotton-field labourers, they also worked as craftsmen, factory hands, and domestic servants, creating, in other words, a division of labour on the plantation. The regional specialization in production led to sectional economic and political differences and ultimately to the American Civil War and to the freeing of the slaves.

Organization of work in the industrial age

THE COMING OF MASS PRODUCTION

Mass production is the name given to the method of producing goods in large quantities at low cost per unit. But mass production, although allowing lower prices, does not have to mean low-quality production. Instead, mass-produced goods are standardized by means of precision-manufactured, interchangeable parts. The mass production process itself is characterized by mechanization to achieve high volume, elaborate organization of materials flow through various stages of manufacturing, careful supervision of quality standards, and minute division of labour.

To make it worthwhile, mass production requires mass consumption. Until relatively recent times the only large-scale demand for standardized, uniform products came from military organizations. The major experiments that eventually led to mass production were first performed under the aegis of the military.

Industry. *Machine tools and interchangeable parts.* The material basis for mass production was laid by the development of the machine-tool industry—that is, the making of machines to make machines. Though some basic devices such as the woodworking lathe had existed for centuries, their translation into industrial machine tools capable of cutting and shaping hard metals to precise tolerances was brought about by a series of 19th-century innovators, first in Britain and later in the United States. With precision equipment, large numbers of identical parts could be produced at low cost and with a small work force.

American System of manufacture

The system of manufacture involving production of many identical parts and their assembly into finished products came to be called the American System, because it achieved its fullest maturity in the United States. Although Eli Whitney has been given credit for this development, his ideas had appeared earlier in Sweden, France, and Britain and were being practiced in arms factories in the United States. During the years 1802–08, for example, the French émigré engineer Marc Brunel, while working for the British Admiralty in the Portsmouth Dockyard, devised a process for producing wooden pulley blocks by sequential machine operations. Ten men, in place of 110 needed previously, were able to make 160,000 pulley blocks per year. British manufacturers, however, ignored Brunel's ideas, and it was not until London's Crystal Palace exhibition of 1851 that British engineers, viewing exhibits of machines used in the United States to produce interchangeable parts, began to apply the system. By the third quarter of the 19th century, the American System was employed in making small arms, clocks, textile machinery, sewing machines, and a host of other industrial products.

The assembly line. Though prototypes of the assembly line can be traced to antiquity, the true ancestor of this industrial technique was the 19th-century meat-packing industry in Cincinnati, Ohio, and in Chicago, where overhead trolleys were employed to convey carcasses from worker to worker. When these trolleys were connected with chains and power was used to move the carcasses past the workers at a steady pace, they formed a true assembly line (or in effect a "disassembly" line in the case of meat cutters). Stationary workers concentrated on one task, performing it at a pace dictated by the machine, minimizing unnecessary movement, and dramatically increasing productivity.

Drawing upon observations of the meat-packing industry, the American automobile manufacturer Henry Ford designed an assembly line that began operation in 1913. The result was a remarkable reduction of manufacturing time for magneto flywheels from 20 minutes to five minutes. This success stimulated Ford to apply the technique to chassis assembly. Under the old system, by which parts were carried to a stationary assembly point, 12½ man-hours were required for each chassis. Using a rope to pull the chassis past stockpiles of components, Ford cut labour time to six man-hours. With improvements—a chain drive to power assembly-line movement, stationary locations for the workmen, and work stations designed for convenience and comfort—assembly time fell to 93 man-minutes by the end of April 1914. Ford's methods drastically reduced the price of a private automobile, bringing it within the reach of the common man.

Ford's spectacular feats forced both his competitors and his parts suppliers to initiate his technique, and the assembly line spread through a large part of U.S. industry, bringing dramatic gains in productivity and causing skilled workers to be replaced with low-cost unskilled labour. Because the pace of the assembly line was dictated by machines, the temptation arose to accelerate the machines, forcing the workers to keep up. Such speedups became a serious point of contention between labour and management, while the dull, repetitive nature of many assembly-line jobs bored employees, reducing their output.

Assembly-line speedups

Effects on the organization of work. The development of mass production transformed the organization of work in three important ways. First, tasks were minutely subdivided and performed by unskilled workers, or at least semiskilled workers, since much of the skill was built into the machine. Second, manufacturing concerns grew to such size that a large hierarchy of supervisors and managers became necessary. Third, the increasing complexity of operations required employment of a large management staff of accountants, engineers, chemists, and, later, social psychologists, in addition to a large distribution and sales force.

Mass production also heightened the trend toward an international division of labour. The huge new factories often needed raw materials from abroad, while saturation of national markets led to a search for customers overseas. Thus, some countries became exporters of raw materials and importers of finished goods, while others did the reverse. In the 1970s and '80s some countries, particularly in Asia and South America, that had hitherto been largely agricultural and that had imported manufactured goods began industrializing. The skills needed by workers on assembly-line tasks were easily acquired, and standards of living in these developing countries were so low that wages could be kept below those of the already industrialized nations. Many large manufacturers in the United States and elsewhere therefore began "outsourcing"—that is, having parts made or whole products assembled in developing nations. Consequently, those countries are rapidly becoming integrated into the world economic community.

Agriculture. Starting in the 19th century, agriculture underwent a transformation comparable to the change from handicraft to mass production in industry. At the beginning of that century, farming was primarily a family

enterprise, employing age-old techniques and organization of work. Despite some technological innovations, such as the plow and seed drill, output per man was relatively small. In the late 19th and especially in the 20th century, output per farmer increased rapidly until, in the most advanced countries, a small minority of farmers supplied entire populations with food. These great changes were brought about by a series of advances in science and technology, including improved power sources, mechanical devices such as the reaper and combine, a scientific approach to plant and animal breeding, better food processing and preservation, more effective fertilizers and pesticides, and application of industrial management techniques to agriculture.

Factory farms. The organization of work developed in mass production industries was not transferred intact to agriculture. The tasks involved in running a farm change in accordance with the cyclical nature of the growing process and may vary greatly among different crops. They depend also upon the degree to which a given farm has been mechanized. The outstanding example of factory-type production in agriculture is in the U.S. poultry industry. A computerized feed bin, programmed with the exact quantities of nutrients required for health and quick growth, mixes the feed and delivers it automatically to the cage in which the bird is confined throughout its life. Water is delivered automatically, and waste is removed by mechanical means. When the chicken reaches precisely the correct weight for processing, that task is performed on an assembly-line basis. Application of these techniques has sharply reduced the cost per pound of chicken, and a form of protein that was once a luxury has become a staple item of diet. Similar methods are used to raise veal calves and other meat-producing animals as well. Capital investment in such factory farms is high, and production is carried on by giant companies.

American poultry farming

Migrant labour. The industrialization of agriculture meant that the small farm was being replaced by larger units, and this had profound consequences for agricultural labour. In the small-scale enterprise that had prevailed since antiquity, the farm family, with perhaps a few hired hands, had done all the work of planting, tending, and harvesting the crop, with neighbours joining to help one another during peak working periods such as the harvest. But the advent of industrialization drew workers from the farms to the cities, and the increase in mechanization required fewer farm labourers on a year-round basis. There was still need, however, for many hands during planting and harvesting, especially for fruit and vegetable crops that matured at the same time and still required hand harvesting.

In the United States the need for seasonal farm workers has been met—especially in Florida, Texas, and California (where approximately half the nation's fruits and vegetables are grown)—by migrant workers, largely from Mexico and the Caribbean countries, although some native-born Americans continue to follow the harvesting season north as temporary labourers. Few city dwellers, however, have deserted the cities to work on farms.

The employment of these seasonal workers raises a number of social, political, and economic problems. Migrants are typically paid low wages with no fringe benefits. Their living and working conditions remain far below standard. In spite of this, they often look to migrant farm labour as a means of escaping the worse conditions of their native countries.

The demand for migrant labour decreases as mechanization increases. In the United States, for instance, the harvesting of wheat and cotton, which required the work of many migrants before World War II, is now largely mechanized and done by regularly employed farmhands. In mature economies migrant labour contributes little to total agricultural output and a negligible amount to non-agricultural output. Nevertheless, the availability at the right time and place of migrant workers can be crucial, as without them large crop losses may occur.

Soviet farming. Agricultural mass production takes many forms. In the Soviet Union *sovkhozy,* or state agricultural farms, are owned by the government; farmers are, in effect, state employees, but work organization resembles that in the West. Soviet collective farms are in theory cooperative associations of farmers who combine their land and capital, sharing proceeds in common. Each family on a collective farm, however, is permitted to own a small plot of land, so that modern and traditional work organization exist side by side. Centrally located machine tractor stations own and operate most of the agricultural machinery, contracting to perform work on nearby collectives. These stations also function as coordinating units; their supervisors control and check on plowing, sowing, cultivation, and harvesting in their region. At the machine tractor stations, organization of work is functional with separate operating, maintenance, and supervisory staffs.

Collective farms

Although the Soviets at first prided themselves on their communal organization of agriculture, it has become evident that the system is not as productive as originally hoped. Despite its fertile soil, the Soviet Union is forced to import agricultural staples such as wheat from nations with capitalist agricultural systems. Indeed, most of the fruits and vegetables consumed in the country come from the small private plots of collective farmers, who, being allowed to grow produce for their own profit, have more incentive to work industriously for themselves than for the collective farms, with their government-set prices and production quotas.

Acknowledging the productive power of private initiative, the Soviet government in the 1980s began to loosen the constraints of collective agriculture. In 1989, individual farmers were given the opportunity to lease land and equipment for 50 years and more. The lessee could decide what to produce and at what price to sell it, and, upon his decease, his children could "inherit" the leased property.

Services. For most of recorded history, the vast majority of the world was engaged in farming. Beginning in the 19th century, industrial employment took primacy over agricultural work in those countries that had successfully industrialized. But in the United States and other highly industrialized nations, the service sector of the economy became the chief employer in the 20th century. By the 1950s the number of people engaged in service occupations in the United States exceeded the number of those employed in industry, and the proportion increased thereafter.

Work in the service sector is highly diverse, including everything from cleaning workers to business consultants, from truck drivers to financiers, from fast-food waiters to maîtres d'hôtel, from office clerks to advertising executives, from kindergarten teachers to university professors, from nurses' aides to surgeons, and government employees ranging from street sweepers and garbage collectors to legislators and heads of government.

Within the service sector, employment trends and working conditions changed throughout the 20th century. For example, the number of domestic servants declined drastically, with full-time, live-in domestic help almost disappearing and even part-time day servants also declining greatly. On the other hand, the number of government employees grew dramatically as government entities, from local to regional to national, took on tasks formerly ignored or left to the private sector.

SOPHISTICATION OF MASS PRODUCTION

Scientific management. According to Henry Ford, the assembly line was based on three simple principles: "the planned, orderly, and continuous progression of the commodity through the shop; the delivery of work instead of leaving it to the workman's initiative to find it; an analysis of operations into their constituent parts." A scientific approach to these principles, the next logical step in the organization of work, had already been enunciated by the American industrial engineer Frederick W. Taylor (1856–1915). From his work an entirely new discipline—industrial engineering or scientific management—emerged, in which the managerial functions of planning and coordination were elevated to a primary position in the productive process.

The engineering view of Frederick Taylor

In Taylor's view, the task of factory management was to determine the best way for the worker to do the job,

to provide the proper tools and training, and to provide incentives for good performance. Taylor broke each job down into its constituent motions, analyzed these to determine which were essential, and timed the workers with a stopwatch. With superfluous motion eliminated, the worker, following a machinelike routine, became much more productive. In some cases, Taylor recommended a further division of labour, delegating some tasks, such as sharpening tools, to specialists.

These studies were complemented by two of Taylor's contemporaries in the United States, Frank B. and Lillian M. Gilbreth, whom many management engineers credit with the invention of motion studies. In 1909 the Gilbreths, studying the task of bricklaying, concluded that much motion was wasted by the worker in reaching down to pick up each brick. They devised an easily adjusted scaffold that eliminated stooping and improved average work performance from 120 to 350 bricks per hour. Industrial engineering ultimately came to include all elements of factory operation within its compass—layout, materials handling, and product design, as well as labour operations.

Taylor regarded his movement as "scientific" because he attempted to apply scientific principles and measurement to the work process. Many previous advances in manufacturing had been made by applying scientific principles to machines in order to make them more efficient, and, through his minute subdivision of labour, Taylor sought to do the same to the work process itself. This scientific approach, however, neglected the human element, so that Taylor in effect converted the work process from a relationship between worker and machine into a relationship between two machines.

Scientific management theorists assumed that workers desired to be used efficiently, to perform their work with a minimum of effort, and to receive more money. They also took for granted that workers would submit without question to standardization of physical movements and thought processes. Their system, however, ignored human feelings and motivations, leaving the worker dissatisfied with the job. Furthermore, some employers omitted the altruistic elements in Taylor's system and employed time and motion studies to set high norms of production and speed up the production line while still keeping wages down.

Industrial psychology. In the decade after 1910, when the principles of scientific management were being applied wholesale in U.S. industry, union opposition arose. Though the unions approved more efficient production arising from better machinery and management, they condemned the speedup practice and complained in particular that Taylorism deprived workers of a voice in the conditions and functions of their work. Complaints were also made that the system caused irritability and fatigue along with physiological and neurological damage among workers. Misuse of the human element in production was causing declines in both quality and productivity. Industrial engineers then faced the problem of motivating the worker so that the combination of human labour and machine technology would achieve its fullest potential. A partial solution came from the social sciences, and, in the process, industrial psychology emerged.

The major premise of this new discipline was that mass production technologies affect the worker both in the immediate job environment and in relations with fellow workers and supervisors. The first important discoveries in the social context of mass production technology resulted from experiments made by the American social scientist Elton Mayo between 1927 and 1932 at the Hawthorne plant of the Western Electric Company, in Cicero, Ill. Mayo, who earlier had studied problems of physical fatigue among textile workers in a Philadelphia plant, was called in to the Hawthorne works, where industrial engineers were considering the potential effect on productivity of changes in illumination. The investigators chose two groups of employees working under similar conditions to produce the same part; the research plan was to vary the intensity of the light for the test group but to keep it constant for the control group. To Mayo's surprise, the output of both groups rose. Even when the researchers told one group that the light was going to be changed and then did not change it, the workers expressed satisfaction, saying that they liked the "increased" illumination, and productivity continued to rise.

Mayo saw that the significant variable was not physiological but psychological. A second series of experiments was performed, involving the assembly of telephone relays; test and control groups were subjected to changes in wages, rest periods, workweeks, temperature, humidity, and other factors. Output continued to increase no matter how physical conditions were varied; indeed, even when conditions were returned to what they had been before, productivity remained 25 percent above its original value. Mayo concluded that the reason for this lay in the attitudes of the workers toward their jobs and toward the company. Merely by asking their cooperation in the test, the investigators had stimulated a new attitude among the employees, who now felt themselves part of an important group whose help and advice were being sought by the company. The name Hawthorne effect was given to such beneficial changes in workers' attitudes, and, within a short time, scientific management incorporated these new findings.

The Hawthorne effect

Mayo's studies had suggested that consultation, usually in the form of interviews between labour and management, gave workers a sense of belonging to a team. Industrial engineers and sociologists have suggested additional approaches toward improving motivation and productivity. These include job alternation to relieve boredom; job enlargement, or having the worker perform several tasks on a project rather than performing a single operation; and job enrichment, redesigning the job to make it more challenging.

In a sense, Mayo's work made scientific management even more scientific, because he brought the new behavioral sciences, like social psychology, into the problems of organizing work and the labour–management relationship. It encouraged the development of human-factors engineering and ergonomics, disciplines that attempt to design "user-friendly" equipment accommodating itself to the human physiology and nervous system. For example, the new engineers try to make certain that a worker's equipment is operable with minimum strain, at a comfortable work level, and with controls easy to reach, see, and manipulate. In brief, they attempt to design the machine around the human mind and body.

AUTOMATION

The logical ultimate in the evolution of mass production processes is automation. In its ideal form, automation implies elimination of all manual labour and the introduction of automatic controls, assuring accuracy and quality beyond human skills. Perfect automation has never been attained, but sufficient equipment has been installed in many industries to alter greatly the pattern of employment. Tasks formerly performed by machine operators on a production line have come to require only maintenance personnel, engineers, office employees, production-control specialists, and some others. Although automation has been described as a "revolutionary" development, it is actually the end result of the trend of mechanization that began with the Industrial Revolution.

The word automation was coined in the 1940s at the Ford Motor Company and was first applied to the automatic handling of parts in metalworking processes. The concept acquired broader meaning when the American mathematician Norbert Wiener wrote about cybernetics, which he defined as control and communication in the animal and the machine. Wiener anticipated the application of computers to a number of manufacturing situations. His prediction that the introduction of automatic machinery would swiftly give rise to mass unemployment was popularized during the 1950s and '60s, causing considerable alarm. But automation was not introduced as rapidly as foreseen, and other economic factors have intervened to lessen the displacement of labourers.

Automation evolved from three interrelated trends in technology: the development of powered machinery for production operations; the introduction of powered equip-

ment to move materials and workpieces during the manufacturing process; and the perfecting of control systems to regulate production, handling, and distribution.

Transfer machines

Devices to move materials from one work station to the next included conveyor-belt systems, monorail trolleys, and various pulley arrangements. The transfer machine, a landmark in progress toward full automation, moves the workpieces to the next work station and accurately positions them for the next machine tool. The first known transfer machine was built by an American firm, the Waltham Watch Company, in 1888; it fed parts to several lathes mounted on a single base. By the mid-20th century, transfer machines were widely employed in the automotive industry, appliance manufacturing, electrical-parts production, and many other metalworking industries, in which they cut labour costs and improved quality by ensuring uniformity and precision.

Automatic controls represented an innovation when applied to all aspects of the production process. The simple cam, automatically adjusting the position of a lever or machine element, was an important control device in many early machines and, during the 19th century, was used to make many machine tools automatic. But cam devices have severe limitations in movement, number of changes, speed, size, and sensitivity. True automatic control cannot be attained unless the machine is sensitive enough to adjust to unpredictably varying conditions. This requirement demands the technique known as feedback, which the microchip computer can perform in a fraction of a second.

Some students of automation maintain that its primary goals are not necessarily increased productivity or cost reduction but product reliability and quality control. Other benefits promised by automation include reduction of waste, improved plant safety, and centralization of control. Still, the most visible initial effects of automation have been reduction of costs and increases in productivity.

The industrial workplace. *Productivity.* Whereas the earlier phase of the Industrial Revolution had resulted in assembly lines mass-producing identical parts for mass markets, the introduction of the computer allowed for custom-made, small-batch production. For example, in the United States the chief investment in plants and equipment in the 1980s went into information technology, such as computers and telecommunications equipment. Such aids have allowed American manufacturers to concentrate on "niche" production—that is, supplying a limited segment of the market with a specialized product and responding speedily to changes in market demand. On the automobile assembly line, niche production enables many cars containing different options demanded by buyers to go down the same assembly line, with the computer making certain that the proper items go into each separate car.

Niche production

These potentialities of automation have created two new fields: computer-aided design (CAD) and computer-aided manufacturing (CAM), often linked as co-disciplines under the title CAD/CAM. In a sense, CAD/CAM allows the mass production system to manufacture customized, "handmade" articles. The machinery can be adapted to a particular product through computer programming, enabling work on small batches to achieve many of the economies previously available only through mass production of identical objects. Computer-aided design itself makes possible the testing of production methods and the design of the product by running tests (of such factors as ability to withstand stress, for example) through the computer. If necessary, the product design or the process can be modified without going to the expense and time required for building actual prototype models.

Automation not only gives flexibility to production, but it also can cut down costly lead times in changing from one production model to another, and it can control inventories to provide a continuous flow of materials without expensive storage requirements or investment in spare parts. Such efficiencies lower production costs and help explain the growing strength in world markets of the Japanese, who first introduced the practice. Automation has also fostered the development of systems engineering, operations research, and linear programming.

Automation has not yet realized the dream of completely robotized production. The first generation of industrial robots could perform only simple tasks, like welding, for they became confused by slight irregularities or differences in the objects on which they worked. To overcome that difficulty, computer scientists and engineers began developing robots with keener sensitivity, thereby enlarging their capabilities. Although progress has been made, it is clear that human beings must be available to back up the robots and maintain their productivity.

When automation was first introduced, each of the robots involved in complicated processes was controlled by a microcomputer programmed to perform only one task. As these processors could not communicate with one another and lacked memory, it was hard to trace which particular processor was responsible when something went awry. In the late 1980s this deficiency was corrected by the introduction of general-purpose computers, which bring together data from all the microprocessing units and make them accessible on one screen. This sped up the detection of problems and reduced the downtime of machinery, thereby increasing productivity and lowering expenses.

Effect on skilled labour. Robotic machines can perform certain unpleasant and dangerous jobs, such as welding and painting, that can be injurious to a worker's health. They can handle loads of up to a ton or more and work efficiently in temperatures ranging from near freezing to uncomfortably hot. Automation has eliminated much of the worker's physical and mental drudgery and has allowed the worker to change from a machine operator to a machine supervisor.

Reduction of drudgery

At the same time, by increasing productivity as measured in output per man-hour, automation can reduce the number of workers. In the 1950s and '60s employment declined in the chemical, steel, meat-packing, and other industries in developed countries that achieved large increases in output. Except in certain older industrial areas in Britain and the United States, however, the widely feared onset of mass unemployment did not materialize. Although certain jobs and skills have been rendered obsolete, a vast array of new jobs calling for different skills has grown up.

Automation has brought about changes in the worker's relation to the job. Here the differences between labour practices in different countries proves instructive. The old management principle that work should be broken down into the smallest operations, so that the worker would not have to use any intelligence in performing a job, was based perhaps upon the notion that the worker is stupid. Hence, when full mechanization was introduced into American factories, the workers were not permitted to stop the moving assembly line if anything went amiss; that was presumed to be the task of supervisory engineering personnel. The result was both low productivity and a loss in quality control. In Japanese factories, on the other hand, assembly-line workers were allowed to stop the process when something went wrong. Indeed, the Japanese companies formed "quality circles," wherein the workers were given a say in the performance of their tasks and in the process of problem solving—an application of Mayo's Hawthorne effect, which they had learned from American management consultants. These practices improved both productivity and quality.

A similar way of enhancing quality and work performance is what is known as group assembly, which started in Swedish automobile plants and was also adopted by the Japanese and then by the Americans. With this system a group of workers is responsible for the entire product, rather than individual workers doing only one small task. If something goes wrong on an assembly line, an individual worker can push a button and hold things in place until the problem is resolved. This approach to production is being increasingly employed throughout the world. It already has had major implications for the labour force and labour–management relations. For one thing, it allows smaller numbers of more highly skilled workers, operating sophisticated computer-controlled equipment, to replace thousands of unskilled workers in assembly-line plants.

As a consequence, the highly skilled worker, who began

to disappear with the introduction of the old-fashioned mass production assembly line, again became indispensable. The increasing use of automated machinery and control systems placed new demands on both the technical skills and the intellectual aptitudes of production workers. While automation may have eliminated many unskilled jobs, it increased the demand for highly skilled mechanical labourers and knowledgeable technicians who could operate the newer automated devices. As a result, the early prophecies that automation would reduce the need for workers' skills have proved to be the contrary of what has been happening. Automation may be seen as improving efficiency and expanding production while relieving drudgery and increasing earnings—precisely the aims of Frederick W. Taylor at the turn of the 20th century.

Need for skilled labour

The office workplace. The introduction of computers also affected the organization of work in the information sector of the production economy. File clerks, bookkeepers, and other skilled office personnel involved in information processing were replaced by semiskilled keypunch and tabulating-machine operators. Office automation represents a further mechanization of office work, a process that began with the typewriter and the adding machine in the 19th century.

The information flow in offices has been likened to the movement of materials in manufacturing. Information, like materials, must be stored; typing or keypunching changes the form of the information, just as a machine operation changes the form of the workpiece; the value of the finished product is changed by adding information to it; and there must be a measure of quality control to make certain that the information is accurate. Just as automated machinery has done away with the jobs of many machine operators, integrated information-processing systems have eliminated many clerical tasks. For the production operation, automation provides an exact control over the inventory of raw materials, parts, and finished goods. Applied to billing operations in the office, it often can drastically reduce accounting costs.

The combination of computers and telecommunications led some to believe that office workers would perform their required functions without leaving their homes, as the computer terminal would take the place of their usual paperwork. Such predictions generally have not materialized, however. Social psychologists explain this by pointing out the social aspect of the work process, in the office as well as on the assembly line. Office workers have revealed their nature as "social animals" who enjoy the companionship of their fellow employees at the workplace.

The social aspect of work

Nevertheless, office automation affects management–worker relationships in a number of ways. For middle-level employees it means that higher management can have the reports of production, costs, and inventory at their fingertips and on the computer screens at their desks instead of depending on their subordinates for information. Automation also gives managers the means to monitor the efficiency of office workers in a way hitherto impossible. Through computerized information they can, for example, count the number of times per hour that a typist strikes a letter on the keyboard or ascertain the number, times, and nature of a worker's telephone calls.

Women in the work force. For most of written history, agriculture was the chief human occupation, and heavy physical labour was not confined to men. Women performed backbreaking chores like grinding grain by hand in a stone quern, drawing and carrying water, gathering wood, and churning milk to make butter. About the only time women were given a respite was immediately before and after giving birth.

With the Industrial Revolution the work situation changed for both men and women. Whereas the hearth and home had formerly been the centre of production and family life, with industrialization the factory became the workplace and the town the dwelling place. The role of women in the family work force did not change overnight, however, for at first whole families were employed in factories almost as teams.

Family work teams

Not until the mid-19th century did the role of the male as the "good provider" emerge, with women relegated to household and domestic tasks. This was caused, perhaps, not so much by the change from farming to industrial technology as by a growing humanitarian protest against the harsh treatment of women and children in the early factory system. Legislation—most notably in Britain—raised the minimum age for child labour in factories, set limits on the working hours of women and children, and barred them from certain dangerous and heavy occupations. Thus, more women were engaged primarily in domestic tasks and child care at home while the men went out to work. Being the sole wage earner in the family reinforced the man's traditional position as the head of the family, who was expected to direct the lives of his dependents. Whatever authority the woman exercised derived from her status as mistress of the household.

The tradition of the deferential female wedded to motherhood and domesticity persisted throughout the 19th century and well into the 20th. After the coming of electric power near the close of the 19th century, advances in technology brought labour-saving devices such as washing machines and vacuum cleaners into the home. Although they changed the nature of some tasks and freed the housewife from some drudgery, they did little to lessen the amount of time she spent on household duties.

At the same time that women's household work was being transformed by technology, changes were occurring in the economy and society leading to the growth of administrative and service jobs. The extension of public education created a demand for more teachers, and growing industrial and commercial enterprises required more office workers and salespeople. While men could have—and earlier had—performed teaching and clerical tasks, it was discovered that it was cheaper to hire women for these occupations. The prevailing attitude was that, since men were the heads of households, they had to be paid enough to support a family. On the other hand, most women who entered the work force in the United States before World War II were single and did not have families to support; hence, they could be paid lower wages. This inequality in men's and women's pay scales, even for equal work, still prevails.

Unequal pay for women

Technological changes affecting the maternal role of women also had an impact on their ability to leave the home. The development of contraceptive devices and the baby bottle and rubber nipple gave women greater freedom, making it easier for them to take outside jobs. Even if her employment was interrupted by childbirth, the woman could, within a relatively short time, reenter the work force.

Many working women performed tasks closely related to their traditional household work. When clothes were less often made at home but purchased ready-made at stores, for example, women were hired as seamstresses in the clothing industry. Even after national emergencies such as the world wars, during which women were encouraged to take manufacturing jobs to replace the men who were in military service, women returned to housekeeping or to traditionally female occupations such as office work and nursing.

In the 1970s married women began entering the labour force in great numbers, and the strict segregation of women into certain occupations began to lessen somewhat as new opportunities arose for female workers in traditionally male occupations. New technology has meant that many tasks that once required heavy physical exertion, and hence were restricted to men, can now be performed simply by pushing buttons. Operating a bulldozer, for instance, does not need muscle power so much as alertness, judgment, and coordination—qualities as plentiful in women as in men. Nevertheless, the entrance of women into occupations formerly the province of men proved to be slower than expected. This persistent occupational segregation by sex is largely responsible for sizable differences in rates of pay that still exist. It would appear that, although rapid technological progress has enabled women in highly industrialized nations to cast off certain traditional roles, technological determinism—or technological rationality—does not always prevail over cultural views and social practices inherited from the past. (M.Kr.)

Delay in entering male occupations

INDUSTRIAL AND ORGANIZATIONAL RELATIONS

Industrial relations, or organizational relations, is a subject of study concerned with the behaviour of workers in organizations in which they earn their living. Scholars in this field attempt to explain variations in the conditions of work, the degree and nature of worker participation in decision making, the role of labour unions and other forms of worker representation, and the patterns of cooperation and conflict resolution that occur among workers and employers. These patterns of interaction are then related to the outputs of organizations. These outputs span the interests and goals of the parties to the employment relationship, ranging from workers' job satisfaction and economic security to the efficiency of the organization and its effects on the community and society.

Worker, manager, and society

CONCEPTIONS OF THE WORKER

19th- and 20th-century views. In classical economics, workers were treated like commodities, subject to the natural laws of supply and demand. Although classical economists readily acknowledged that workers are not motivated by money alone, they abstracted out of reality only the economic factors. This led them to consider workers as undifferentiated and passive instruments in the production process.

Karl Marx in the mid-1800s was the first to challenge this view of labour. He rejected the notion that workers should bear the costs of market forces and argued that all the value of production comes from workers' input; therefore, he insisted, labour should own the means of production. Since under a market system the means of production are owned and controlled by capitalists, workers would be exploited. Eventually, according to Marx and his followers, the injustice of this exploitation would lead to a revolutionary overthrow of the capitalist system and its replacement by a socialist state.

Later, around the turn of the century, two British political economists, Sidney and Beatrice Webb, joined this debate by arguing that a combination of worker and community forces would gradually achieve a socialist state. They shared with Marx a belief that workers and employers are separated by class interests and that only by organizing into trade unions would workers amass the bargaining power needed to improve their economic and social conditions. They did not believe, however, that a revolutionary overthrow of the capitalist system is necessary for social progress. Instead, through union representation, collective bargaining, and legislative protections, worker, employer, and community interests would eventually be harmonized.

American institutional economists

About the same time the Webbs were developing their views in Britain, an American view was taking shape under the work of John R. Commons and his associates at the University of Wisconsin. Unlike classical economists, these institutional economists believed that the laws of supply and demand can be influenced by the policies, values, structures, and processes used to govern employment relationships. Like Marx and the Webbs, Commons rejected the commodity view of labour and believed that an inherent conflict of interests separates workers and employers. He also believed, however, that these conflicts are a natural and legitimate part of any employment relationship and would not disappear if capitalism were replaced by socialism.

Consequently, these American scholars and social activists emphasized the importance of legislation that protects workers' interests in safety and health, unemployment and workers' compensation insurance, minimum wages, and retirement benefits. They also believed in the value of labour unions and in the need for negotiation and compromise between workers and employers. These institutional economists not only served as the forefathers of modern industrial relations, they also provided many of the ideas behind the labour legislation enacted as part of President Franklin D. Roosevelt's New Deal in the 1930s.

The advent of industrial relations. The New Deal paved the way for the development of modern industrial relations research and practice. In response to the economic and social crisis of the Great Depression, the U.S. Congress and the Roosevelt administration enacted a series of laws granting workers the right to organize into unions and to engage in collective bargaining with employers. Other New Deal legislation set minimum wages and provided a system of unemployment insurance and social security. In subsequent years unions organized large numbers of workers in the growing manufacturing, transportation, and communications industries. By the end of World War II nearly one-third of all American workers were union members.

After the war an exceptionally large number of strikes brought national attention to the problems of labour–management relations and led to the formation of a number of industrial relations research and teaching programs at leading American universities. The goal of these programs was to draw together the theories and insights of economists, labour and management specialists, and other social scientists to find ways to encourage greater cooperation and improved conflict resolution among workers and employers. Thus, the modern field of industrial relations was born.

Studies of worker behaviour. *Scientific management.* While Marx, the Webbs, and Commons focused on the role of labour in the late 1800s and early 1900s, others were developing theories of management. Frederick W. Taylor's engineering approach, later known as scientific management, was similar to that of the classical economists in regarding workers as passive instruments of production, but he did recognize differentiation among workers, at least insofar as degrees of skill were concerned. He developed methods for time-and-motion studies to determine the elements of particular jobs and the way in which these elements should be put together for the greatest efficiency. His focus was upon the individual worker; there was no place in his model for group membership or for the effects of groups upon individual behaviour.

Industrial psychology. A step further in the recognition of differentiation among workers came with the emergence of industrial psychologists, who are concerned with the measurement of the skills and aptitudes of individuals. At least in the early stages of these developments, workers were viewed as isolated individuals, and no attention was given to group phenomena.

Human relations. In the 1930s the emphasis of management researchers shifted from individuals to the work group. Of primary importance was the human relations research program carried out by Elton Mayo and his associates at the Hawthorne Western Electric plant and their discovery of the "Hawthorne effect"—an increase in worker productivity produced by the psychological stimulus of being singled out and made to feel important. The ideas that this team developed about the social dynamics of groups in the work setting had lasting influence.

Although the interpretations that Mayo and his group made of the test results have come under critical fire, the concern here is not with the validity of the conclusions but rather with the effects of the Hawthorne studies upon the subsequent development of management and organizational behaviour. Publication of some of Mayo's writings and of the major research report in *Management and the Worker,* by F.J. Roethlisberger and William J. Dickson in 1934, attracted large numbers of sociologists, social psychologists, and social anthropologists into a field that had previously been limited to economists, engineers, and industrial psychologists.

Four general conclusions were drawn from the Hawthorne studies:

1. The aptitudes of individuals (as measured by industrial psychologists) are imperfect predictors of job performance. Although such measures may give some indication of the physical and mental potential of the individual, the amount produced is strongly influenced by social factors.

2. Informal organization affects productivity. Although previous students of industry had looked upon workers either as isolated individuals or as an undifferentiated mass organized according to the formal chart of hierarchical positions and responsibilities established by management, the Hawthorne researchers discovered a group life among the workers. The studies also showed that the relations that supervisors develop with workers tend to influence the manner in which the workers carry out—or fail to carry out—directives.

3. Work-group norms affect productivity. The Hawthorne researchers were not the first to recognize that work groups tend to arrive at norms of what is "a fair day's work," restricting their production below that point even when they are physically able to exceed the norm and would be financially rewarded for it. However, the Hawthorne study provided the best systematic description and interpretation of this phenomenon.

4. The workplace as a social system. The Hawthorne researchers came to view the workplace as a social system made up of interdependent parts.

Behavioral science. Behavioral scientists had made their entry into the field by attacking the then-prevailing oversimplified notions of the individualistic economic person and the formalistic engineering notions of organizational structure, technology, and efficiency. As often happens in arguments between members of competing schools of thought, the force of the behavioral science attack carried some of its proponents so far as to view the work organization as simply a system of social relations and to downplay the role of economic forces. During the 1950s and '60s the field underwent a major process of redefinition, which consequently affected conceptions of the worker.

Behavioral scientists now recognize the importance of economic factors, but they see material rewards as having an effect upon behaviour in combination with social and psychological factors, and they study the pattern in this combination. Thus, over the years behavioral scientists have deepened the understanding of the role played by interpersonal, structural, and technological forces in organizations and industrial relations.

CONCEPTIONS OF THE MANAGER

Classical economists made no distinction between the manager and the entrepreneur, the person who brings together land, labour, and capital and puts them to work. This distinction did not take hold in the literature until the appearance in 1933 of the classic study by Adolf Berle and Gardiner Means, *The Modern Corporation and Private Property.* When the authors demonstrated that in most American corporations the owners (that is, the stockholders) played no direct role in the management of the concern and that the managers generally had insignificant holdings of stock, it became apparent that theories of entrepreneurial behaviour had little to contribute to the understanding of the behaviour of managers.

Specialized management. Somewhat earlier Max Weber, a German economist and sociologist, had approached the study of managerial behaviour through his concept of bureaucracy. To Weber the term bureaucracy did not have the negative connotation often heard in casual conversations. He used the term simply to point to a phenomenon of growing importance even in his time: the large organization with fixed positions linked together in a hierarchical pyramid, with specialization and division of labour, and with established rules and regulations governing behaviour. To Weber the manager was the individual who interpreted and applied the rules of the organization.

Later organizational sociologists, though recognizing the importance of Weber's contribution in focusing attention on the impersonality and rationality of modern industrial and governmental organizations, pointed out that Weber's model failed to take into account some of the most important features of the modern business organization. They argued that it gave an unduly rigid picture of organizations, that it failed to devote attention to processes of change, and that it built so exclusively on the hierarchy of authority as to neglect relations not explicitly defined by the structure. In any case, Weber's formulations were of interest primarily to social scientists. Practicing managers and students in business schools at that time were likely to have little familiarity with the Weberian approach to managerial behaviour.

The early model of the manager taught in American business schools applied these formal structural concepts by emphasizing functional specializations. In these terms the manager was the one who had mastered such subjects as accounting, marketing, production, finance, and so on. Later it was recognized by theoreticians and practicing managers alike that management was a good deal more than the sum of these specialized functions, and this realization in turn led to the conception of the manager as generalist, who would understand the various specialized functions and the people engaged in them. The emphasis turned to decision making, leadership, and the relation of the firm to its environment.

Participative management. In 1960 Douglas McGregor of the Massachusetts Institute of Technology published *The Human Side of Enterprise,* a classic book that had a major impact on subsequent management education and practice. McGregor challenged many of the prevailing managerial assumptions about worker motivation and behaviour. According to the prevailing view, which he labeled "Theory X," workers were viewed as uninformed, lazy, and untrustworthy members of the organization. Management's task was to control workers and motivate them through a combination of control systems, fear of discipline or dismissal, and organizational rules. McGregor contrasted this with a "Theory Y" assumption, namely, that workers are highly motivated and can be trusted to contribute to the organization's objectives if given the opportunity to participate in organizational decision making. Out of the work of McGregor and others, such as Rensis Likert, has evolved "participative management," a process in which managers consult with and involve subordinates in organizational problem solving and decision making.

McGregor's Theory Y

McGregor's views were supplemented by theories that argued for the importance of integrating considerations of workers and their social environment into the design and implementation of new technologies and production systems. These sociotechnical concepts originated in Europe and had substantial impacts on the design of innovative work systems in Scandinavia in the 1960s and '70s. By the early 1980s they had achieved significant acceptance and use in American firms.

Sociotechnical and worker-participation models of decision making have grown in importance and popularity in response to the rise of global competition and the rapid pace of technological change. Most contemporary organizational and industrial relations scholars have concluded that the full potential of new information and manufacturing technologies can only be realized through management processes that support participation and communication across functional lines and departments, flexibility in how work is organized, and effective problem solving. Yet there is still considerable debate among practitioners over the feasibility or wisdom of involving workers in organizational decision making. Therefore, vestiges of both Theory X and Theory Y concepts and practices can be found in organizations today.

RESPONSIBILITY TO SOCIETY

The debate over the appropriate role for workers in organizational decision making is part of a larger debate over the extent of the firm's responsibilities to its community and society. This debate has been going on since the days of the Industrial Revolution.

The Industrial Revolution brought about great accumulations of wealth and also focused public attention on the apparent negative effects of rapid industrialization on working people. To what extent workers in the new factories were worse off than they had been in the much smaller-scale cottage industries may be a matter of continuing debate, but there is no question that large concentrations of workers—men, women, and children—crowded together in oppressive physical conditions and working long hours for low pay made the problem much more publicly visible. In the earlier period workers had dealt with owners

and agents of whom they had some personal knowledge. The establishment of large factories destroyed that direct relationship, and it became less credible for the owner to claim that he took a personal interest in his workers.

Relationship between managers and workers

Two models of social responsibility. In the past two centuries managers of industry have taken, in general, two broadly different positions regarding management's social responsibilities:

Laissez-faire. The first stance represents a sort of combination of laissez-faire economic theory and the Protestant ethic. In this view the owner or manager has no responsibility for the welfare of the workers outside the immediate plant situation; a person's station in life is a reflection of his intrinsic merit in the eyes of God; the wages and other labour costs incurred by the firm are the result of competitive market conditions. In this view, then, the owner's or manager's responsibility to his employees begins and ends with operating the firm in such an efficient manner that it is able to meet competition in the marketplace, and, if all business managers similarly followed a policy of intelligent self-interest, the broad social interests of society would be better served than by any other policy.

The expression of this point of view has undergone changes in style over the years. Today one hardly expects business leaders to state the position with religious overtones, and even the executive most inclined toward a laissez-faire viewpoint is likely to concede that there are some social problems that are not resolved by private initiative in pursuit of enlightened self-interest. However, managers with this view of the world tend to take a defensive position regarding the responsibilities of their firm beyond the gates of the plant. They recognize that popular opinion and government policies and programs may require them to take on activities not dictated by immediate material interest, but the tendency is to do what has to be done to keep out of trouble with the outside world and nothing more.

Paternalism. The other stance begins with the assumption that management has a social responsibility to the communities in which its plants are located. If one states the situation in this general way, hardly a management spokesperson today would deny this social responsibility. Yet, when one gets beyond rhetoric, one finds a wide variety of views as to what actions—if any—management should take. In assessing the present scene, one might do well to examine the historical evolution of conceptions of management's social responsibilities.

In the early part of the 19th century, the industrialist and social reformer Robert Owen was the first manufacturer to back up words about management's social responsibilities with a program of action. Having risen out of the work force in a textile mill himself, he was concerned with the social and economic conditions of workers and believed that the economic success of an enterprise did not have to depend upon exploitation of labour. In the mill town of New Lanark, Scot., Owen built workers' housing, schools, and a store that were far superior to contemporary standards for workers' communities. His philosophy was influential in the development of the cooperative movement in England.

Owen's ideas and the successful operation of his plant and community during his lifetime impressed many social reformers and the business community as well. His influence was clearly visible in the establishment of the industrial city of Lowell, Mass. Francis Cabot Lowell had made a trip to England and Scotland to study textile mills and related community problems before launching his own enterprises in Massachusetts. He had found New Lanark far more in harmony with American ideals regarding the dignity of the individual than was the average English industrial plant of the time. Lowell faced a social problem of an immediate practical nature: he had to recruit a labour force, largely female, not available in the towns where he was building his plants. To meet this need, the firm built, in what came to be the city of Lowell, a number of boardinghouses especially for young women. Each house was under the control of a woman who was supposed to ensure the morality of her charges, and the young women were not allowed out of the house after 10 PM except with special permission. In addition, Lowell made liberal provisions for the building of schools and churches. He and his associates also gave stimulus to the Middlesex Mechanics Association, which sponsored cultural and educational programs.

Housing for female textile workers

Later in the century George M. Pullman undertook to build around his Pullman Palace Car Company a complete community (the town of Pullman, now part of Chicago) that would house all the employees and provide for all the essential facilities. In the early period of the Pullman Company, the quality of worker housing was notably superior to that of most other industrial workers.

When Henry Ford startled the industrial world with his announcement of the $5-a-day wage in 1914, he followed it with steps designed to help workers make good use of their increasing affluence. The company already had a small legal department set up to help workers with the complicated problem of home buying, and then Ford established what he called a sociology department. It was staffed with social workers who made home visits to workers' families to provide advice and help on family problems. Members of the department were also free to talk with workers within the plant during working hours in efforts to straighten out family problems.

How does one assess the success of such paternalistic efforts? In the United States the Lowell project was the longest lived and the most admired by foreign visitors. Charles Dickens compared Lowell very favourably with the typical English industrial city. Nevertheless, the distinctive character of Lowell had been lost before the end of the 19th century. Native New Englanders were replaced in the work force by the new immigrants and their children, and the women of these immigrant families had no rural homes to return to after several years of labour in Lowell, as had the New England girls. The boardinghouses thus lost their primary rationale, and by 1900 all of them had been sold by the companies. Even in the early years of the Lowell textile industry, conditions could not have been quite so idyllic as described by some of the visitors, for there were nine recorded strikes or lockouts in the mills in the period 1834–79.

The strikes greatly increased in frequency after this period, when management was faced with changes in social composition of the work force brought about by successive waves of immigration. One of the most famous of these was the "Bread and Roses" strike in 1912 of women textile workers in Lawrence, Mass. It got its name from the banner carried by some of the striking women indicating that they wanted both respect and improved working conditions as well as higher pay. It would seem to be more than coincidental that some of the most bitter strikes in the United States—from Pullman in 1894, through the Southern mill towns in the 1930s, to Kohler, Wis., in the 1950s—have taken place in company towns. (In fact, the Kohler strike holds the record as the longest strike in American history, lasting from 1954 to 1965.) Whatever economic grievances workers have had in these situations, it is clear that economic exploitation is not a complete explanation of the bitterness of the disputes. Whether benevolent or oppressive, the manager in the company town has far more extensive control over the lives of the workers than is found in other towns and cities. In the company town the owner or his agents not only run the plant; they also run the government, provide the local services, run the hospital, and so on. Any grievance a resident of the town may have is seen to be the fault of the company. When clashes in labour relations do arise in these situations, they are thus likely to look more like a struggle for independence than a standard union–management dispute.

Strikes in company towns

The rise of unions in the mass production industries of the United States in the 1930s helped to persuade executives that a paternalistic approach to labour and community relations was no longer feasible. Extensions of management's social responsibilities were now achieved through collective bargaining. Still, these broader benefits, such as pensions and health insurance, were limited to the workers and their immediate families. There was a tendency to assume that any responsibility for the wel-

fare of the community as a whole should be assumed by government.

Modern objects of social concern. Although management's withdrawal from paternalistic community responsibility was accelerated by the growth of unionization, company officials began increasingly to see the need for the company to take part in the affairs of the communities where its plants were located. As the plants became part of large corporate organizations with headquarters in distant cities, company executives found themselves increasingly estranged from the local people. A common reaction to this estrangement was the development of public relations and community service programs. The plant manager's role was redefined to include the function of representing the company in the community, and this meant participating in community activities along with locally based business and professional persons. Thus, it became common to find the manager and other plant executives playing prominent roles in community fund drives and other service programs. Such activities, however, involved simply a commitment to maintaining the community rather than any initiative toward changing it.

In the 1960s managers began to recognize that the community service orientation was not adequate for coping with the problems of cities that were erupting in violence. By this period the leaders of the powerful unions that were negotiating with the leading companies had come themselves to be part of the new industrial elite. No longer were the unions seen as leading a broad movement for social change and a possible radical reordering of society. They came increasingly to be viewed as organizations devoted to the defense and improvement of status of workers who held a relatively privileged position in society compared with the "underclass" of workers who were employed in low-wage situations, who worked only sporadically, or who were unemployed. It became evident that the acute and chronic problems affecting society as a whole and industry in particular were not going to be solved simply by having a "good" management relating to its union and indulging in "good" relations with the communities.

Societal problems

Plant closings. Debates over a firm's responsibilities to the communities in which its establishments are located intensified in light of the increased number of plant closings and corporate restructurings that occurred in the United States in the 1980s. Many of these closings resulted from intensified international and domestic competition; others were caused by technological changes; and still others reflected shifting corporate business strategies and priorities. Regardless of the cause, community leaders argued that employers should provide their workers and communities advance notice and should work with employee and community leaders either to avoid the closing and job loss or to ease the process of worker and community adjustment. These pressures were voiced most strongly by leaders in the older industrialized cities of the Midwest and Northeast that were hardest hit by the migration of manufacturing jobs to the South and to countries with a lower wage scale.

Many companies responded to these calls by providing advance notice voluntarily or through collective bargaining and by working cooperatively with community and union representatives to ease the transition. In a small number of cases workers and unions took the initiative by purchasing businesses and transferring them into employee-owned enterprises. The record of this latter strategy is at best mixed. Some of these plants survived and thereby continued to provide jobs, albeit often at reduced wages and pension benefits. Others, however, failed for the same reasons that prompted the original owner to want to abandon or move the business.

Concern over this issue led nearly all industrialized countries to enact legislation requiring companies to notify workers and communities of impending plant closings or mass layoffs.

Demographic changes. Over the years employers have had to broaden the scope of their responsibilities in answer to changes in the demographic makeup of the labour force and to various social issues that affect the employment relationship. For example, in the 1960s the United States enacted a series of equal employment opportunity laws, which forbid discrimination in employment on the basis of race, colour, creed, sex, age, or handicapped status. Companies that do business with the federal government have an additional obligation to demonstrate that they have taken affirmative action to provide job opportunities for women and minorities.

Efforts to eliminate discrimination

Since the 1960s, therefore, firms have carefully reexamined and upgraded their recruitment, selection, training, and promotion policies to eliminate discriminatory practices. The evidence on the job market status of blacks and women shows that, while these legislative and company-level initiatives have helped to reduce the income and employment differentials of blacks and women, sizable gaps in wages and occupational status still remain. Moreover, while most employers have eliminated overt forms of discrimination from their formal personnel policies, many observers believe that there is still considerable subtle discrimination that holds back women and minorities in organizations. Research has shown, for example, that some managers tend to systematically bias their performance evaluations of women or minorities. Others unconsciously hold lower expectations for women or minorities or are uncomfortable in dealing with them as equals or superiors.

Because of their subtle nature these forms of discrimination are especially hard to eliminate from organizational life. New practices such as the use of mentors (senior managers who look out for and provide career advice for junior employees), ombudsmen (third-party neutrals who help to solve conflicts and resolve problems in organizations), and peer support groups or networks are now being used in leading firms to address the remaining discriminatory practices that hold back the full utilization of all members of the work force.

Stakeholder versus stockholder. Debates over the scope of corporate responsibilities have raised an important theoretical question that goes to the heart of the purposes and roles of the modern corporation. Is the corporation's only, or primary, responsibility simply to maximize the value of its shareholders' investment? Or is it more appropriate to recognize that the modern corporation has a responsibility to serve the interests of multiple stakeholders, including not only its stockholders but also its employees, communities, customers, suppliers, and the broader society in which it is located?

This question takes on added importance in a world where capital is more mobile, corporate takeovers have increased in number and frequency, and multinational corporations may have, by definition, multiple national allegiances. A stakeholder model would argue for development of new forms of corporate governance where these multiple interests are represented in organizational decision making. Workers might, for example, be represented on corporate boards of directors, as they are in many European countries and in a small number of American firms where employees either own substantial stock in the corporation or have taken significant wage concessions to save the company and their jobs.

American managers are often criticized for their failure to adopt a long-run view of the corporation's objectives, following instead the strong tendency built into financial markets and institutions to maximize short-run profits. The adoption of a stakeholder view of the corporation's responsibility would challenge this tendency. The evidence most often cited for this belief comes from comparisons between the behaviour of American and Japanese firms. Japanese executives appear to take a longer and broader view of corporate objectives than do their American counterparts. One reason (but not the only one) for this is that the cost of capital is lower in Japan than in the United States. Therefore, Japanese firms can achieve more favourable economic returns for long-term investments than can comparable American companies. This allows Japanese executives to invest in new products and processes that will build and protect a stable employment base and allow the firm to grow over the long run, even though short-term profits may be sacrificed.

Japanese and American views

A stockholder wealth-maximizing model would argue for a more laissez-faire, or free market, approach by limiting government regulations and worker or community repre-

sentation in corporate decision making. While the long-run trend has been to broaden gradually the scope of the corporation's responsibilities to its communities and work force, corporate law still treats the maximization of shareholder wealth as the primary responsibility of firms and their top executives and directors.

The changing work force

In the past when one wanted to describe the demographic and social characteristics of the work force and the career patterns of its members, it was common to divide individuals into two categories: managers, or "salaried" employees, and workers, or "hourly" employees. The laws governing employment practices still make this distinction, as salaried employees are "exempt" from much of the wage-and-hour legislation that governs the rights of "nonexempt" hourly employees. The increasing diversity, however, in both the characteristics of the labour force and the organization of work make these categories less helpful.

Consider, for example, the following projections issued in 1987 by the U.S. Bureau of Labor Statistics for the future up to the year 2000. Over 85 percent of all new entrants to the work force will not be white males. Instead they will be women, minority group members, and immigrants. A majority (55 percent) will be women. At the same time, most of the new jobs will be created not in the large manufacturing firms once thought of as the typical place of employment but in the service sector. In addition, the majority of these new jobs will be created by small rather than large firms. Furthermore, the educational requirements of the "typical" job are expected to continue to increase.

Taken together, these trends worry many industrial relations and personnel experts and managers. The fear is that there may be a mismatch developing between the characteristics of future labour force entrants and the types of jobs that will be in high demand. If this is true, considerable efforts will be required to improve the match between supply and demand. This in turn implies that individuals will need to engage in a lifelong learning, training, and retraining process and that firms will need to increase their investment in training. The changing nature of the labour force further implies that employers will experience intensified pressure to open up opportunities for advancement to women, minorities, and immigrants.

INTERESTS, VALUES, AND EXPECTATIONS

The interests, values, and expectations that workers bring to the workplace provide a useful point of departure for understanding how employees respond to managerial policies. While these psychological features vary among individuals, over time as workers move through different stages of their family and career cycle, and across nationalities, they do reveal certain similarities.

Assessing workers' interests. There is a long-standing debate between psychologists and economists over how best to ascertain worker interests and expectations. Psychologists have traditionally used survey questionnaires and interviews to measure worker attitudes, values, and beliefs and then examined the relationships of these attitudes to observable workplace behaviours such as job search, turnover, absenteeism, union organizing, and withdrawal from the labour force. The value of this approach is that it provides a direct measure of an individual worker's expressed attitudes and beliefs. Economists favour direct observation and measurement of these observable behaviours. This provides evidence of what economists call "revealed preferences," preferences that are revealed by actions taken. Both approaches are helpful in painting a complete picture of workers' views and the workplace outcomes that result from these views.

Since work is the most important source of income, it is no surprise to find that all workers place a high value on the income and security provided by their jobs. Both survey evidence and labour market behaviour demonstrate that workers expect their jobs to provide both adequate and fair compensation. Fairness, or equity, is normally assessed by comparing one's wages and fringe benefits with those of others in the same occupation, area, industry, or organization. Failure to provide adequate and equitable wages has consistently been shown to lower workers' job satisfaction and to increase the likelihood that workers will prefer either to look for another job or to act on the job to increase wages through organizing a union or striking. Furthermore, there is no evidence that the expectation of high and equitable wages weakens as individuals move up the occupational ladder and receive higher pay. Even among professionals, pay dissatisfaction continues to be a strong predictor of job turnover.

Adequate and fair wages

Most workers expect much more from their jobs than high and fair pay. In fact, perhaps the most important long-run trend in worker values is the gradual expansion and broadening of worker expectations. Survey data have shown that the vast majority of workers not only in the United States but all over the industrialized world place a high value on autonomy, opportunity for advancement, and the ability to have a say in how they do their work. Moreover, the higher the level of education, the higher the value workers tend to place on these aspects of their jobs. Given that educational attainment levels are gradually rising, these dimensions of employment are becoming more central to behaviour at the workplace. It is not surprising, therefore, that leading employers throughout the world have been experimenting with innovations in the organization and management of the workplace that provide workers with these satisfactions.

Voicing workers' interests. With broader expectations and higher levels of education also comes a more assertive work force—one that is able and willing to voice as individuals or groups their demands or expectations. The means chosen to exercise voice at the workplace vary with the legal status and availability of collective forms of representation, degree of employer resistance to unions, and workers' preferences for individual or collective action.

There is a widely shared belief that workers' rights to challenge or seek changes in employer policies either as individuals or collectively should be protected by government policy and respected by employers. Even in the United States, for example, where the tradition of individualism is stronger than that of most other national cultures, surveys consistently show that more than 80 percent of the population believe workers should have the right to join unions or employee associations designed to give them a voice in shaping their employment conditions. While this right is provided by law in all industrialized democracies around the world, this is not always the case in developing nations or in totalitarian states.

Individual and collective action. Despite this widely shared belief in the right of workers to form unions or engage in collective bargaining, there are wide variations in the means workers prefer to use to assert their interests at the workplace. Generally, the higher the education and occupational status, the more workers prefer to assert their interests individually. Or, when organized, higher-level professionals such as teachers, doctors, lawyers, engineers, scientists, and middle managers tend to act collectively through occupational associations rather than in broad-based unions with blue-collar workers.

Professional workers

This occupational or professional approach helps to create and reinforce the professional ties and status of these groups as well as to bring their special needs to the attention of employers. Moreover, these groups tend to rely on the power they derive from their labour market and geographic mobility along with professional norms, licensing or certification procedures, and government-passed standards as much as, if not more than, they rely on collective bargaining. Still, in some countries of Europe an increasing number of white-collar and professional employees are organized into unions and negotiate collectively with their employers.

Blue-collar workers who have highly marketable skills also derive individual bargaining power from their potential mobility. Blue-collar workers around the world, however, are also more likely to form unions and bargain collectively to promote and protect their interests. In turn, while most employers would prefer to not have

their employees join or form unions, in the vast majority of industrialized democracies employers have respected workers' rights to organize.

Perhaps the strongest employer opposition has been found in the United States. American labour history is replete with pitched and often violent battles between workers seeking to organize and employers seeking either to avoid or eliminate unions. Even after the violence associated with these historic battles abated, both survey data and employers' revealed preferences continued to show that American employers placed a high priority on discouraging their employees from unionizing. The current and future status of unions and other forms of worker representation continue to be one of the most controversial issues in American industrial relations.

Participative decision making. How strongly do workers wish to take part in decisions that affect them? Do they want to be coequals with management on issues, or are their interests more limited? Such questions have been at the centre of many historic debates among industrial relations scholars, practicing managers, union leaders, and public policymakers. The evidence is surprisingly robust over time and across national boundaries: workers reveal the greatest interest in participating in decisions that affect their immediate economic concerns and those that directly affect their specific job.

Survey data collected from workers across 12 European and North American countries show that the majority of employees want a say in workplace decisions such as how they are to perform their jobs, how jobs are organized, and how problems related to their immediate environment are solved. An equally strong majority want a say on bread-and-butter economic issues such as wages, benefits, and safety and health conditions. Only a minority favour direct participation or indirect representation in the broad strategic business decisions normally made by high-level executives or the boards of directors of a firm. The one strategic issue that workers demonstrate real interest in influencing, however, is the role of new technologies at the workplace. When they can see a link between strategic managerial decisions and their own long-term economic and career interests, workers want to have a voice in those decisions.

Economic and career interests

Given these broadly shared values and expectations, some of the particular work and career concerns of various occupational and demographic groups are examined below.

WORK CAREERS

Managers. In all industrialized countries managers are typically recruited from among university or postsecondary technical-school graduates. Although there are exceptions to this pattern, it is becoming rare for blue-collar workers without a college or technical-school degree to rise beyond the level of first-line supervision into the ranks of higher management. Yet because few graduates fresh out of a university or technical school have the experience or background necessary to assume broad-based or high-level managerial responsibilities, most organizations invest heavily in systematic management training and development efforts.

Training and promotion. An initial part of the training often involves socialization into the practices, values, and culture of the organization. Another source of training and development lies in the career paths and job rotation policies of the firm. One large multinational firm, for example, devised a 10-year management development plan for all its junior managers, assuming that within those 10 years the manager would change jobs at least five times. Each job change was expected to expose the junior manager to a different functional area, such as marketing, finance, technology or product development, and manufacturing. Each job change also was expected to increase the number of people the manager supervised or the level of responsibility. This firm, like an increasing number of others, attempted to include international experience in the career path, especially for those young managers targeted early in their careers as having the potential to rise to the level of senior management.

Many firms make an explicit judgment of this potential early in a manager's career and put those thought to have the most potential onto a "fast track" developmental path. In smaller professional organizations, such as law offices or consulting firms, a similar decision is made within the first five to seven years on whether to promote an individual to the status of partner in the firm. This "up or out" decision is analogous to promotion to tenure of faculty members in most universities at some point prior to the seventh year of service.

Researchers have shown that managerial career patterns can be predicted quite accurately by the results of these early promotional outcomes. Some have used the analogy of a tournament to describe the process, in which "losing" at any step along the way significantly reduces one's chance of "winning"—that is, getting to the top of an organization or profession. Thus, a failure to get a promotion one expects (or that others expect a manager to get) often is a signal for the manager to look for opportunities in another organization.

The tournament analogy

Competitive pressures. The competitive career environment described above can lead to considerable tension and stress among middle managers. This stress is intensified by the desire of many firms to reduce the number of levels in the management hierarchy and the number of middle managers.

Rapid changes in business practices and knowledge also create a strong demand for continuing education programs for middle and senior managers. Most leading business schools and numerous consulting firms offer various short refresher courses or short conferences to practicing managers. Firms spend a significant amount of their training and management development resources on such programs.

In large companies that have plants or offices in many different locations, moving up the managerial hierarchy usually requires a number of geographic moves. While employers normally give a manager the option to accept or reject a geographic transfer and promotion, individuals who want to rise in their organizations tend to be reluctant to reject such offers. Yet the process of selling a house and moving one's family to another community can be difficult, especially if both the husband and wife have careers or their children are in early adolescence. This tension between work and the responsibilities and priorities of family life is a growing concern in many leading companies, especially as the number of women managers and dual-career couples increases.

Blue-collar workers. *Career structure.* The career of blue-collar workers can be divided into four parts: initial education and entry-level training period, trial or job-matching period, stable period, and retirement.

The long-run income prospects of a blue-collar worker are heavily dependent on the amount and quality of basic education. Failure to complete high school reduces significantly one's expected lifetime earnings. Obtaining post-high-school technical training through vocational schools, community colleges, or apprenticeship programs that involve both formal schooling and on-the-job experience increases expected long-run earnings. Finally, choosing a job that provides additional training opportunities either on the job or through part-time outside course work further increases a worker's earnings potential.

Thus, the initial career stage is one in which an individual is investing in education or, as social scientists put it, building human capital. Failure to complete high school or to acquire basic mathematical, verbal, and analytical skills not only limits long-run earnings but also increases the risk of being unemployed for longer periods than those who invested more time and energy in this period of education and training.

Building human capital

Following the completion of schooling and entry-level training, most workers experience a trial period in which they change jobs a number of times in search of a good match between their abilities and aspirations and the opportunities available to them. The average worker changes jobs six to eight times before settling into a stable employment relationship. Because most firms follow a seniority rule in laying off workers (that is, the most junior workers are laid off first), some of this job movement may be

involuntary. In searching for a job, blue-collar workers tend to rely heavily on informal contacts and information provided by friends, family members, or school advisers.

The most stable period of employment for this group of workers occurs between the ages of 30 and 60. As family responsibilities become greater and seniority increases the security of employment within an organization, the likelihood of staying with a given firm likewise increases. The potential costs of job loss also tend to mount over time, as it becomes harder to find a job with another company that will replace the wages and benefits often achieved after years of service and internal promotion.

Two different factors affect workers as they approach the retirement stage of their careers. The first is that, while age 65 still serves as a customary age for many to retire, firms are no longer allowed to impose a mandatory retirement age. Thus, theoretically, employees can continue working as long as they are able to do the job required of them. Few blue-collar workers, however, choose to stay on beyond age 65. Instead, an increasing number of them retire and then take part-time jobs. The second factor is the growth in early retirement among blue-collar workers, prompted in part by the early retirement incentives many firms offer to employees between the ages of 55 and 65. This also has led to a growing number of older workers employed part-time.

Job, union, and community. Along with the stages in workers' careers go shifting attitudes toward their jobs. When workers remain with the same company, their outlook on the job and the company tends to follow a curvilinear pattern: high at first, then dropping through the middle period, and rising in the later parts of their career. Individuals tend to begin work with such unrealistically high expectations as to the nature of the jobs and the opportunities before them that disillusionment later sets in; but after some years they adjust themselves, lower their expectations, and express more satisfaction with the work situation.

Interests at different career stages

Interest in joining a union or in becoming a leader in the union tends to follow a reverse curvilinear path. Interest is low at the beginning of tenure with a company because of the uncertainty over how long the worker will stay with the firm and because job satisfaction is generally high. Then, as job satisfaction declines over time and the worker recognizes that the costs of leaving the firm are rising as he or she accumulates seniority benefits and family obligations, interest in changing conditions in a given job or firm rises and stays fairly high. Only as retirement approaches and the costs of leaving the firm become untenable does job satisfaction again rise, thereby lowering the worker's tendency to participate in aggressive efforts to change the status quo.

A number of studies have shown that few blue-collar workers want to leave their community when the plant or office shuts down. It is clear that the availability of jobs at high levels of pay elsewhere is not in itself enough to move people. Ties with friends and family make workers reluctant to leave, but there can be good economic reasons as well. They may, for instance, find that housing costs are much higher in communities where job opportunities are expanding. Blue-collar workers and their families are therefore likely to conclude that it is best to stay where they are in the hope that the job market will pick up.

Service-sector workers. Most research on the careers and expectations of workers comes from blue- and white-collar workers employed in manufacturing industries. Yet this sector is shrinking in comparison with the service sector. In most advanced industrialized economies, more than half of private-sector workers are employed in services, compared with approximately 20 percent in manufacturing.

It is difficult to generalize on the nature of service-sector employment and careers, as they vary widely. For example, while average wages in service-sector jobs are lower than average wages in manufacturing, the wage differential between the best and the worst jobs in services is also larger than the comparable differential in manufacturing. This greater inequality of income (and skill requirements) helps explain why workers who are displaced from jobs in manufacturing experience, on average, significant pay cuts in the jobs they are able to find in the service sector. The best predictor of the size of the difference in pay between the job lost and the new job is the amount of education and transferable training the worker possesses. Again, education and training are critical to income and employment security and career advancement.

Service-sector jobs differ in several other important dimensions. First, more women are employed in services than in manufacturing. Second, services employ a relatively large and growing number of part-time workers, some who work part-time by choice and others who move to a full-time job when and if one is available. Third, low-paid service jobs tend to have high rates of turnover and lack many of the fringe benefits, training opportunities, seniority rules, and union protections found in the more stable and better-paying manufacturing, professional service, or public-sector jobs. Finally, service firms tend to be smaller in size and more vulnerable to changes in market or technological developments outside the control of their owners. For all of these reasons, the shift in the labour base from manufacturing to services has engendered vigorous debate over the ability of these new jobs to meet the high and ever-expanding expectations of the work force.

Technical professionals. The first research in industrial relations focused on blue-collar workers. Gradually attention spread to foremen and then to higher levels of management. Considerable attention has also been devoted to the study of scientists and engineers who work in industrial organizations. Interest in these technical professionals in part reflects the importance organizations attach to the development and use of new scientific discoveries and technologies. How well these technical professionals and the research and development processes they engage in are managed can have substantial effects on the long-run profitability of a firm and on the competitiveness of the larger economy.

Role in research and development

Scientists or engineers are often thought of as solitary individuals who work in a laboratory on some abstract problem or idea. While this may be an accurate image of the relatively small number of scientists who work on basic research, the vast majority of technical professionals in organizations actually work together in teams or project groups on applied research and development tasks. Their primary role is to transfer new scientific discoveries or ideas from the laboratory to manufacturing and out to the marketplace by creating new products or technologies. These project teams include representatives of diverse functional groups within the organization, including specialists in marketing, manufacturing, and human resource management as well as representatives of various scientific disciplines or technical specialties.

Interests and concerns. What do technical professionals want from their jobs and careers? Like all other workers, scientists and engineers are concerned about their employment security and long-term career opportunities. Concern over employment security arises in part from their dependence on winning contracts from customers or on obtaining budget funds from top management to support their project ideas. Because of the uncertainty over project funding, some firms try to keep their permanent research and development staffs rather small and hire engineers and other technicians as consultants on a contract-by-contract basis.

Like other professionals, scientists and engineers also want to gain the respect of peers in their field of work. Recognition of this desire led many early researchers to argue that these professionals were externally oriented—that is, they wanted to contribute to science and placed a relatively low priority on the needs of their particular employer. Later research showed, however, that most technical professionals also want to work on problems that are critical to the success of the firm. They want to know what the firm's priorities are and be given an opportunity to contribute to those goals. Above all else they want to be assigned to important and challenging projects and then be given the resources, influence, and autonomy needed to complete the projects successfully.

Career paths. Not all technical professionals want to

remain in technical jobs throughout their careers. Some aspire to move into management; others want to continue to do technical work but want the status and economic rewards that normally come with promotions to higher management. This has led many organizations to establish a dual-ladder progression system. Individuals in mid-career can seek promotions to more senior assignments on the technical ladder or to administrative positions on the management ladder. In theory the steps on each ladder are supposed to provide equivalent economic rewards, influence, and status. In practice, however, the management ladder usually provides broader exposure within the organization and thus better access to senior executive positions. Experience has shown that dual-ladder systems are extremely hard to administer.

Organized research and development. In the past, research and development work was organized in a serial fashion with a project passing from one group of specialists to another until it was ready to be given to the manufacturing section. The metaphor of "throwing it over the wall" was often used to describe this mode of organization, signifying both the serial and the isolated nature of each stage of the process. Research evidence convinced most organizations that this was very inefficient and time-consuming, and firms now encourage more cross-functional communication and participation by bringing together teams of representatives from each stage of the development process. The goal is to coordinate the process better and to identify and avoid problems that otherwise might be discovered only at a later stage.

To work effectively in these cross-functional project teams, scientists and engineers must have both up-to-date knowledge of their technical disciplines and skill in the communications, problem-solving, and group decision-making processes needed for successful teamwork. Universities are now placing more emphasis on the development of these skills in their curricula, and organizations are seeking ways to reinforce these attributes in the career development and reward systems.

The project leader

One key to the success of the research and development process is the project leader, who must motivate, lead, and coordinate team members. At the same time, the leader must represent the group's interests in the larger organization by serving as a champion for the team's project and negotiating with management for the resources needed to get the job done. In the end, it is the project leader who is held accountable for keeping the project on schedule and within budget.

Organizational design

A central task of management is to motivate individual workers and coordinate their collective efforts in ways that achieve the organization's goals. Over the years the concepts and methods used to structure work and to design organizations have changed considerably. In fact, it has been noted that an organization's age can often be predicted by the way work is structured, as work practices tend to reflect the organizational design theory in vogue at the time the organization was founded.

TWO MODELS OF WORK STRUCTURE

Specialization of function and separation of authority. Much of the organizational design in the United States can be traced to the influence of Frederick Taylor's scientific management movement and the division-of-labour concepts found in Max Weber's description of the ideal bureaucracy. One of the reasons for the durability of these concepts is that they helped make the modern corporation with its multiple divisions and functions both manageable and successful. Formal bureaucratic rules, specialization of functions, and close supervision proved suitable for disciplining and directing an immigrant and poorly educated labour force in factories geared to mass production markets. The phenomenal success of American manufacturing organizations in the first half of the 20th century reinforced managerial faith in these systems and provided workers with sufficient improvements in income and standard of living to support their continuity. Furthermore, the American labour movement adapted well to these principles, and the collective bargaining system that grew out of the New Deal of the 1930s provided workers with the opportunity to have their voices heard, if only indirectly, through union representatives.

Scientific management was based on the principle of a clear separation of authority between the engineers and supervisors, who decided how to organize work, and the production employees, who carried out the boss's orders. Also, it emphasized narrow job definitions and clear divisions of labour between jobs, thereby accommodating the presumed lack of education or skills of production workers. Finally, scientific management emphasized individual incentive wages. Paying each worker for the output he or she produced was expected to maximize employee motivation and overcome any presumed conflict of interest between the worker and the firm.

When the industrial unions that grew rapidly after the 1930s inherited this form of work organization, they generally accepted it, but they codified job descriptions, negotiated wage rates for each job, and established principles of seniority to govern worker rights to different jobs and workplace benefits. All these provisions were written into a collective bargaining contract, and disputes over interpretation of the contract were resolved through grievance arbitration.

The tightly structured workplace

This workplace system was combined with an equally tightly structured and differentiated managerial and technical hierarchy and departmental or functional division of labour. Specialization of function and clear lines of authority separated managers assigned to such various departments as marketing, sales, finance, personnel, production, and engineering. Within the engineering and new-product development process similar specialized tasks separated design engineers, manufacturing engineers, industrial engineers, and so on. A large cadre of middle managers produced the accounting and other financial reports needed by top executives to monitor and control operations and to allocate resources for new initiatives.

In a growing domestic economy and expanding mass markets, these organizational design principles allowed American firms to use their economies of scale to improve productivity and reap large profits. Sharing the fruits of these economic returns with the labour force in turn produced a stable industrial relations system that provided the labour peace and predictability needed to take advantage of expanding market opportunities. By the late 1940s and early 1950s American managers, union leaders, and policymakers alike were generally convinced that the American industrial relations system was superior to any other. Therefore, after World War II, efforts were made to export American institutions and practices to the nation's war-torn allies, as well as to its former adversaries Germany and Japan and its Cold War adversaries in communist-bloc nations.

Participatory management and flexible work systems. By the 1960s many of these traditional principles of organization and work group design were being challenged by early advocates of participatory management. Arguments for enlarging the scope of responsibilities and influence of individual workers were presented as better means of motivating workers and increasing job satisfaction. While these ideas gained favour in a number of the new companies and high-technology industries that grew rapidly through the 1960s and '70s, it was not until the following decade that they began to gain support within existing organizations in the manufacturing and service sectors.

Challenges to American organizational design

In the 1980s, when American firms faced significant productivity and quality performance problems, large numbers of American managers challenged traditional principles of organizational design. Also important was the recognition that Japanese and some European firms were outperforming American firms by using more flexible work systems and participatory management practices.

The development and implementation of new microelectronic technologies in that decade reinforced the pressure to introduce flexibility into work organization and to break down many of the traditional dividing lines between jobs. A number of scholars argued that the full benefits

of these new technologies could be realized only when, as one of them put it, workers are given the training, motivation, and organizational influence needed to "give wisdom to the machines." By this is meant that worker input into decision making must be encouraged and valued and that workers, technicians, and managers must work together to solve problems and resolve conflicts that arise between workers and managers and across functional groups. At the same time, experience showed that maintaining high levels of trust, commitment to problem solving, and flexibility requires a corresponding commitment from the firm to give a high priority to employment security and related employee interests.

Managerial critics have argued that new technologies also provide the option to "de-skill" many clerical and blue-collar jobs and give management new methods for controlling workers and invading their privacy. For example, computers can be programmed to monitor the work of machine operators and therefore to serve as a new form of electronic surveillance that replaces the personal presence and control of the supervisor or production foreman. The introduction of new technologies also displaces—and in some cases replaces—labour, posing a threat to the job security and economic livelihood of the workers affected. Thus, a critical challenge facing managers, worker representatives, and public policymakers lies in the management of technological and organizational change in ways that benefit not only individual firms but also the work force and the larger society.

CHANGE IN ORGANIZATIONS

The important issue of organizational change has always been a central topic of interest to theorists of industrial relations and organizational behaviour. Early studies focused on a presumed resistance to change among employees. It soon became apparent, however, that much of the resistance could be overcome or avoided by involving those affected in the design and implementation of the change. But it was also discovered that the management of change is an ongoing and complex political process that requires considerable leadership and resources.

The most widely acclaimed theoretical model of the change process was developed in the 1940s by social psychologist Kurt Lewin. He argued that successful organizational changes move through three stages, or cycles. The first stage involves the unfreezing of existing practices or behaviour patterns. This is most easily achieved if the organization is subject to some serious external threat or economic crisis. In the absence of a perceived crisis, employees are likely to see insufficient reason to change, and managers are often unwilling to commit the needed resources or to accept the risks that accompany change.

Lewin's theory of the change process

If practices or behaviours are successfully unfrozen, there is a period of experimentation, or trial and error, with new practices. Here the challenges are to implement changes in ways that respect and accommodate the interests and concerns of those affected and to demonstrate that the changes produce tangible improvements in organizational performance or employee well-being.

The final stage comes as a new set of practices is institutionalized and becomes the standard or accepted way of doing things.

Research has shown that each stage of change can take on the characteristics of an intense political process, in which those advocating the change must attract sufficient support from top management, union leaders, and rank-and-file workers to keep the process moving forward and to avoid the tendency of organizations to fall back into past patterns or practices. The job of a contemporary manager or union leader, therefore, has been described as a change agent. As the pace of technological and social change intensifies, the ability to manage organizational change and innovation successfully grows in importance.

Union–management relations

Of the variety of conflicts found in industrial organizations, those involving unions and management have received the most attention. Labour unions are the primary means workers have for advancing their collective interests at the workplace and in the broader society. Much of the history of industrial relations is filled with efforts on the part of workers to gain the right to organize into free trade unions—that is, worker organizations that are not controlled either by employers or by the government.

While the actual percentage of workers who are organized into unions varies considerably from country to country and over time within individual countries, it is safe to say that there is no democratic country in the world where independent trade unions are not present. Unions serve an essential role in a democratic society by giving independent voice to worker interests. The best evidence of the importance of this function is that unions are often among the first democratic institutions (along with the church and the press) attacked by totalitarian regimes.

Unions and union–management relations are also of special importance in that, through collective bargaining and other formal and informal means of interaction, unions and employers establish the wages, hours, and working conditions of large numbers of workers. In countries such as Sweden, Denmark, and Norway collective bargaining covers more than 80 percent of the labour force. In Britain, Germany, and Japan it covers between one-third and two-thirds. Even in countries like France or the United States, where less than 20 percent of the workers are unionized, collective bargaining often sets new patterns in wages and other conditions of employment that are then adopted by nonunion employers as well.

It should be noted that employers are often reluctant participants in collective bargaining. While the degree of opposition to unions varies among countries, most employers would prefer not to deal with unions. This opposition is perhaps strongest in the United States, where employers have aggressively opposed unionization of their employees. This is one of the reasons why the right of employees to organize and bargain collectively is normally protected by law.

Employer opposition to unions

The decade of the 1980s was a time of tremendous pressure for change in union–management relations. This pressure came from increases in market competition within and between countries, the rapid rate of technological progress, the changing nature of the work force, shifts in jobs from highly unionized large manufacturing firms and industries to smaller, newer firms and service industries, and, in some countries, the election of more conservative governments less supportive of unions. As a result unions in the majority of industrialized countries have lost membership and continue to debate how best to adjust their strategies and practices to their changing environments. The following discussion, therefore, focuses both on the traditional union–management practices that have dominated relations since the 1930s and on how these practices have responded to pressures for change.

UNION ORGANIZING

The typical way in which workers become organized into a union in the United States is through an election campaign and vote on representation. A majority of workers must vote in favour of union representation and collective bargaining. In these campaigns arguments about the need for a union and the benefits of collective bargaining are countered by employer efforts to convince workers that they do not need a union. American workers historically have taken a pragmatic approach to this choice: they vote in favour of union representation only when they are highly dissatisfied with their employer and when they see union representation as a viable means of improving their employment conditions.

Attempts to unionize white-collar workers

In the case of clerical and professional employees, unions have appealed by arguing that one need not see the employer as hostile or untrustworthy to believe in the need for collective representation. When an organizing drive took place among clerical and technical employees at Harvard University, for instance, the union campaigned on the slogan, "It's not anti-Harvard to be pro-union." While this approach has gained favour among white-collar and professional workers, it still is the exception rather than the rule for these workers to join a union.

Sometimes employers voluntarily recognize the union or remain neutral in the election process. This is most often the case in the public sector. Some private-sector employers have voluntarily recognized unions in new establishments in return for union cooperation and participation in the task of designing the work system, training the work force, and starting up operations. Such accommodation, however, is the exception in the United States; the typical picture of an organizing drive is still one of aggressive union campaigning in the face of aggressive employer opposition.

The more adversarial the organizing campaign, the more likely it is that the bargaining relationship will develop along similar adversarial lines. Conversely, the less resistance to organizing by the employer, the higher the likelihood that the union–management relationship will evolve along cooperative lines. For example, one large manufacturing company that voluntarily recognized a union in the 1940s, and has remained neutral in organizing drives held in new plants opened since then, has experienced only one brief strike in its entire history. This record stands in marked contrast to the pitched organizing battles and frequent strikes experienced over the years in the rubber, meat-packing, and coal-mining industries.

COLLECTIVE BARGAINING

What effects do unions and collective bargaining have on the outputs of the employment relationship that are of greatest interest to workers, employers, and the larger society? The historical evidence is that unions improve the wages, hours, and working conditions of their members. Perhaps the biggest and most direct effects have been on wages and fringe benefits; estimates indicate that unions have raised the wages and benefits of their members by 15 to 30 percent above those of comparable nonunion workers. Unions have also pioneered over the years in introducing an expanded array of fringe benefits such as paid vacations, sick leave, pensions, seniority provisions, apprenticeship and training programs, and grievance procedures for resolving conflicts on a day-to-day basis.

Assessing the effects of collective bargaining on the goals of the firm is a more difficult task. Historically, unions have served to encourage greater formalization and professionalization of personnel management practices. By increasing wages and related labour costs, unions have also encouraged employers to take actions that improve labour productivity. But the evidence is that, overall, unions reduce returns to shareholders, in part because they successfully capture a higher portion of a firm's profits than are normally passed on to employees by nonunion firms. These effects vary considerably, however, depending on the quality of the relationship that evolves between labour and management.

The American behavioral scientists Richard Walton and Robert McKersie captured the different patterns of relations that may evolve by distinguishing between "distributive" and "integrative" bargaining. Distributive bargaining is essentially a win–lose engagement. What one party "wins" through hard bargaining comes at the expense of the interests or goals of the "losing" party. In contrast, with an integrative bargaining approach the parties engage in cooperative problem solving in an effort to achieve a resolution from which each party benefits.

Distributive and integrative bargaining

In reality, most bargaining relations are mixed-motive in nature; that is, they have both distributive and integrative features. In the 1980s, however, the pressures on labour and management to solve complex problems intensified and therefore strengthened the efforts of many unions and companies to develop integrative relationships. The scope of labour–management relations expanded to include more opportunities for employee participation and union consultation in managerial decision making. Again, these innovative relationships did not spread to large numbers of bargaining relationships. Instead, sustained innovation and cooperation tended to be limited to environments in which the economic pressures for change were intense and the company was willing to share influence and power with the union and accept union leaders as joint partners in the enterprise.

The workplace in different cultures

Do the principles of organizational behaviour and industrial relations apply universally across nations and cultures? This issue not only has fascinated scholars and policymakers but, at critical points in history, has influenced the course of international events. After World War II, for example, the head of the U.S. forces occupying Japan imposed American-style labour laws and industrial relations practices under the belief that they would help ensure that Japan would not fall back into a militaristic or totalitarian state. By the 1980s the situation had reversed. Many American experts called for adoption of Japanese management practices in hopes of achieving the same high productivity, quality, and cooperative labour–management relations found in leading Japanese firms.

Transplanting national organizational practices

In both of these instances some practices were effectively transplanted to the other country. Free trade unions and collective bargaining did evolve in postwar Japan, albeit not in the same fashion as they had in the United States. The success of Japanese management did spur many American firms to reexamine their own policies and practices and to emphasize many of the principles underlying the Japanese system. This has been especially true in American manufacturing industries. In the auto industry, for example, the success of Japanese "transplants" (Japanese-managed plants operating in the United States and staffed with American workers) has convinced American auto executives that such practices as the use of teams to organize work, labour–management cooperation, worker participation, training in quality control, and just-in-time inventory systems result in higher quality and productivity than do traditional American practices.

However, the specific practices found in one country are hard to imitate exactly in another. To understand why and to explain why practices vary among nations, one needs to consider differences in national cultures, political and economic conditions, timing of the industrialization process, and key historical events that affect different countries. The comparative analysis that follows briefly reviews how these factors have influenced developments in the United States and compares their role in shaping the evolution of industrial relations in Japan and West Germany. These two countries are often compared with the United States because they have both achieved high rates of economic growth, productivity, labour peace, and improvements in workers' standards of living, yet these results have been achieved with very different institutions and practices.

THE UNITED STATES

Perhaps the value most closely identified with American culture is individualism. The importance of individualism can be seen in organizational systems of authority and conflict resolution, where subordinates are expected to question orders of superiors and attempt to resolve differences in a one-on-one fashion. The expected response to individual ambition and achievement is reward and promotion, and individuals normally turn to collective actions only when frustrated with organizational responses to individual efforts.

The broader economic and political context in which organizational and industrial relations developed has been one that places a high value on the role of the free market and keeps government intervention in private enterprise to a minimum. This ethos was particularly strong during the period of rapid industrialization between the late 1800s and the 1920s. The economic and social shock of the Great Depression modified this position considerably, however, and since then the American public has expected the government to play a more active role in regulating economic policy and industrial relations practices. Still, the view favouring decentralized institutions, industrial self-governance, and free enterprise has kept industrial relations focused at the level of the firm.

Free enterprise

Given these values, it is not surprising that the greatest conflicts in American industrial relations tend to arise over efforts to unionize a company and over negotiation of the specific terms of an employment contract. The value Americans place on individualism and mobility also helps

explain why turnover rates tend to be higher in American firms than in many other countries and why cooperative labour–management relations are difficult to sustain.

JAPAN

Shimada Haruo, a leading Japanese industrial relations scholar, has maintained that one can begin to comprehend Japanese industrial and organizational practices only by first recognizing that Japanese managers operate on the assumption that human resources are the one critical asset affecting the performance of Japanese enterprises. Therefore, management is deeply committed to developing and sustaining effective human resource and industrial relations practices. Many Japanese observers go on to argue that this assumption grows out of Japanese culture and traditions. Shimada points out, however, that this cultural thesis fails to explain the changes observed in management and labour practices over the years. Thus, he and most other contemporary Japanese experts stress the interactions of cultural, economic, and political events that shape organizational relations in the country's industries.

Japanese culture places a high value on family relations and obligations, and this family model carries over into the workplace. Employers are expected to show the same regard for their workers as a parent shows for other family members. Japanese expert Robert Cole goes a step further in noting that Japanese corporations can best be described as "semiclosed corporate groups"—that is, the firm is the key economic unit and should not be subject to significant outside interference. Unity within the firm becomes a central value and corporate objective. In turn, employees are expected to show strong loyalty to their employer.

Emphasis on the group

There are fewer traces of individualism and more emphasis on group relationships in the design of work and day-to-day workplace interactions in Japanese firms than in most Western nations. Direct conflict in organizational decision making is discouraged in favour of a more informal group consensus building. Authority is respected highly, so even in group problem solving the unstated views or preferences of senior managers tend to weigh heavily.

Prewar industrial relations. From the early days of industrialization, Japanese employers, labour leaders, and bureaucrats were divided over whether Western-style conflicts between management and labour were inevitable and whether Western models of unionization and dispute resolution were appropriate models for Japan. Many employers (and, in the nationalistic 1930s, some labour leaders) argued that Japan's "beautiful customs" of benevolence from superiors and loyalty from subordinates made the Japanese family a more appropriate model for industrial enterprise. Between 1920 and 1931 government policymakers brought forward eight proposals to provide a legal framework for the establishment of labour unions, but each was defeated by vigorous opposition from employers' associations and conservative politicians. At its peak in 1931, the prewar union movement had reached only 7.9 percent of the total industrial labour force. Large-scale enterprises were particularly successful in forestalling the formation of unions, and several developed alternative "Japanist" models of paternalistic management.

Postwar industrial relations. Japan's rapid economic growth from the mid-1950s through the 1980s propelled its industrial relations and organizational practices into the centre of international attention and debate. Three interrelated features of the system have attracted the most attention: (1) enterprise unions, (2) high levels of labour–management cooperation and cross-functional problem solving, and (3) lifetime employment security.

Enterprise unions. In the immediate postwar period the lifting of restrictions on unionization resulted in a wave of labour activism and unrest. Alarmed by the radicalism of the industrial union movement and the active involvement of the Communist Party at the movement's national level, the Japanese government and the American occupation authorities launched a counteroffensive (the "Red Purge" of 1947–48) to deny union rights to Communist-backed organizations. The newly formed Japan Federation of Employers' Associations (Nikkeiren) embarked on a campaign to form moderate, anti-Communist enterprise unions that included lower level management personnel as well as workers.

Employers made important concessions to the labour movement, including employment security, seniority-based wage systems, and twice-yearly bonuses negotiated each year along with base-pay increases. These accommodations provided the foundation for industrial relations in the large-scale enterprises that led Japan's remarkable economic growth from the mid-1950s to the mid-1970s. Even the slowdown in growth that followed the 1973 oil crisis did little to shake the implicit security of employment. Those industries that faced the most severe business declines—shipbuilding, steel, aluminum processing, and petrochemicals—undertook to cushion the inevitable reduction of employment through outplacement programs (often in cooperation with affiliated companies), government-subsidized retraining programs, and diversification.

Labour–management cooperation. The low level of conflict even in declining industries is an indicator of the generally cooperative relationship between management and labour that characterizes Japan's large private-sector firms (in contrast to the more conflictual relations in the public sector). Because blue- and white-collar workers belong to the same union, there are fewer lines of demarcation between these groups. In most enterprises, for example, the scale of management bonuses is tied to the size of bonuses for blue-collar workers. Many senior Japanese executives have been union leaders in their company at an earlier stage in their career.

In part because the union leader of today may well be the manager of tomorrow, large firms generally practice union–management consultation over broad strategic decisions. They also endeavour to elicit employee participation in day-to-day problem solving and quality improvements in the workplace. Quality circles and employee suggestion systems are widespread. Problems in product and technological development are more easily solved by employing cross-functional teams and by a career development strategy that provides engineers and managers with job experience in multiple functions, including working on the factory floor.

Problem solving

Employment security. Lifetime employment security is not guaranteed by law or contract but is embedded in the business and human resource planning policies of large firms. Growth in company size and stabilization of employment are high priorities for Japanese executives. In this sense, Japanese firms accept the stakeholder view of corporate objectives more readily than do their American counterparts. Recruitment, training, compensation, and internal promotion policies are designed in ways to facilitate lifetime employment.

Employees are hired into the firm immediately upon completion of their education and are expected to stay with the firm until they retire. In return the firm invests heavily in employee training and development. Layoffs are carried out only as a last resort, even during periods of technological change or a downturn in the business cycle. Wages in Japanese companies tend to rise with seniority, and most job openings within the blue-collar and managerial ranks are filled through internal promotions rather than by hiring from the external labour market. All these features limit the opportunities for workers or managers to transfer to other companies in the middle stages of their careers and increase the costs of leaving a firm that offers lifetime employment security.

It should be noted that these features of Japanese employment relations do not apply to all firms or all workers. While the law forbids discrimination against women and minorities, Japanese women traditionally have been excluded from the lifetime employment system and from higher-level jobs in corporations. Security of employment is also supported by a large number of small firms and subcontractors that employ many retired workers, immigrants, women, and those who have not found work or have lost their jobs in the large firms.

WEST GERMANY

The industrial relations system of the Federal Republic of Germany presents an interesting contrast to both the

American and Japanese models. The key characteristics of the German system are (1) industrial unions and industrywide collective bargaining, (2) formal structures for employee representation in management decision-making processes, and (3) the close integration of formal education and training with human resource practices within firms.

German cultural values

In contrast to Japan, few scholars emphasize the centrality of culture in explaining the character of industrial relations in West Germany. Instead, attention focuses on the legal framework and institutional structures created in the aftermath of World War II. Yet traces of German culture can be seen in these institutions, including a deep respect for the values of community and authority and a strong work ethic.

These qualities can be seen in the industrial relations system that began to emerge during the time of the Weimar Republic, between 1919 and 1933. The factory came to reflect the values of the society and to serve as an industrial community or plant family. In 1918 a compromise was reached between the ruling authorities and the German labour movement in which unions were recognized by the government and employers. In return unions accepted the basic rules of a capitalistic economy despite their socialist rhetoric to the contrary.

Nazi rule from 1933 to 1945 suppressed free trade unions. Yet when a democratic form of government was reestablished after the defeat of Nazism, the basic patterns of management and worker relationships initiated in the prewar era reemerged. A system was established that encouraged each party to assert its interests through a combination of political pressure for legal rights, formal negotiations, and cooperation within the enterprise. West Germany today exhibits a management style that respects workers' rights to representation and consultation and reveals a worker preference for negotiation rather than open conflict and challenge to existing authority.

Industrial unions. Following the war, the German labour movement was rebuilt with American aid and chose to develop along industrial lines similar to the emerging structure of unions in the American manufacturing sector. In the years that followed, union membership grew to approximately 40 percent of the labour force and has remained stable. An industry-level system of collective bargaining evolved. Firms within each industry have formed employer associations to serve as their bargaining agents with the industrial unions.

Government policy supported industry-level bargaining by enacting legislation that extends the basic wage and fringe benefit patterns negotiated in collective bargaining to cover workers in the nonunionized firms of each industry. These industry-level negotiations are supplemented with tripartite (government–union–employer) consultations at the national level over larger economic, social, and employment policy issues.

Consultation and codetermination. In addition to collective bargaining, both union and nonunion German workers are represented formally by works councils that are required by law to exist within establishments employing 20 or more workers. Works councils are representative bodies elected by all the employees in an enterprise. Management must consult with works councils on a broad range of human resource issues, including questions of adjustment to technological change, safety and health, training, and layoffs.

Representation on supervisory boards

German law also provides for a system of "codetermination," or worker representation, on the supervisory boards of large companies. These boards function in a fashion that is similar to the boards of directors in American firms. The supervisory board appoints executives to top management positions. One of these top managers is the *Arbeitsdirektor,* or personnel director; this person must be approved by the majority of the worker representatives who sit on the supervisory board. Consequently, by providing workers with a voice at the highest level of managerial decision making in large companies, German law and practice institutionalize a stakeholder view of the corporation. For Americans, this stands out as the most distinctive—and controversial—feature of the German industrial relations system.

Education and training. The starting point for human resource practices in German firms is the country's highly structured education and apprentice-training system. Tracking begins at age 10, when a small percentage of the most academically talented students (most of whom do not come from working-class families) enter a college preparatory program and later go on to obtain university degrees and jobs in their chosen profession. About 70 percent of German students are tracked into a vocational education and training system. At age 15 those in this track begin a three-year apprenticeship program in which they spend part of their time participating in on-the-job training in German companies and part of their time in classroom instruction. Upon completion of this apprenticeship they are certified in their trade. Further occupational mobility at later stages of a worker's career depends in large part on receiving additional training and professional certification. This system therefore provides considerable general training that is transferable to other enterprises and thus makes it possible for workers to move from one firm to another.

The emphasis on professional credentials and merit-based promotional policies in German firms is matched by relatively few status or other distinctions between blue- and white-collar workers and few rules or boundaries that separate occupations or job titles. The high degree of skill training combined with a strong work ethic reduces the need for close supervision. Studies have shown that German firms tend to have fewer supervisors than are typically found in comparable concerns elsewhere in Europe or in America. Finally, the heavy role that business enterprises play in the training and socialization of their workers helps explain why surveys have found German workers to be deeply committed to their jobs and to exhibit strong allegiance to their organizations.

COMMON CHALLENGES

Together, comparisons of the American, Japanese, and German models illustrate that, while each country fashions institutions that are consistent with its unique culture and with its economic and political environment, all industrial relations systems must deal with the same fundamental issues. They all must devise policies and institutions to meet workers' expectations, to provide them with a voice at the workplace and in political affairs, and to resolve conflicts that inevitably arise between workers and employers. How well an industrial relations system performs these functions has a major effect on the welfare of individual workers, their employers, and the society in which they live. (W.F.Wh./T.A.K.)

LABOUR ECONOMICS

Labour economics is the study of the labour force as an element in the process of production. The labour force comprises all those who work for gain, whether as employees, employers, or as self-employed, and it includes the unemployed who are seeking work. Labour economics involves the study of the factors affecting the efficiency of these workers, their deployment between different industries and occupations, and the determination of their pay. In developing models for the study of these factors, this section deals with the labour force of contemporary industrialized economies.

The influence of non-market forces

The economist cannot study the capabilities, jobs, and earnings of men and women without taking account of psychology, social structures, cultures, and the activities of government. Indeed, these forces often play a more conspicuous part in the field of labour than do the market forces with which economic theory is mainly concerned. The most important reason for this arises from the pecu-

liar nature of labour as a commodity. The act of hiring of labour, unlike that of hiring a machine, is necessary but not sufficient for the completion of work. Employees have to be motivated to work to an acceptable standard; the employment contract is, in effect, open-ended. This may be no problem when employees are weak and easily replaced, but the more skilled, organized, and indispensible they are, the more the care that must be given to creating an institutional setting that will win their compliance and meet their notions of fairness.

A second major reason for looking beyond straightforward labour market forces is the often highly imperfect nature of the industrialized labour market. The majority of jobs are occupied by the same employees for many years, and only a small minority of employees quits their jobs in order to move to a comparable job that is better paid. Studies in a number of countries have all revealed substantial variation in the level of pay offered for the same job by different firms in the same local labour market. This sluggishness of labour market response is particularly notable for more skilled labour and for labour employed by firms in strong product market positions. The main thrust of competition in many instances comes not through the labour market but through the product market, with an employee's pay being determined less by what the job is than by who the employer is.

In discussing both market and nonmarket forces in labour economics, the following discussion poses them not as alternatives but as complementary explanations. The difference in pay between, for example, a craftsman and the labourer who works alongside him may be fixed by custom, an arbitrator, a job evaluation system, or a bargain with a trade union. In their different ways these are far from being merely passive agents through which market forces are transmitted into human behaviour. They may, for instance, shape the market by defining its categories of labour. Also, they may differ greatly in their speed and extent of response.

The comparative study of wage movements in different periods and countries does show many similarities and regularities that are more marked than the variety of their settings would lead one to expect. This evidence of the influence of persistent forces working within an equilibrating system is one justification for the economist's speaking of a labour market. But there is much in labour that can be understood only with the aid of the psychologist, the sociologist, the historian, the labour lawyer, or the political scientist. Depending upon both the circumstances and the purpose for which the explanation is required, it is an empirical question how far the forces that these scientists study might interact with the market forces that are the special province of the economist.

Quantity and quality of the labour force

The size of a country's labour force, within a given total population, depends on two factors: the proportion of the total population that is of working age and the proportion of these who work for gain.

The limits of working age are usually taken to be established by the minimum school-leaving age and the prevailing pensionable age. Allowance must then be made for those persons who continue to work for gain after attaining pensionable age. Typically, some two-thirds of the population of an industrial country lies within these limits.

ACTIVITY RATES

The employed labour force may be characterized by particular activity rates. An activity rate is the proportion of the whole number in a given age and sex group—for example, females aged 30–34—who work for gain. Among males, activity rates in the earlier years of working age are as a rule low, because so many remain in education and training. Between the ages of 25 and 50, male activity rates approach 100 per cent, but from 50 onward they fall as men begin to retire. The pattern of female activity rates is very different and has changed greatly in the second half of the 20th century. Formerly, female rates ran higher than male in the earlier years because fewer girls enjoyed extended education, but from the age of 20 onward they fell sharply as women married and withdrew to domestic duties. Women so occupied remain by far the largest contingent of persons of working age but not in the labour force. Since World War II, however, it has been less usual for women to leave paid employment immediately on marriage. A fall in the age of marriage, moreover, together with a smaller size of family, has enabled many women to return to paid work in their 30s, and female activity rates have come to show a second peak between the mid-30s and the mid-40s, after which they decline more steeply than male rates. From these various activity rates there emerges an overall proportion of the gainfully occupied among all of working age that is typically in the region of two-thirds.

QUALITY OF LABOUR

The quality of the labour force depends on education and training, physique, and health. There is evidence that physique has been greatly improved by increases in the standard of living in the 20th century. Because of the reduction in family size, this rise has been even more marked for children than for adults, and the effects have been seen in the greater height and weight attained by children at a given age. The beneficial effects of stronger physique on health have been enhanced by the advance of medical knowledge and the increased availability of medical services. Better health has raised productivity by a reduction in absenteeism and by a prolongation of the working life during which the economy reaps the benefit of the education and training the worker has received.

Investing in education and training

Education and training can be regarded as a kind of investment, and the rate of return it yields can be estimated. The amount of the investment is the value of the student's use of resources—buildings, equipment, and instructors—together with the output that the economy would have enjoyed from work had the student been gainfully occupied rather than studying. The yield, in turn, is calculated by assuming that the average subsequent earnings of those who completed a given course of education, compared with the average earnings of those who stopped just short of it, provide a valuation of the increase in productivity that the course confers. From this difference in earnings there must be deducted the contributions to the sinking fund required to replace the amount of the investment by the end of the student's working life. The net yield so calculated can then be expressed as a rate of return on the investment. Estimates suggest that this rate of return is not less than that generally obtained from investment in physical capital. They also indicate that a great part of the productive resources of the economy consists in the education and training embodied in its labour force.

Though estimates of this kind are subject to some objections in principle, they do serve a useful purpose in stressing the potential of education and technical training in raising productivity and the risk of investing too little in them relative to other forms of investment. There is no less a risk of underinvestment in training in industry. The great obstacle there is that the employer is not assured of retaining the services of workers in whose training he has invested. Employers generally follow one of two strategies. They may provide training in-house and seek to retain the employee by inducements such as the prospect of career progression, pension entitlements, and other devices designed to encourage loyalty and an "organizational orientation." Or, alternatively, employers may combine to establish industrywide training arrangements, sometimes with statutory support, thereby permitting ample skilled employees for easy movement between firms and more of a "market orientation" in their work forces.

Deployment of the labour force

Shifting patterns of employment

The contribution of education and training to economic development is apparent in the changes that have taken place in the deployment of labour in the developing economies. When the deployment of the labour force is followed over a period of time, certain patterns appear. One of these arises from changes in methods of production.

In farming, improvements in technique and equipment have made possible an increasing output from a declining labour force. In industry, the extension of research and development, the increased complexity of products and equipment, and new methods of collecting, storing, and processing information, along with other developments of management procedures, have all acted to increase the numbers of administrative, clerical, and technical workers relative to manual workers. A second course of change has affected occupations linked with particular industries, when those industries have contracted or expanded as compared with others. Coal mining and cotton textiles are examples of contraction. The service industries, on the other hand, have expanded: a greater proportion of household expenditure is devoted to services; education has extended; governments have provided more social services. A third course of change has its origins in relation to supply. Domestic services, for instance, have contracted because improved education and the opening up of other occupations to women has enabled many to take up work that they prefer. One general tendency is that as standards of living rise the service industries absorb a greater proportion of the labour force, because the rising demand for their output is not generally offset, as in manufacturing, by a progressive reduction in the amount of labour required to produce a given output.

The far-reaching changes that have come about in the relative numbers in different occupations and industries have called for corresponding changes in the training and allocation of young entrants to employment and for the movement of workers already in employment to other kinds of work and, often, other places. Though part of this adaptation has been unplanned and undirected, a number of governments have undertaken to foster the process of adaptation by a labour-market policy. One means of applying this policy is the provision of information to job seekers as to vacancies immediately available, and to workers at large as to the prospects and requirements of particular occupations. Labour-market policy also tries to guide entrants toward those occupations for which an expansion of demand is expected. One way of doing this is by promoting the training and retraining of selected persons for selected occupations. The function of retraining may be extended, as in Sweden, to offer all workers opportunities to qualify themselves for better-paid jobs throughout their working lives.

Fixing rates of pay

Wages may be fixed by collective bargaining between unions and management or by individual bargaining between worker and employer or simply by custom. When the status of wage earner became distinguished from other forms of labour, it was marked by the existence of an individual agreement about the rate of pay between wage earner and employer. The law still recognizes the individual contract of service even where the rate of pay has been fixed collectively. In earlier days there was often not even individual bargaining, because customary rates of pay prevailed that might be unchanged for many years at a time. In southern England, for instance, the prevailing rate for building craftsmen remained at sixpence a day for 120 years after 1412; for most of the 500 years after 1412, the building craftsman's rate was half again as great as the labourer's, or nearly so.

Customary rates

After industrialization had set in, custom continued in some measure to regulate rates of pay and to protect workers who entered into individual agreements. But its sway was much less extensive: from time to time rates changed. Although there was at first no reference to the cost of living, when price increases were general and sustained, there must have been informal understandings among the wage earners of a locality that each in making his own agreement would hold out for a higher rate. At times of increased demand for labour, moreover, the employer would have to offer a rate sufficient to attract and retain the wage earners against the competition of other employers. The necessity of holding needed labour is today the governing factor for employers who have workers with whom they do not negotiate either collectively or individually—generally clerical and administrative workers.

Frequently where the safeguards both of custom and of competition for workers have been missing, workers have felt the need to combine in order to bargain collectively. The force of custom declined as industrialization created new jobs and moved workers into new localities. Business fluctuations brought unemployment so that instead of employers competing for labour, workers were often competing for jobs. Thus industrialization has been universally associated with the rise of trade unions. (See below *Organized labour: Trade unionism.*)

TRADE UNIONS AND BARGAINING AREAS

A main purpose of the trade union was to maintain a minimum rate of pay for its members, a purpose that led unions to extend or delimit both their membership and the number of employers with whom they bargained. The starting point was typically the club of craftsmen in a certain locality, concerned to ensure that none of its members worked for less than the rate it recognized from time to time as a minimum and to raise that rate when opportunity offered. By bringing all who worked in the same craft and district into membership, the club could reduce the risk of their bidding against each other; and if it could also limit the number entering the craft—by controlling the number of apprentices—it would be more likely to be able to raise the rates. However, since it was still likely to be subject to the competition of members of the same craft coming in from other places, and some of its own members might move in search of work, it had an interest in extending its coverage over all members of the craft throughout the labour market.

Representing all craftsmen within a market area

If the labour market was not coextensive with the product market, however, the union might still find itself exposed to the competition of workers at a distance if these worked at lower rates and so enabled their products to be sold at lower prices. Thus there was reason to extend the coverage of the union up to the boundary of the market for the product, though it was not practicable to organize workers in other countries. However, the union would see no advantage in bringing workers of other occupations into membership: on the contrary, it was felt that one could expect employers to concede a rise more readily if it would have to be paid to only a restricted membership. What has been said here of the craft union applies to all unions insofar as their aim is to maintain and raise the pay of members of a given occupation: the pursuit of that aim will lead them to embrace all the members of the occupation throughout the market for their product and to establish a basic rate throughout this bargaining area.

The reactions of employers both reinforce and modify this tendency. The ability of any one employer to pay a given rate depends largely on what rates are being paid by other employers who compete in the product market. When competition is close and labour costs are a substantial proportion of total costs, all employers selling in a given product market have a strong inducement to negotiate only through an employers' association that embraces them all. Most employers' associations are in fact industrywide, though some are limited to particular regions or sectors of an industry. Employers also know that what is conceded to employees in one occupation will commonly be demanded by those in others, unless they are divided by such a gulf as used to separate the manual from the clerical workers. Employers therefore commonly prefer to reach an agreement with all their workers in common and may make this a condition of negotiation. They thereby put pressure on occupational unions either to extend, amalgamate, and divide up until they form industrial unions each embracing all the manual workers in a given industry, as the Swedish unions have done, or to enter into confederations that provide all the unions having members in a given industry with a common front for the purpose of bargaining—the course followed by British unions.

Many semiskilled and unskilled workers are unable to seek bargaining advantage by restricting the membership of their unions to one defined occupation: they have to

Organizing all workers within an industry

seek it rather through the accumulation of funds and the force of numbers—for them, "unity is strength." Some unions have therefore adopted the principle of industrial unionism from the outset, in accordance with the tendency noted above toward establishing industrywide bargaining areas. Others, the general unions, have set out to recruit workers from every occupation and industry; but for bargaining purposes they have commonly had to act on behalf of their members in each industry separately. In any clash between the forces delimiting the bargaining area and those delimiting the trade union, the former generally prove the stronger.

EFFECTS OF COLLECTIVE BARGAINING

Collective bargaining developed with the growth of trade unionism, especially from 1890 onward. It impinged upon labour markets in which the trend of money wages was upward: in years of good business, money wages generally rose, and though in the years of falling or low activity they were often cut, the cuts were generally smaller than the preceding rises had been.

Leveling of pay rates. A first effect of the extension of collective bargaining was to reduce pay differences, which had been large, between the wages a given grade of labour received at any one time in different regions and in different firms in the same region, and even between one worker and another under the same employer. The unions at first had to accept the prevailing regional differences, but their pressure to bring up the lower-paid regions has reinforced the effect of improved communications and information in reducing these differences greatly, especially since World War II. Assurance of "the rate for the job" raised the wages of particular groups or individuals who lacked access to alternative employers, either spatially or because of their lack of information and mobility. In general, the extension of collective bargaining brought about greater uniformity in the rates of pay received by workers of a given grade, and it did so by raising the lower rates.

Collective bargaining has also affected the forms in which improvements in pay are realized. It has borne particularly on those parts of the terms and conditions of employment that of their nature require to be regulated collectively. Chief among these are the hours of work. The extension of such fringe benefits as insurance and pensions paid for by the employer has also reflected trade union pressure.

Raising the level. Studies of differences between the movements of wages in unionized and nonunionized sectors of employment, especially in the United States, have brought out three other effects of the extension of collective bargaining. One is an impact and once-for-all effect: the introduction of collective bargaining has raised the wages of the workers concerned, relative to the general level prevailing around them, by some 10–15 percent. A second effect has been in the timing of changes: when wage rises were the order of the day, unionized workers achieved them earlier than nonunionized; and when the market was moving the other way, cuts of unionized workers were put off longer. When the cost of living has risen rapidly, as in wartime, the unionists' ability to secure compensatory rises in money wages more promptly promoted the extension of unionism, especially among white-collar workers who had previously stood aloof from it. The third effect has been in the ability not only to defer wage cuts in depression but also to reduce their amount. In the United States, for example, the differential between wages in the unionized and nonunionized sectors was at its highest in the 1932 depression trough. A major effect on the general level of pay in terms of purchasing power and on its share in the product of industry seems to have stemmed from the resistance to pay cuts in the world economic depression of 1921: though pay was cut severely, often after protracted struggles, it could not be brought down as far as product prices had fallen, and in more than one country the distribution of the product of industry between pay and profit seems to have been permanently shifted.

Advantages over nonunion pay rates

Limitations. By raising the pay of particular workers and by modifying fluctuations in the workers' favour, over a period of time collective bargaining has made the total of pay higher than it would have been otherwise in the same conditions of the market. But the effect has been limited. Before World War II the movements of the general level of pay continued to depend mainly on market conditions, and the points at which the effects of collective bargaining can be distinguished clearly are fewer than might be expected. Collective bargaining provided the arena in which market forces took their effect, rather than a shelter from or alternative to them.

After World War II, however, the bearing of market forces on collective bargaining changed. One important influence was full employment (at least until the 1970s), but others were the increased importance of governments as employers, an apparent diminution of the significance of labour cost in product market competition, and, from the 1970s, the floating of national exchange rates. Employers gained the expectation that if they agreed to rises in pay that would exceed the rise in productivity, and so raise unit costs, they would still be able to preserve profit margins by raising the prices of their products—and do this without losing business, provided only that the initial rise in pay was not greater than that which was being conceded at the time by other employers. Some countries, such as Sweden and West Germany, had employer organizations that were sufficiently united to resist these pressures. Other countries with little employer solidarity and highly fragmented bargaining, such as Britain and Italy, suffered persistently high cost inflation. Thus, the impact of trade unions cannot be assessed in isolation from that of employers.

Division between pay and profit

A second limitation is that, even where collective bargaining has affected the movement of money wages, it has had only transient effect on the division of the national income between pay and profits. Whatever the course from time to time of rates of pay in money, the pay per person in real terms (*i.e.*, in terms of purchasing power) has risen with remarkable regularity in much the same proportion as output per person, save for the one major exception of the displacement in favour of pay in the early 1920s. It appears that firms take advantage of opportunities to restore profit margins either by maintaining their selling prices while productivity rises or by raising those prices. A rise in real pay initially conferred by any one rise in money pay, therefore, will be reduced as the cost of living rises.

THEORY OF BARGAINING

Limitations on the scope of bargaining are also suggested by theory. Collective bargaining can be seen as the reduction of two risks to which the worker is exposed through individual bargaining. There is first the risk that the worker will be merely one of a number of applicants for a single vacancy and that competition between them will force the pay down. Even as the sole applicant for the vacancy, there remains the second risk that the job will be offered only on terms that are unacceptable; in the event of failure to agree, going without the job will inflict more hardship on the worker than not filling the vacancy will on the firm. Bargaining through a trade union removes the first risk by ensuring that whichever applicant the firm engages it must pay not less than the union rate: in this sense the union exercises monopoly power. Membership in a trade union reduces the second risk by increasing the workers' relative power to change proffered terms by withholding consent: in this sense the union confers bargaining power.

Constraints of supply and demand. The scope of the monopoly power that the union exercises by maintaining the rate for the job may be seen by supposing that this rate is simply announced by the union, which leaves firms to hire as many or as few people as they choose at that rate. In deciding how high it can set the rate, the union must have regard for the consequences for employment. Firms may be able to alter the design of the product and the method of production so as to use less labour. To the extent that they cannot economize in the use of labour and that the pay of this labour enters into the total cost of production, a higher cost arises that firms may be obliged to pass on to their customers through higher product prices. The customers are then likely to buy less from them, especially if there is international competition

in the markets for the product, and again employment will suffer. Thus a union that dictates its own terms is still subject to the constraint of the demand curve for the labour concerned. Equally, if the employers dictate the rate of pay, they could not set it so low as to make it impossible to attract and retain the required labour force: they would be subject to the constraint of the supply curve of the labour concerned.

The costs of work stoppages. When neither side dictates the terms and an agreement must be negotiated, failure to agree results in a stoppage that causes losses to both parties. Attempts have been made to develop the pluses and minuses of these losses into a theory of bargaining. If, for example, it is assumed simply that the continuance of a stoppage progressively increases the wish of the parties to end it, and so causes firms to raise and the union to lower the rate at which each is prepared to settle, then the stoppage will end on the day when the two rates have been brought into equality. Further, if the parties agree in their forecasts of how the wish to settle will be affected by the continuance of the strike, they will find it in their interests to reach agreement on what would be the terms of the ultimate settlement without resorting to coercion by stoppage. A more elaborate theory has been developed in which each party is seen as weighing the cost to itself of a stoppage of given length, the benefit to it of a given concession by the other party, and its estimate of the effect of a given extension of the stoppage on the willingness of the other party to make a concession.

Influence of non-market forces

In practice much more is involved—internal political pressures, for instance, personal prestige, or the tactic of involving the government and public opinion. Many of the costs of a stoppage, moreover, are hard to express in terms of money. The above three variables must always figure prominently in the parties' consideration. A stoppage is unlikely when on a consideration of these variables it appears that there will be no net gain; this situation exists when bargaining power is evenly balanced or when negotiation has already brought the parties' positions close together. One party is likely to see a clear advantage in a stoppage only when market forces are working in its favour, and these will have told already in the course of negotiation. In particular the cost of a stoppage will be high to employers when they are busiest, whereas in a recession a stoppage may be a positive benefit to them. Insofar as bargaining power is thus conferred by market forces, it injects no distinct factor into the determination of rates of pay. Bargaining power may also be conferred by determination, loyalty, and leadership on either side. It has also been conferred on trade unions by the expectation, engendered among employers by the experience of sustained full employment, that rises in pay can be covered by higher prices so as to maintain profit margins without loss of business.

METHODS OF PAYMENT

The productivity of a work force depends to a substantial extent upon the successful management of its payment system. Employees generally judge the "fairness" of their pay not by its absolute level but by its level relative to that of other employees, and in particular those with whom they are in close proximity. Their criteria for fairness are generally very conservative; the fair pay differential is the one to which they have become accustomed. A disturbed differential can be a source of discontent and lack of motivation. The technique usually used for managing the internal pay structure, as the relative pay of occupations within an organization is called, is job evaluation.

Job evaluation. This term covers a range of procedures used to develop and maintain a consistent internal pay structure that is acceptable to the work force. Ranking methods use surveys of the work force's preconceptions of fairness to arrive at a comprehensive pay structure. Analytic methods score the requirements of different jobs according to distinct criteria such as physical effort, mental skills, responsibility, and working conditions and then use weighted averages of these scores to establish the final pay structure. Job evaluation is typically participative, methodical, and ponderous. It offers a means to legitimate a pay structure and a procedure whereby changes in that structure can be negotiated and implemented. As such, it is a defense against the effect of disturbed pay differentials on employee motivation.

Pay incentives. By contrast, there are a great variety of devices that use pay as a positive motivator. The most common method of payment is according to the duration of time worked—by hour, week, month, or year. But additional merit payments may be added on at the discretion of management as rewards for good performance. The hazard here is that, if employees feel the criteria on which these are based are inconsistent, the effect may be negative.

Salary structures are more formal devices that offer a range of pay levels for different job grades. The employee's position within the range may depend upon managerial discretion, or it may be formalized into automatic annual increments. Promotion between job grades depends upon criteria over which managerial discretion has stronger incentive effects.

Piecework and standard times

Payment by results most commonly relates money payment to physical output for a part of the wage. This may be done for an individual as piecework or for a group of workers. In order that the incentive effect be seen as fair for employees engaged on different tasks, it is necessary to develop common standards to provide the same rewards to comparable increases in effort. The work study techniques devised for this use a combination of accurate timing and the observer's judgment of the effort being applied over many repetitions of the job to arrive at a standard time, which is then directly comparable with the standard times for other jobs. This provides a basis for incentive payment, with the same bonus being earned by workers who complete their different tasks in the same percentage briefer than their standard time. In practice, there is ample opportunity for dispute and for the emergence of contentious anomalies, particularly as a result of minor changes in production technology. The incentive effect usually fades with time, and most payment-by-results systems have a limited life.

Payment related to corporate performance has become increasingly popular since the 1970s. Rather than linking employees' bonuses to their own performance, it is tied to profits or some other indicator of the state of the company. The main advantage of this is didactic; it is believed to increase loyalty and to educate the work force about the commercial circumstances within which the company operates. For similar reasons many governments have encouraged employee share ownership schemes.

Single-employer bargaining. In the United States, Japan, Great Britain, and a growing number of other countries, the scope of pay bargaining is often no greater than a single employer or even a single plant. This has the advantage that the wage structure and incentive system can be closely tailored to a broader package that includes training, motivation, and career development. It requires the employer to sever links with the multiemployer industrywide agreements that have often prevailed previously. It also implies that the trade union unit of organization is focused on the single firm as well—as a "local" in the United States, as an "enterprise union" in Japan, or as a "joint shop stewards' committee" in Britain. Such organizations enjoy considerable or complete autonomy from the wider union movement, making them in some respects weaker and more pliable.

Single-employer bargaining is a strategy that offers a firm greater freedom to manipulate the productivity of its work force by isolating its trade union (if any) and developing organization-oriented attitudes and company-specific training and job descriptions. It does not, however, provide the employer with any influence over the generally prevailing level of pay settlements. This is offered by the alternative multiemployer strategy, which also permits a more market-oriented approach to labour with industry-wide wage and training agreements. Multiemployer strategies do not imply complete uniformity of payment across all firms: in practice they tend to have discretion to vary the agreement somewhat at plant level. In some countries this is a fairly disciplined two-tier arrangement; in others,

local bargaining pressures cause the plant-level element to dominate in what becomes known as wage drift.

PUBLIC REGULATION OF RATES OF PAY

Minimum-wage laws. Governments have intervened in three ways to enforce minimum rates for workers who lacked both the protection of trade unions and competition between employers for their services and whose wages in consequence were regarded as needlessly low. One way has been to provide by law that "recognized terms and conditions of employment," such as those reached by collective bargaining for workers of a particular description, shall be applied to all others engaged in the same kind of work. A second way, followed by the United Kingdom since 1909 and by a number of state legislatures in the United States, has been to set up boards of representatives of the workers concerned and their employers, together with independent members, charged with determining rates of pay and hours of work that are legally binding as minimal on all employers within the scope of the board. The board discusses and negotiates wage claims in much the same way as in collective bargaining, albeit if the parties cannot reach agreement, the independent members have a deciding vote.

These two forms of intervention are calculated to raise the pay of particular groups of unorganized workers only to the extent that it would be raised by the extension of collective bargaining to cover them. A third way, followed notably by the United States in its Fair Labor Standards Act since 1938, has been to specify by statute the actual minimum wage applicable to wide categories of employment—the amount set being such that only a relatively small number of workers, namely the lowest paid, are immediately affected. When such measures were first proposed, critics argued that they would only result in the workers they were intended to protect losing their jobs. In some cases this has happened, as when the United States minimum wage was applied to the needleworkers of Puerto Rico. More often, however, the workers concerned were receiving lower pay than a competitive market would have afforded them—that is, if they had had more access to alternative employers. Minimum-wage measures tend to discourage labour-intensive methods of production, so that while they may cost jobs in the short term, they tend to force employers into more advanced production technologies, which create greater long-term growth and employment potential.

Costs and benefits of the minimum wage

Arbitration. Another way of regulating rates of pay is a by-product of arbitration systems set up originally as a means of avoiding strikes and lockouts. In Australia it has become the practice, accepted by both employers and trade unions, to have the main proportions of the wage structure and the movements of the general level of wages determined by the awards of arbitrators to whom these issues are submitted in the form of disputes. In setting rates for particular occupations or industries relatively to others, arbitrators must in practice have regard to what is acceptable to the parties; for even where arbitration is compulsory, its awards would cease to be observed if either party had cause to believe that the terms of the awards were persistently less favourable than it could obtain by its own bargaining power. In regulating the movement of the general level of pay, the arbitrators have more discretion; but the government, and the employers insofar as they meet international competition at home and abroad, will make them aware of the effects of the awards on the level of domestic costs and prices and on the balance of payments.

National incomes policy. Under full employment the rise in effective rates of pay has generally been inflationary in that it has exceeded the rise of productivity. The consequent rise in costs and prices has at times been disturbing domestically and has been particularly embarrassing to governments that face difficulties in balancing their external payments. Governments in general have been unwilling to check the rise of inflation by applying fiscal and monetary restraints to the degree that unemployment would be substantially raised. In the belief that at least part of the rise is due not to excess purchasing power but to the pushing up of costs and prices, governments have appealed to those who make decisions affecting labour costs and product prices to moderate the rise in pay and profits. Some governments have formulated norms that would, in theory, keep the general level of prices constant and would keep the general level of pay rising only at the rate of the expected rise in productivity—allowing, of course, for specific exceptions. Agencies have been set up to apply these principles, but usually only by way of investigation, assessment, and advice. Governments have preferred to rely on the acceptance of the policy in principle by employers and trade unions, and on their efforts to secure its observance by their affiliates. Even where statutory powers of control exist, they have usually been kept in reserve. During times of recession, such as those experienced in the 1980s, governments have generally suspended their efforts to enforce a national incomes policy.

The structure of pay

Systematic differences are found in the average earnings obtained in different regions, industries, and occupations. The average earnings prevailing in different regions of a country show a considerable range between the highest and the lowest, even when the same procedures for fixing rates apply everywhere. Much of the dispersion is due to differences in the localization of industry: if the relatively high-paying iron and steel industry, for instance, is concentrated in a particular region, then the average pay of the region will be raised to that extent. But regional differences also exist because work of the same kind frequently commands different rates of pay in different regions. Such differences are sometimes necessary to maintain the balance of payments between regions; they may also be in some measure a legacy of history and are likely to be reduced as communications improve and labour becomes more mobile. As noted above, this process of reduction has been expedited by trade union pressures.

Factors in pay variation

Average earnings also vary from industry to industry, and the considerable range that appears can again be attributed largely to differences in the composition of the labour force: an industry such as printing, which by the nature of its processes employs a high proportion of skilled workers, will on that account alone show higher average earnings than, say, the textile industry, which employs a higher proportion of the semiskilled. The similarity between the structure of earnings by industry in different countries—with printing, iron and steel, and engineering near the top, and textiles and food processing low down—is thus attributable to common processes requiring similar compositions of the labour force.

OCCUPATIONAL PAY THEORIES

This occupational structure, therefore, presents the main object of economic analysis. International comparisons show that the ranking order of the rates of pay prevailing in different occupations is similar in different countries, but that the range, whether between professional and manual occupations or, within the manual, between the skilled and unskilled, is much wider in economies at an early stage of development and diminishes in the course of development. These are the principal observations to be accounted for by any theory of the differences in the rates of pay that different kinds of work command. Several such theories have been propounded.

The narrowing gap between rates of pay

Status. One theory stresses the link between occupations and their status in the community, some having higher status than others. The community believes, according to this theory, that pay should correspond to status; and the rate of pay for each occupation is assigned to it by common consent, reflecting the place it occupies in the hierarchy of esteem. The implications are that the community's discretion is not as a rule subject to other factors such as the market forces of supply and demand and that, if people came to make less distinction of status between occupations, then the rates of pay for different jobs could be more nearly equal. Doubtless, many people do think in the way the theory supposes, feeling it anomalous, for example, when an occupation that is commonly

accorded a higher status than another ceases to command a higher rate of pay. But the correspondence between status and pay is ambiguous; it is not clear to what extent pay is made to fit the status and to what extent status follows from pay. Moreover, the occupational pay structures of different countries show more similarity than do their social values. Nor does the theory explain why the pay structure has generally become compressed in the course of development, and the ranking order sometimes inverted, as when some clerical occupations drop below some manual ones—changes that can otherwise be explained as the effect of extended education in increasing the relative supply of more qualified labour.

If the theory is not acceptable as an explanation of the pay structure as a whole, it does call attention to a factor that appears to affect parts of that structure. One of these parts is that of the higher administrative posts. It is generally accepted that any such post must carry a higher salary than any post below it in the chain of command; and when this chain is long, as it is in a big corporation, the salaries set for the posts in it, and the high level reached at the top, are to be accounted for by this principle.

Different rates of pay for men and women

The same theory also suggests a cause of prevailing differences between men's and women's rates of pay. Some women's work is different in kind from men's irrespective of the fact that it is done by women; and, where men and women both do work of the same description, some disabilities attaching to women as employees, in particular the likelihood that they will not stay in the job as long as men, may make them worth less to the employer. There are many jobs, however, in which these considerations do not apply and in which there is no difference in the productivity of men and women adequate to account for the actual difference between men's and women's rates. The difference seems attributable rather to customary attitudes and valuations—in particular, the assumption that women's productivity is lower in all jobs and the belief that pay should be proportioned to need (women workers generally needing less than the man who has a family to support). If such factors as these account for differences in pay between men and women where there is no corresponding difference in the work they do or the efficiency with which they do it, one may speculate that the same factors account for some part of the pay differential where the two kinds of job are distinct.

Unions and the rising pay of less-skilled workers

Power. A second theory lays its stress on power: the ways in which organized groups can protect and advance the pay of their members. Any group that restricts entry into its occupation can keep its labour relatively scarce and thereby support the rate of pay that that kind of labour commands. The discussion above of the economic effects of the trade union indicated the circumstances in which a trade union would be able to raise the relative pay of its members by the exercise of monopoly and bargaining power. The general increase in the pay of the less skilled relative to that of the skilled manual worker has been attributed in part at least to the increased unionization of the unskilled. Evidently, policies of organized groups will account for some part at least of the position of particular occupations in the pay structure. Power can also be endowed by management in the form of managerial authority, and at least a part of a hierarchy of pay within an organization is likely to be an imposed means of distinguishing and upholding the strata of the organization's internal power structure.

Value. A third theory treats the differences in pay for different jobs as corresponding to differences in their content or requirements. The simplest form of this theory was embodied in the labour theory of value, whether in the system of Adam Smith or of Karl Marx, by the assumption that different kinds of labour can be reduced to different quantities of "homogeneous labour time," and that rates of pay are then simply proportional to those quantities. Job evaluation, discussed above, purports to condense the varied requirements of each job to a single figure in a common scale in order that the ranking order of the rates of pay of the jobs may be brought into conformity with that of those figures. But the assumption that, if two articles are priced in the same currency, they must contain quantities of a common substance is gratuitous. The impossibility of establishing the existence of such a substance and measuring the amount of it in any article drives both the labour theory of value and job evaluation into the circular argument of inferring the job content from the rate of pay and then explaining the rate of pay by the job content.

THE SUPPLY PRICE OF LABOUR

The foregoing directs attention to the supply price of labour to the job—the rate that must be paid if employers are to be able to attract and retain the quantity of labour that they wish to employ at that rate. Entry into an occupation generally imposes certain monetary costs; there may also be subjective costs, for example, in the effort of concentration required by preparation for examinations. The exercise of any occupation may be attended by disadvantages that require monetary compensation or may provide satisfactions and amenities that make workers willing to accept lower pay. For each occupation the various costs and benefits can be set off against the pay, and entrants will choose the one in which the prospective balance of advantage seems greatest. If more workers are to be attracted to and retained in a given occupation with unchanged conditions on the side of supply, the rate of pay in that occupation must be raised relative to others. An extension of supply will work to the opposite effect: for instance, if there is more public provision for secondary and tertiary education, and if rising standards of living enable more families to bear the costs of training, then a given number of workers will come to be available in a given occupation at a lower relative rate of pay. Here is to be found the reason for the occupational pay structure extending over a smaller range in developed than in poor countries and for the reduction in the margins for skill and the relative rate of pay for clerical work in the developed economies during the present century.

Education and the lowering pay of skilled workers

A number of considerations thus indicate that changes in the supply of labour influence its relative wage, although it is quite another thing to affirm the general theory of prices and assert that the rate of pay in any occupation tends to equality with the long-run supply price of labour to that occupation. The highly subjective nature of many of the costs and benefits involved in labour supply, and their dependence upon socially determined norms, strips the notion of a long-run supply price of any practical meaning. Nevertheless, in the absence of an extension of supply, a fall in the relative rate of pay of an occupation will bring a check to recruitment, followed by some withdrawal to other jobs of those already in the occupation. A rise in the relative rate of pay needs longer to take effect where proficiency takes long to acquire. Some types of proficiency may be limited by nature, and the rise in the rate of pay that follows on an extension of demand for them constitutes an economic rent—*i.e.*, a payment that is not required to maintain supply. In general, however, given time, the number of proficient workers available to follow a given occupation will be increased by a rise in the relative rate of pay it offers.

PRODUCTIVITY AND THE PRICE OF LABOUR

Orthodox economic theories. The complex character of labour as a commodity is nowhere more evident than in the relationship between pay and productivity. According to conventional economic theory, productivity should be the straightforward determinant of the employer's demand for labour. An employer who wishes to maximize profits will continue to recruit only up to the point at which the extra output gained from another worker equals the wage that worker is to be paid. This theory of marginal productivity lies at the heart of the orthodox economic theory of labour.

The theory of marginal productivity

The value of the theory, however, has come under question. Empirically minded economists note the profound difficulty of applying the theory when the productivity of individual labour in most organizations is unmeasurable and wage structures are internally connected. Aware of its weak analytic power, contemporary theorists in the orthodox tradition have suggested minor elaborations. Noting

a number of apparently discordant empirical and institutional features of the labour market, they have tried to bring them into the scope of formal economic analysis. Thus, taking the finding that local labour markets support a wide range of wage rates for a given grade of labour, search theory has tried to explain the phenomenon as a product of imperfections in information about available jobs and the consequent cost of searching. The same phenomenon is addressed by efficiency wage theories, which propose that the higher-paid occupants of a job grade are also achieving above-average productivity. Implicit contract theories, noting the considerable duration of most labour contracts, account for it as a necessary cost in the effort to overcome the difficulty of monitoring an employee's performance. That wages do not fall to levels that might, according to orthodox theory, eliminate unemployment is explained by insider–outsider theorists as a reflection of collusion between self-interested parties—in particular, those possessing jobs.

Such theorizing is promising, but it has shown relatively little explanatory power. It remains limited by a highly constraining view of the worker as an individual of purely rational motives and by an inability to grasp the significance of collective norms and behaviour in labour matters. It fails to explore the consequences of a world of imperfect product and labour markets, and its blindness to the open-ended character of the employment relationship prevents it from analyzing the significance of the varied institutional devices with which managers try to elicit productivity.

Empirical, multidisciplinary analysis. Something can be learned from the way in which employers manage productivity in practice. Employers pay great attention to internal pay structures, using job evaluation and other techniques to assure a stable and controlled structure of status within the work force. They give less detailed attention to what other employers are paying, so long as they keep the general level of pay increases broadly in line. They generally use specific incentive schemes more to generate an atmosphere of cooperation and flexibility than to make people work harder. Improvement in labour productivity comes overwhelmingly from technological change, requiring employees not to work harder but to work differently. The stress of such change to the employee is essentially temporary. It involves working with different workmates, facing the daunting challenge of learning a new skill, mourning for the lost opportunity to perform a skill of which one was once proud, and so on. Managers typically respond to the stress of meeting technological change with a temporary payment, although it may not seem temporary at the time. They may, for instance, introduce a new skill grade to match the new technique but remove it in a grade restructuring after a few years when memories of the upheaval have dimmed.

Employers are primarily concerned with unit costs, which involve both absolute wage levels and labour productivity. This concern arises from the pressures of the product market, which tend to override opportunities to pay what the labour market will bear. Employees, by contrast, think primarily of relative wages, especially very local relativities, and have only a fitful, vague, and temporary concern with their own productivity. They also place a relatively low priority on the opportunities offered by alternative wage levels in the labour market. It is in meeting this asymmetry of aspirations that the successful management of productivity lies, requiring constant tactical skill and personal attention on the part of employers.

Weakness of marginal-productivity theory

Such management implies the antithesis of marginal-productivity theory for two reasons. The first is that in a complex organization the productivity of the individual means nothing, and the productivity of the overall organization means everything. The second is that, in a world of imperfect markets, expecting prices to approach equilibrium in just one—the labour market—misses the important fact that competition is a total process, pursued on many fronts, such as design, marketing, and labour productivity—of which a competitive price for labour is only one.

Labour is by any standards an exceptional commodity. The quality of it is molded by its social context, and it is able to influence the shape of its own markets. Only a multidisciplinary analytic approach can unravel this complexity. The competitive forces of the economists' marketplace do indeed have a substantial impact upon the price of labour, although through more than just the specific market for labour itself. The level at which these forces are most evident is at that of broad aggregates and long time spans.

Movement of the general level of pay

A wage is a price, and the rise of the general level of wages or rates of pay in the course of time has, to some extent, been part of the long-term rise in the general level of all prices—that is, of the cumulative depreciation of the purchasing power of money, largely attributable to increases in its quantity. In another way, however, the movements of rates of pay have been an independent cause of the rising trend of prices. At times those rates rose in common with prices under the pull of monetary demand (in times of inflation, during war, or in the rising phase of the trade cycle), but when the demand fell off they were resistant to cuts; and though they were cut somewhat, they commonly remained at a higher level than when the preceding rise began. A graph of product prices shows big falls as well as big rises, and sometimes a falling trend persisting for many years together; but a graph of money wage rates is more like a flight of steps. This characteristic of wage movements puts a floor under prices and provides a higher starting point for the next upward movement, so that the fluctuations of monetary demand impose a rising trend on prices. In addition, the analysis of cost inflation under full employment, noted above, has shown that when employers generally expect demand to be sustained, rises in pay may occur in the absence of excess demand, thereby initiating rises in prices; and it is possible that the same process may have played some part in the rising phase of the trade cycles of earlier years.

The level of real earnings

The rise of real earnings may be traced by comparing the movements of earnings in money with those of an index number of the prices of the articles on which pay is typically expended. Such comparisons indicate that between 1860 and 1960 the real earnings of manual workers rose fourfold in France, Germany, and the United Kingdom; more than fivefold in the United States; and more than sevenfold in Sweden. In considering the standard of living attendant on these movements, it is necessary also to take account of the prevailing reduction in the size of the family, the complex effects of urbanization on the amenities of life, the effects of changed techniques and deployment between occupations on the strains and satisfactions experienced in work, and the reduction of hours of work. The last element has been extensive: it appears that down to World War II the wage earners of the five countries mentioned, save the United States, gave up from a third to a half of the potential increase in annual purchasing power in favour of a shorter working week and longer vacations.

To the extent that real earnings are measured simply by the quantity of consumables that money earnings will buy, their rise has depended on three factors: productivity, or the output per worker in terms of his own product; the share of this product that accrues to the worker; and the rate of exchange between the worker's own product and the goods and services he buys. In the industrialized countries, the last factor has presented itself largely in the form of the terms of trade between manufactured products and primary products, especially foodstuffs: real earnings have risen faster or slower according as a representative consignment of manufacturers may be exchanged, at the prices of the day, for a greater or smaller quantity of foodstuffs and raw materials. There have also been variations from time to time in the second factor: the share of the product accruing to the worker. The effect of the last two factors, however, has been small in comparison with that of the first, the rise of productivity. The salient finding from the statistical record of the last hundred years is that real earnings per worker have risen very nearly in the same proportion as output per worker. (E.H.P.B./W.A.Br.)

STRUCTURE OF THE LABOUR FORCE: EMPLOYMENT AND UNEMPLOYMENT

The three sectors

In more traditional societies, those in which the productivity of agricultural labour is very low, virtually the entire population must be employed in farming. As a country develops, it becomes able to feed its people more easily, so that a smaller proportion need be employed in farming. When productivity reaches a certain point, the demand for primary goods (agricultural products) drops in relation to the demand for other goods, such as clothing, shelter, and manufactured products. The production of these so-called secondary goods ultimately becomes organized in factories and expands dramatically. As the demand for other manufactured goods grows and remains high, employment in the secondary sector remains high as well. There is also a third, or tertiary, sector of employment comprising such activities as the service trades, teaching, administration, scientific research, medical care, art, and tourism, as well as other pursuits that are not carried on in factories. Technical progress in this sector is slight, compared to that in the primary and secondary sectors, and so the activities of the tertiary sector require a large number of people. In countries with high living standards, the demand for products of the tertiary sector keeps increasing; thus employment in that sector increases more rapidly than in the others.

The effect of technology on employment patterns

The great decrease of employment in the primary sector is one of the most important phenomena of modern history. People who abandon the soil must change not only their means of livelihood but also their residence and their way of life, all in the course of a single lifetime. This migration from a rural culture to an urban industrial one has involved many millions of people. For a long time the migration was chiefly toward the factories; as recently as 1930 it was imagined that the world of the future would be one immense factory, but this has proved unfounded because, though the output of factories has continued to grow, factory employment has not increased. Employment in the secondary sector in the United States reached its peak in the period from 1920 to 1970 at about one-third of the labour force. Since then the expansion in employment in the United States has been in the tertiary sector; without that expansion technological progress would have led to much greater unemployment or to much-reduced working hours.

Since the beginning of the Industrial Revolution people have feared new machinery would lead to unemployment, with machines replacing people in their work. This fear often led to violent attacks upon machines, as in England in the 18th and early 19th centuries. It now appears that the fear of technological unemployment is exaggerated.

Two kinds of technological progress

To understand the effect of technological progress on employment it is necessary to distinguish between two kinds of such progress. First, there is so-called intensive progress, which involves an increase in the efficiency with which people exploit nature: in agriculture, for example, all of the techniques that make it possible to extract products from the soil in a shorter period of time per worker. Intensive progress obviously results in a reduced need for workers. But there is another kind of progress, called extensive progress, which enlarges the exploitation of nature. Thus the discovery of raw materials enlarges the framework of the economy, and the industrial development based on it creates new kinds of employment. The invention of a new product, such as the automobile or the radio, has the same effect. In France more than 1,400,000 people make their livings from the automobile, for example, whether as factory workers, garage mechanics, or drivers.

Intensive progress allows people to satisfy their existing needs with less labour; extensive progress, on the other hand, satisfies needs that either did not exist or were not satisfied previously and, consequently, creates new jobs. But the two kinds of progress do not automatically balance each other. In an economy that is not highly developed, technological progress in agriculture may allow the population to consume more food. In a more developed economy, one in which the standards of living are already at a high level, the same progress does not lead to increased food consumption; the end result, on the other hand, may be glut, with lower farm prices and unemployed agricultural workers. In a period of general technological progress, employment is likely to be unstable because production must adapt itself to the changing demands of the market.

The consumption of primary, or agricultural, goods eventually reaches a point of saturation. The consumption of manufactured products passes through a phase of increase and then one of relative decline, except when new products appear on the market that have the extensive effect previously described. It is the tertiary sector, however, that absorbs most of the labour freed by technological progress in the other sectors. Employment as a whole does not decline because of technological progress. In most wealthy countries neither the size of the work force nor the length of the work week has declined since about 1920. There have been economic crises resulting in unemployment, but the unemployment was eventually absorbed. A dynamic economy requires a work force that is mobile enough to move out of sectors in which technological progress has reduced the need for labour and into sectors in which labour is in short supply. This migration is not pleasant for those involved, since it is inevitably accompanied by some degree of unemployment or underemployment.

Measuring unemployment

In its simplest sense, unemployment means being without a job. But for the statistician unemployment is difficult to define precisely. In compiling unemployment statistics for the United States and other developed countries an unemployed person is defined as anyone who is capable of working and is actively seeking work but is unable to find it. Before a person can be unemployed in this sense he must be an active member of the labour force in search of a job. Students and homemakers perform work, but they are not considered employed unless they are paid; however, they are not considered unemployed unless they are actively seeking gainful jobs. Any person who, because of a handicap or discouragement, has not tried to find work—even if only to the extent of registering at a public employment office—is not counted in unemployment statistics. There is a category of young people who leave schools or universities and who do not find jobs commensurate with their education; in some countries the government recognizes them as unemployed and tries to assist them.

UNDEREMPLOYMENT

Outside of the basic category of unemployment, there are various degrees of nonemployment and underemployment. At one time unemployment was not clearly distinguished from nonemployment. Many people were not employed or, more often, were underemployed—that is, working sporadically and without a precise contract. A period of inclement weather can put day labourers out of work, and they often receive no aid of any kind. In the rural areas of many countries there are large numbers of people who may be called underemployed because they do not have full-time work even though they are fully recognized as part of the labour force.

Concealed unemployment

In modern industrial societies the statistics of employment sometimes conceal substantial amounts of unemployment or underemployment. This is because the un-

employed may be brought into factories where they are not needed and have insufficient work to do. If the economy is a competitive one, each firm or enterprise has an incentive to keep its own operations efficient enough so that it makes a profit, and this tends to prevent the hiring and maintaining of surplus workers. But where enterprises are subsidized or owned by the state, it is necessary to develop administrative and political controls to assure that work is carried on efficiently and in the interest of the public. These controls often do not work as well in enforcing efficiency as do those of a profit-making economy; they are often biased or erratic, and they may be subject to the interference of pressure groups, politicians, or bureaucrats.

Ever since the 1950s the communist countries of eastern Europe have been concerned about the problem of running socialized enterprises efficiently; the planners and economists of these countries have conceded that those enterprises that are not subject to the pressures of the market have more difficulty than others in adjusting their production to the needs of consumers and consequently in working out optimal plans and raising the efficiency of their labour force. It is possible to reduce unemployment and underemployment on a national scale, even to the point where there is an apparent shortage of labour, while at the same time permitting a considerable amount of misemployment with its accompanying problem of inefficiency.

RATES OF UNEMPLOYMENT

During the Great Depression of the 1930s, unemployment in the Western capitalist countries reached extremely high rates. In the United States the proportion of unemployed as counted by government statistics reached 25 percent of the labour force in 1933. In Sweden unemployment reached 25 percent twice—in 1921 and 1931. In the United Kingdom the rate was more than 15 percent in those two years. From the 1940s to the early 1970s unemployment rates were far below catastrophic levels.

Varying incidences of unemployment

Even when the overall rate of unemployment is low, it remains a serious social problem because its incidence is not the same for all categories of workers. While some may be only temporarily out of work or moving from one job to another, others may be unemployed for long periods of time. The most damaging unemployment, for an individual or for a family, is that which persists for a long period of time. The statistics published by most countries do not distinguish among categories of unemployed workers. The U.S. Bureau of Labor Statistics, however, has thoroughly studied the incidence of unemployment. Its figures show that unemployment is higher for women than for men, for blacks than for whites, for young people than for adults, and for those who lack education and special skills as compared to those who have such qualifications.

In developed countries governments collect increasing quantities of information on the labour force. Concise statistics, however, cannot adequately portray the whole spectrum of employment problems. Some workers remain in the same occupation, the same business, and even the same firm for decades, while others change jobs frequently. Some people work at more than one job at a time. Some have seasonal jobs. Some people work only a few hours a month, while others may work as many as 65 hours a week. Some pass through a number of jobs searching for one that is enjoyable and pays well. A substantial number of people cannot be easily classified as being inside or outside of the labour force. Many who are outside the active labour force might be drawn into it if jobs were obtainable or if help were available to assist them in qualifying for jobs that do exist. This sort of hidden unemployment may be more serious than overt unemployment if personal and social handicaps are combined with a shortage of work opportunity.

Government efforts to increase employment

Since the end of World War II the governments of most of the developed countries have become committed to a program of reducing unemployment and underemployment. In France, for example, the constitution explicitly charges the state with assuring full employment. Similar goals have been affirmed with varying degrees of concreteness in other countries. The Charter of the United Nations makes full employment a major objective of its members.

Approaches to employment policy

Governments have followed various policies in pursuit of full employment. One general approach is to try to improve the quality of the labour supply. Another is to try to alleviate the effects of unemployment and underemployment. A third approach seeks to maintain economic activity at a high level through monetary and fiscal policies. Finally, there is the method of economic planning attempted in France, where the government planning commission sets targets for the various sectors of the economy that are linked to forecasts of manpower, in the belief that the difficult and complex problems of employment and unemployment cannot be separated from other problems of economic and social development (see ECONOMIC GROWTH AND PLANNING).

Labour policies. The first consistent efforts to deal with the instability of employment caused by technical and economic progress involved providing information on the state of the labour market and on the qualifications of those seeking work. Public employment services developed, particularly during the 1930s, which, on the basis of this information, were to direct applicants more efficiently to existing jobs or to help them prepare for occupations and careers in which labour needs seemed likely. Most of the large industrial nations now have such agencies, which are designed to bring together the two sides of the labour market. Information leads to guidance, and guidance leads to education and training or retraining. The rapid progress of technology has required a perceptible raising of the qualifications of workers: the number of skilled labourers has increased more than that of the unskilled, technicians more than skilled labourers, and engineers more than technicians. In response to this trend, many governments in industrial nations have endeavoured to either provide or subsidize training programs for those members of the work force who might be able to benefit from them.

Often unemployment is a regional matter. Attempts have been made to mitigate it through regional development programs in a number of countries.

Ameliorative policies. All of the developed countries try to soften the impact of unemployment through unemployment compensation and other forms of social insurance. In some cases governments seek to induce employers to retrain workers for new jobs rather than to lay workers off.

The maintenance of aggregate demand. The use of monetary and fiscal policies to keep the economy functioning at a high level of employment has been undertaken in a number of countries since the Great Depression (see GOVERNMENT FINANCE: *Fiscal and monetary policy*). In periods of recession or of growing unemployment the government may increase aggregate demand by expanding the money supply or by increasing its own spending. This approach had some success in the United States during the 1960s, when a major income tax reduction, tax incentives for business investment, and a large increase in federal expenditures brought the unemployment rate down to 3.5 percent in 1969. In the years that followed, however, the United States was faced with serious inflation, and government policies that were then undertaken for the purpose of stabilizing prices had the effect of causing unemployment to rise.

National economic planning. In the 1950s and '60s the French government undertook national planning on a broad scale and succeeded in maintaining rapid economic growth, full employment, and a stable currency; French national planning was much less effective during the severe recession of the 1970s. It seems likely that employment policies in most industrial countries will aim in the future toward a context much broader than that of placement bureaus, labour mobility, and fiscal and monetary policies. The speed which society is evolving, the variety of possible ways of life, and the unforeseeable changes in the human condition make it increasingly necessary to view economic phenomena as only one part of civilization, inseparable from the rest. (J.Fo.)

MIGRANT LABOUR

Workers who move about systematically, seeking and engaging in seasonal temporary employment without becoming residents of the areas in which they work, are called migrant labourers. Throughout the world the wages, working conditions, and standard of life of migrant workers are low in comparison with other groups of workers, almost always the lowest. The union- and government-established labour standards available to industrial and regular agricultural workers generally do not exist for migrant labour; and where there are statutes their implementation is limited and irregular. Migrant labourers are the rural equivalents of the urban poor.

Migrant labour is primarily but not exclusively agricultural; it generally involves harvesting activities and is mostly manual, repetitive, easily learned, and demanding of almost no skill. Migrant workers pick cotton, cocoa beans, grapes, tea or tobacco leaves. They cut sugarcane and pineapples and gather artichokes, beans, carrots, sugar beets, or other cash crops. They also harvest corn, small grains, or wheat. Because all of these activities require physical strength and call for much bending and reaching, migrant labour is also called "stoop labour."

Characteristics of migrant labour

ASSOCIATION WITH INDUSTRIALIZATION

Migrant labour is associated with the developed or middle stages of industrialization and its technology and organization. While the demand for migrant labour stems, in part, from the seasonal nature of worker requirements in agriculture and related industries, it is not a phenomenon characteristic of agricultural societies. It was probably unknown in antiquity and is not significant today in the developing economies of Asia and Africa; high levels of demand for migrant workers are related to rapid increases in farm output, decreases in permanent farm employment, long periods of rural migration, increasing urban wages, and fairly rapid mechanization and commercialization of agriculture. The demand for migrant labour is particularly high in the case of market farming, where production is increasing or where numerous small farmers specialize in one or two cash crops and are joined in production or marketing associations.

The propensity to work as a migrant labourer stems from the unfavourable economic and social conditions of the worker's "home base." Technological displacement, radical changes in sharecropping systems, long business recessions, drought and crop failures, and various forms of discrimination are factors inducing workers to enter the migrant stream. Lack of education, industrial accidents, and ill health also force many to become migrant. The supply of migrant labour thus develops in response to the push of need, not from a romantic desire to wander. In a minority of cases, particularly those in which a poor region, state, or nation is adjacent to a much richer one, workers respond also to the pull of better opportunities; in spite of its misery and degradation, migrant labour appears to them as a way to a higher standard of life in the future.

Effect of economic development

As national economies mature and move through the secondary (mostly industrial) stage to more developed stages, both the demand for and the supply of migrant labour decrease. The demand for it decreases because of mechanization; in the United States, for example, the harvesting of wheat and cotton that required the employment of many migrants before World War II is now largely mechanized and done by regularly employed farmhands. The supply decreases as higher real standards of living and greater employment possibilities make migrant work less attractive. For a small number of employed workers in affluent areas, migrant labour provides secondary (or additional) employment opportunities.

In mature economies migrant labour contributes little to total agricultural output and a negligible amount to nonagricultural output. Nevertheless, the availability at the right time and place of migrant workers can be crucial, because without them large crop losses may occur. This may spur private employers, planning authorities, and governments to take remedial actions to improve the conditions of migrant labourers in order to assure their continued availability. If native migrant workers are in short supply, the gap may be filled by the temporary—legalized or tolerated—migration of foreign workers from poorer countries. In the communist countries, where the use of migrant foreign labour is politically unacceptable, the flow of migrant workers is supplemented with organized brigades of urban youth encouraged to contribute their labour voluntarily to the common good.

THE MIGRANT FLOW

Most migrant workers move in well-established patterns or "streams." The migrants usually flow from the south to the north (in the Southern Hemisphere, however, they move in the opposite direction), as geography makes for later harvests in northern latitudes; also, the poorer areas with the labour surpluses tend to be in the south. Migrants make detours and side trips as they hear of better or new opportunities. The various crews, families, or individuals that make up the migrant stream do not all follow the same pattern: some move from place to place, others make a big journey to a set of places with known jobs, while others make only one move and then go back to home base.

Economic and social conditions of migrant workers

In all employment relationships the matching of unemployed workers and job vacancies is a haphazard and complicated process, but the labour market for migrant workers is notably disorderly. One basic reason is that the employment relationship is ephemeral. With minor exceptions, employers do not systematically recruit migrant workers and have no regular personnel specialists or procedures.

LIVING CONDITIONS AND WAGES

The conditions under which migrant labourers live are abysmal. Their misery has been formally documented by investigative bodies down the years: the U.S. Industrial Commission of 1901, the Royal Commission of Labour in India of 1933, the U.S. President's Commission on Migratory Labor of 1951, and others. Novelists, moved by the migrant workers' plight, have aroused sympathy with such masterpieces as John Steinbeck's *The Grapes of Wrath* or Mulk Raj Anand's *Two Leaves and a Bud.* Social scientists, particularly economists, however, have tended to ignore the issues posed by migrant labour. Governmental statistics on migratory labour are scarce, often contradictory, and approximate at best.

Migrant workers are often cheated even of the low wages they earn. They work long hours and under exacting requirements. Total employment is always limited; in a year most migrants work less than one-quarter of the working days. Child labour is widespread, and the children who do not work are not adequately provided for; they often do not go to school, as in most places schools are open only to residents. Housing amounts to hardly more than a roof over the head, often tar or tent. Tuberculosis, infant and maternal mortality, dysentery, enteritis, smallpox, and typhoid are all more prevalent among migrant labourers than among the general population. Literacy levels, social cohesion, and political participation rates are always exceptionally low among migrant workers, while crime and all other indexes of unacceptable social behaviour are high. Contributing to this picture of misery is the fact that the migrants, whether national or foreign born, are fundamentally alien to the community in which they work. The local population, if eager to see them come, is even more eager to see them go. They have difficulty in gaining access to local health and social services, are often

deprived of their rights, do not enjoy easy recourse to the courts, and are abused by exploiters and sometimes the local police.

DIFFICULTIES OF REGULATION

The fact that the migrant is "here today and gone tomorrow" makes the regulation of his working and living conditions difficult. Local, state, and national authorities generally acknowledge the existence of serious problems associated with migrant labour, yet in the face of these, generally claim they are unable to deal with them effectively. The International Labour Organization has consistently prodded governments to expand their legislation in this area. Efforts are being made to develop regulatory standards protecting workers during their journeys, to discourage migratory movements of undesirables, and to stabilize workers and their families near their place of employment.

The task of providing migrant workers with governmental protection and assistance is full of complex difficulties. Regulatory legislation ideally should be applied at the home base of the migrants as well as along the stream of migration and at the place of employment. However, advances in one area often create difficulties in another due to uneven social and economic conditions. On the one hand, mechanization and permanent migration is reducing the number of migrant labourers, while, on the other, relocating them is difficult because their low educational attainments and lack of skills force them into urban poverty. Where legislation has been passed and improvements have been achieved, the major effect has been to force employers to provide more housing and sanitary facilities and to regularize the activities of the labour contractors.

Patterns of movement

THE UNITED STATES

In the United States in the 1960s about 400,000 people yearly were occupied in migrant work. In the previous decade the number was about 600,000 yearly; in 1940 there were 1,000,000; and it is possible that in the middle 1920s as many as 2,000,000 workers were in the migration streams. There are not many reliable statistics. Interpretation of scattered data suggests that the following trends and relationships have occurred since World War I: (1) the migrant labour population has been declining, albeit at a decreasing rate, (2) migrant agricultural workers have tended to become a larger proportion of all agricultural (nonfamily) workers, (3) foreign workers—while their participation varies with the rulings of the federal immigration administration and with requirements of foreign policy—constitute an increasing proportion of all migrant workers, and (4) the size of the national migratory labour force varies inversely, and the foreign labour force directly, with improvements in the total level of employment.

In the 1960s more than half of the migrant workers were 25 years or younger; fewer than one in 10 was 55 years or older. Three out of four migrant workers were males, with males outnumbering females in all age groups. Almost half of the female migratory workers were between the ages of 25 and 54. In terms of areas of permanent residence a little less than two out of four migrant workers originate in the South. Over 55 percent of all migratory workers had eight years of education or less.

Three major patterns characterize the movement of migrant workers in the United States. (1) On the East Coast, migration begins in winter when workers from Southern states move south to Florida to work on the citrus crop. From there, joined by others coming from Texas and Puerto Rico, they move north along the coast ending up in August and September gathering tomatoes and other vegetables on Long Island and potatoes as far north as Maine. Blacks constitute a large proportion of those on the East Coast. (2) The largest stream of migrants flows north and west from Texas in the spring to the north central, mountain, and Pacific Coast states. The crops they harvest are fruits and vegetables, sugar beets, and some cotton. The workers are natives of Mexican descent with whom some immigrants from Mexico join; they travel mostly with their families and end work early in December. (3) A large number of Spanish-Americans work on harvesting vegetables, moving from southern California northward through the Pacific Coast states. Thus some migratory farm work is done in almost every state, with most labourers coming from Texas and the Southeast. About half of the mandays of such employment occur in California, Michigan, Texas, and Florida.

The principal U.S. migrant flow

These migrant workers suffer from unemployment, inadequate housing, little education, exclusion from normal community life, and the denial of services that others take for granted. In the middle 1960s they grossed an annual average income of about $1,700, working an average of 150 days if they worked in nonfarm activities as well. Those who worked only as migrant farm workers averaged $800 for about 82 days of work.

In spite of this, it is probably true that the 1960s saw more improvements in the living and working conditions of migrant workers in America than any previous decade. A number of states and localities established special migrant committees to implement and expand legislation. "Crew-registration laws" were increasingly adopted by the states, and their implementation is having beneficial effects. Most important, after years of inactivity attempts were being made by labour unions to organize migrants, with the United Farm Workers achieving particular success in California.

WESTERN EUROPE

Migrancy as described above does not involve very large numbers in Europe because of the longer history of urbanization. Migrant agricultural workers are still active in the rice harvest of northern Italy and in some of the vegetable harvest of Portugal and France.

The European migrant worker frequently moves across national boundaries in response to seasonal fluctuations in the labour demand of specific industries. This type of migration rarely involves the workers' families and generally calls for only one move. Thus Italian workers go to work in Swiss hotels, Polish workers go to mine English coal, and so on. The problems of European migrants are less acute than those described above; they generally involve difficulties in acculturation and in qualifying for national health and social service programs.

AFRICA

Migrant labour is one of the most important problems of the developing nations of Africa, where migrancy is clearly an intermediate stage of the movement from rural agricultural to urban industrial life. The phenomenon does not conform to the general characteristics described earlier. Workers migrate to urban rather than to agricultural employment; they work with one rather than many employers; and they often stay in one job as long as three years.

Migrancy in Africa is very clearly a response to the pull of expanding industrial opportunity and the lure of city life rather than to a push away from the poverty of rural life; it is thus a step in the process of building an industrial labour force. The worker who moves does so alone and with considerable reluctance. He often goes because he needs cash for a particular purpose (*e.g.*, the bride-price). He is not prepared to give up his contacts with the village, where there is security in old age, sickness, or periods of unemployment. African migrancy is nonetheless quite similar to the migration that occurs in the middle stages of industrialization. In the regions where the migrant workers come from one finds the decline of traditional subsistence agriculture, changes in ownership patterns, and drought; one rarely finds workers displaced by mechanization of production.

Pull versus push

Migrant labour is also to be found in many other nations. India has probably more migrant workers than any other Asian country; these are involved mainly in the harvesting of tea, cotton, and rice. In Australia and the southern nations of Latin America migrants work on ranches more often than on farms and are engaged in wool shearing and meat processing. (O.A.O.)

ORGANIZED LABOUR: TRADE UNIONISM

To be understood in its international context, trade unionism must be examined as part of a wider concept—the labour movement as a whole. That movement consists of several more or less intimately related organizations such as labour parties, workers' mutual insurance organizations, producers' or consumers' cooperatives, and workers' education and sports associations. All have the common objective of improving the material, cultural, and social status of their members.

What distinguishes one organization from another is the particular aspect of that broad objective it is endeavouring to pursue, and the particular method it employs. The relationship among the various parts of the labour movement varies from country to country and from period to period. Not all countries have produced the entire gamut of organizations referred to above; in some countries the term labour movement is virtually synonymous with trade unionism.

Origins and background of the trade union movement

Early forms of labour organizations. Unions originated, mainly in Great Britain and the United States in the late 18th and early 19th centuries, as associations of workers using the same skill. There is no connection between trade unions and medieval craft guilds, for the latter were composed of master craftsmen who owned capital and often employed several workers. The early unions were formed partly as social clubs but soon became increasingly concerned with improving wages and working conditions, primarily by the device of collective bargaining. Progressing from trade to trade within the same city or area, the clubs formed local associations, which, because they carried on their main activities on a purely local level, were almost self-sufficient. With industrial development, however, local associations sooner or later followed the expansion of production beyond the local market and developed into national unions of the same trade. These in turn formed national union federations.

Factors favouring unionism. The unions of the early 19th century were almost exclusively based upon a particular craft. However, as mass production industries—which required large numbers of rapidly trained, semiskilled workers—developed, a trend toward large-scale union organization grew, and toward the end of the 19th century Great Britain was including unskilled workers. Unions that recruited members from such groups—whose ranks were expanding rapidly as a result of new technologies—emerged either as industrial unions or as general unions. Industrial unions attempted to organize all workers employed in producing a given product or service, sometimes including even the general office or white-collar workers. General unions included skilled workers and labourers of all grades from different industries, even though they usually started from a base in one particular industry. Changing technologies, union mergers, and ideological factors led to the development of various kinds of unions that would not fit easily into any of the above categories.

Labour unions as protest movements

Obstacles to union organization. In most Western countries, labour movements arose out of the protest of workers and intellectuals against a social and political system based upon discrimination according to ancestry, social status, income, and property. Such a system offered few avenues for individual or collective advancement. Discrimination in political franchise (restrictions on or outright denials of the vote) and a lack of educational opportunities, antiunion legislation, and the whole spirit of a society founded upon acknowledged class distinctions were the main sources of the social protest at the root of modern labour movements.

Antiunion laws such as the Combination Laws in Britain, the French Loi Le Chapelier (forbidding combination), similar laws in most Germanic countries, and court decisions under common law in the United States in the 19th century made it inevitable that reformers would first turn their energies to the removal of those legal impediments. The demand for equal voting rights became an early preoccupation of the labour movements, partly because of a desire for recognition of the worker as a full citizen and partly to give the worker political power to remove antiunion laws and to obtain legislation for improving working conditions, especially for children and women. At a much later stage, when democratic equality was substantially achieved, unification of unions that had been divided according to political or religious sympathies became the next goal of the labour movements, with less attention paid to anticapitalist ideologies.

Union development in industrialized nations

General differences

Among the various labour movements the roles assigned to collective bargaining and to labour legislation vary substantially; such differences even occur within a single movement. In Great Britain and the United States the role of collective bargaining is stressed. Legislation is regarded primarily as a device to bring about, facilitate, and supplement orderly collective bargaining. Most continental European countries, though with considerable variations among them, rely on substantive labour legislation to a far higher degree than do Great Britain or the United States.

Collective agreements in Britain and many other Western countries, either traditionally or under the impact of tight labour markets, frequently designate the minimum wages allowable; the actual wage rates may be considerably above those listed in the contract. Prolonged full employment and high prosperity may tend to weaken the allegiance of workers to their union, because wages then tend to rise more from an increasing demand for labour than from union pressure. Such a weakening of allegiance to the union is particularly noticeable among younger members who did not live through the "heroic era" of a union's early struggles for recognition and improvements in low wages and poor working conditions.

Changes in the character of unions

The "generation problem" within the unions is complicated by rapid and substantial changes in the structure of the labour force in the industrial countries. The decline in agricultural jobs, the emergence of the semiskilled workers in the mass production industries, and the rise in the numbers of white-collar workers and employed women have changed not only the economic and social milieu within which the unions operate but also their recruiting potential and their character. White-collar workers have formed or joined unions in increasing numbers, though in lesser proportions than their blue-collar colleagues. A number of quasi unions also have emerged—professional associations that, while often refusing to be classified as unions, increasingly behave as unions do and are sometimes prepared to cooperate with or even join confederations of manual workers.

GREAT BRITAIN

Early history. Britain was a pioneer in the development of modern industry, and its unions are among the oldest in the world. They go back to the late 18th century when artisans, threatened by the establishment of mechanized factories, formed local clubs for their common defense. Their efforts soon met with government hostility, based at first upon the common law concept that unions were in restraint of trade. Later, under the impact of the French Revolution, the Combination Acts of 1799 and 1800 declared unions to be criminal conspiracies against the public. Even after those laws were repealed in 1824, legal and administrative measures against unions continued for a long time.

In 1834, for example, six Dorsetshire labourers from the village of Tolpuddle (the Tolpuddle Martyrs) were sentenced to seven years' transportation to a penal colony in Australia for administering unlawful oaths in their trade union activities. It was only in 1871 that legislation was

passed that exempted unions from prosecution because of "restraint of trade." Local trade clubs had survived intermittent prosecution, and in 1834 the reformer and socialist Robert Owen attempted, without lasting results, to organize them in a national movement—the Grand National Consolidated Trades Union.

In 1851 the Amalgamated Society of Engineers established the pattern for the "new model" unions—national associations of skilled workers who had passed through a regular apprenticeship. Their main economic weapon was their ability to restrict the number of apprentices, thereby limiting the supply of skilled labour. They also functioned as "friendly societies"; that is, they provided mutual assistance in case of sickness or for emigration.

During the same period some of the leaders of the craft unions formed the so-called Junta, which engaged in political action and joined Marx's International Workingmen's Association (the First International), even though most of them failed to understand—much less approve—the principles of Marxian socialism. This was not the first attempt of labour leaders to engage in political action. In the 1830s some unionists had joined a small group of radical members of Parliament in drawing up a petition for universal male suffrage, which became part of the platform of the Chartist reform movement. The Chartists were concerned mainly with electoral reform, but trade unions played an important part in their activities. Chartism was the first specifically working-class national movement in Britain. But the emergence of the working class as a major political force had to await the entrance of the miners into the political arena, the extension of the right to vote to large numbers of workers, and the organization of less skilled workers.

The legalization of unionism. After passage of the Trade-Union Act of 1871, which effectively legalized trade unions, unions of agricultural workers, seamen, gas workers, general and municipal workers, and dockworkers were organized. This period of union activity culminated in the successful dockworkers' strike of 1889. The new unions emphasized nationwide collective bargaining and demanded a legal minimum wage and an eight-hour day. Socialist ideas entered the movement, and political organizations of the working class—such as the Independent Labour Party—came to play an increasing role. Shortly after the turn of the century, the Labour Party was born.

The legal status of unions became endangered by the Taff Vale court decision of 1901, which held that unions could be sued for losses incurred by employers as a result of union action. That decision was nullified by the Trade Disputes Act of 1906. Shortly before World War I, financial support of the Labour Party by trade unions was legalized, provided that individual union members could "contract out." This meant that union members could choose not to pay the so-called political levy; that is, not make a contribution to the special fund that most unions established for support of the Labour Party. After the unsuccessful General Strike of 1926, that provision was changed into the "contracting in" system, which required union members to state expressly their wish to pay the political levy. As a result, a sizable proportion of union members failed to pay the levy, and in 1945 the newly elected Labour government once again established the principle of contracting out.

Contracting out

The growth of union membership. Between 1910 and 1920, trade union membership more than trebled, from 2,565,000 to 8,334,000. But the postwar slump of 1921 and the return to the gold standard in 1925 led to a great crisis in the history of British trade unionism—the General Strike of 1926. The failure of the strike resulted in widespread disillusionment, which was intensified by the start of the Great Depression in 1930. By 1933 union membership had fallen to 4,392,000, half of what it had been in 1920. Afterward union membership gradually increased until by 1969 it stood at more than 10,000,000, about 9,500,000 of whom were affiliated with the Trades Union Congress (TUC), the national organization representing British trade unions. The diversity in size and organization of British unions is tremendous. They range from small local craft groups to giant general unions with memberships in excess of 1,400,000. A trend toward large-scale union mergers became apparent in the 1960s.

Dual system of bargaining. The basic institutional factor underlying problems of industrial relations in Britain is the existence of two bargaining systems: the industry-wide contract negotiated by the union and the employers' association, and the unofficial agreements arrived at in the various plants by managers, shop stewards, and individual workers. The terms reached under the unofficial system are consistently more favourable for the workers than those in the official contract. Such a dual system has its roots in the failure of most unions to control the shop stewards, and of the employers' associations to control the plant managers.

Shop steward control of the going pay rates

Union control over the shop stewards proved weak during World War I when the stewards assumed an authority that often brought them into conflict with the union leadership. World War II, and the long period of full employment that followed, reinforced that trend. The Industrial Relations Act of 1971 brought the first comprehensive legislation in this field in the country's history. It provided, among other things, for legally enforceable collective agreements between unions and employers.

Role of government. Caught between rising earnings, rising prices, and a great reluctance to permit a devaluation of the currency, the post-World War II governments embarked upon a series of experiments with methods of wage restraint. The first, undertaken in 1948–50, had ended in failure; and from 1951 to 1964 the Conservative governments made several attempts to limit wage and salary increases, but the unions refused to cooperate. When a Labour government was returned in 1964, some of the unions became more willing to accept an incomes policy. As a result, union (and business) representatives took their places on a Prices and Incomes Board empowered to investigate and suggest reforms, which were mainly related to enhancing productivity.

A searching investigation of the entire industrial relations system by the Donovan Commission (Royal Commission on Trade Unions and Employers' Associations) was also undertaken. Among other things, the commission recommended an obligatory 28-day conciliation period before any unofficial strike (any walkout not sanctioned by the official union leadership) could begin, and also a poll of the union's members if a major official strike was imminent and if the government doubted that the union leaders had rank-and-file support.

Various emergency measures were taken when the exchange rate of sterling was threatened in 1966 and again after sterling's devaluation in 1967. They included the suspension of all wage, price, and dividend increases for various periods. However, in 1970 a government white paper admitted that "the use of a productivity, prices, and incomes policy for short-term purposes can only have a short-term effect. The long-term role of the policy is essentially an educational one."

The incomes policy put the responsibility for wage restraint on the unions. However, as mentioned above, effective earnings are determined not so much by the unions as by shop stewards, whose activities the union leaders do not entirely control. It was hardly surprising, therefore, that the incomes policy proved ineffective. Nor was it conceivable that the unions, created to improve their members' earnings, could for long be turned into devices for preventing their members from obtaining the wage increases that management was willing to concede. The creation of an Industrial Relations Commission and of a Commission for Industry and Manpower in 1969–70 indicated that new means were being sought to promote and preserve collective bargaining under conditions of full employment while at the same time keeping price increases at an acceptable rate.

UNITED STATES

The beginnings. Unionism in the United States goes back to the early days of national independence when craftsman employees in various trades—carpenters, masons, shoemakers, and printers—formed local groups to obtain shorter hours and higher wages. The Mechanics'

First steps toward wider organization

Union of Trade Associations formed in Philadelphia in 1827 represented an early attempt to combine unions of different crafts on a local level. The first steps toward extending the organization beyond the local area were made in the 1830s but were only temporarily successful. Much emphasis was placed on political action and the formation of labour parties. The main objectives included free education and the abolition of property restrictions on the right to vote, which had survived from colonial times in some states down to the Jacksonian era (1828–36). Once those objectives were achieved in the Northern states, the labour parties disappeared. Most of the unions—including the National Trades Union formed in 1834—were dissolved in the depression of 1837.

In the 1850s a new departure occurred. With improvements in transport and the consequent expansion of markets beyond local areas, unions in the same trade began to expand their fields of operation and to establish regional or even national federations. The National Typographical Union was formed in 1852 and was followed by five others. Concentration on job-connected problems rather than on broad, ideological issues was characteristic of this period.

The Knights of Labor. The only large reform movement among the early labour organizations was backed by the Noble Order of the Knights of Labor, founded in 1869. Its objectives, many of them more political than industrial, were vague and moderate and included the eight-hour day and the abolition of child labour. Its rise was spectacular, especially after it won a strike against the Wabash Railroad, controlled by Jay Gould, a New York multimillionaire, in 1885. After that victory, however, the organization soon declined—in part because of a growing antiradical mood in the country. That mood was intensified by the Haymarket bomb episode in Chicago (1886), in which seven policemen were killed and with which the Knights were quite unjustly associated. A more significant factor in their decline was the fact that the Knights combined skilled and unskilled workers in one organization. The mass immigration of unskilled workers weakened their economic power to the point that union victories could be won only with the assistance of skilled workers; and when they were no longer willing to give battle on behalf of the unskilled immigrants, the decline of the Knights became inevitable.

The American Federation of Labor. In the meantime, however, the American Federation of Labor began to supplant the Knights as a national federation of unions of skilled workers. Started with modest success as the Federation of Organized Trades and Labor Unions of the United States and Canada in 1881, it was transformed into the American Federation of Labor (AFL) five years later. Samuel Gompers, a cigar maker, was elected president, and his personality decisively shaped the AFL's evolution. It was to be a practical, nonideological movement, accepting the existing social system but attempting to change it in favour of the organized skilled workers. The main instrument of change was to be collective bargaining. The affiliated organizations were to be autonomous within their jurisdiction—that is, the job territory over which their charter gave them full authority. Organization was limited to skilled workers, who alone were regarded as capable of forming effective unions.

Membership of the AFL grew only slowly until the turn of the century, partly because of a predominantly unfavourable economic situation and partly because of government hostility. Between 1900 and the outbreak of World War I, however, membership expanded from less than 600,000 to about 2,000,000. The war itself opened new opportunities for organized labour. President Woodrow Wilson's administration gave Gompers semiofficial status by making him a member of a small advisory commission to the Council of National Defense, and the president himself appeared at the 1917 AFL convention. By 1920 the membership of AFL-affiliated unions had doubled to 4,000,000.

The next decade was a period of crisis and stagnation for unionism. Serious opposition to Gompers arose in 1921 in the person of John L. Lewis, newly elected head of the United Mine Workers. Gompers won, but his death in 1924 brought William Green, a member of the mine workers' union, to the presidency. The liquidation of the war industry and a sharp antiunion offensive by many employers—who were supported by conservative U.S. presidents, hostile courts, and antilabour sentiments of great sectors of the general public—combined to bring about a sharp decline in union membership. Under the slogan of the "open shop," by the use of so-called yellow dog contracts that required workers not to join unions, and by court injunctions against unions, employers staged a major offensive against organized labour. This not only kept the unions out of the steel industry but also reduced union membership to little more than its prewar level.

Rise of antiunion sentiment in the 1920s

The Wobblies. The 1920s also saw the decline of an important union movement that had attempted to organize unskilled workers into industrial unions—the Industrial Workers of the World (IWW). The Wobblies, as they were popularly known, had been founded in 1905 in Chicago in response to the AFL's opposition to unionism for unskilled workers. Consisting largely of itinerant industrial labourers, the IWW had concentrated on militant strike action, especially against employers hostile to unionism. The movement included a considerable ideological element and eventually came under the control of anarchists and syndicalists. That fact, together with the IWW's strong opposition to World War I, led to the suppression of its press and to prosecutions under the Espionage and Sedition Acts of 1917 and 1918. After World War I many states adopted legislation outlawing the IWW, thus driving it underground. It was further weakened by disputes with the communists, so that by the mid-1920s it had ceased to be a significant force within the labour movement.

The Great Depression and the New Deal. The Great Depression, which began in 1929, almost destroyed some of the unions, but it also gave unionism a new start. The reasons for this new turning point were several. With the depression and the shifting political climate, there came a radical change in labour legislation—the Norris-LaGuardia Act of 1932, which removed the legal basis for the use of court injunctions in labour disputes. (In a partial reinterpretation of that 1932 act, the U.S. Supreme Court in 1970 ruled that unions could be enjoined from permitting walkouts in violation of no-strike pledges contained in their contracts.)

The New Deal legislation that followed the victory of Franklin D. Roosevelt in the presidential election of 1932 guaranteed the right of workers to join unions of their own choosing and required management to engage in collective bargaining if the majority of the workers so desired. A National Labor Relations Board was set up to enforce the law. Although the law was revised later by the Taft-Hartley Act and the Landrum-Griffin Act, its basic principle that collective bargaining is the most desirable way of determining wages and working conditions was not changed. Even when strikes threaten national security, only delaying action is legally provided for; in some situations, however, governmental pressure to avoid or settle strikes has, in fact, come close to coercion. Minimum-wage legislation was passed on both the federal and state levels, and government encouragement of unionism was manifested not only during the New Deal but repeatedly throughout the presidencies of Harry S. Truman, John F. Kennedy, and Lyndon B. Johnson.

A basic change in unionism itself

The change in the governmental attitude toward unionism was accompanied by a basic transformation of unionism itself. Adjusting itself to the new, enlarged role of the government in economic affairs, as well as to the catastrophic consequences of the Great Depression, the AFL abandoned its traditional distrust of government regulation of industrial relations. With some reluctance it accepted and finally supported a government-sponsored social security system providing for gradual expansion of old age pensions, unemployment benefits, and—much later—health insurance.

The Congress of Industrial Organizations. No less important was the belated adjustment of the structure of unions to accommodate the growing number of semiskilled workers in the mass production industries, whose employ-

ees the traditional craft unions of the AFL had failed to organize. Under the leadership of the United Mine Workers' president John L. Lewis, eight unions within the AFL promoted organizing drives to establish industrial unions in such industries as steel, automobiles, rubber, and chemicals. When endorsement of his plan was defeated at the AFL convention in 1935, Lewis set up a Committee for Industrial Organization, which three years later was transformed into the Congress of Industrial Organizations (CIO). For having formed a rival organization the unions of the CIO were expelled by the AFL.

The decisive battles of the new organization were fought in the steel and automobile industries. A favourite weapon of the organizers—mainly mine workers' leaders assisted by intellectuals and a few Communists—was the sit-down strike, a strike during which workers do not walk out of the plant but remain idle at their work positions, thus paralyzing the plant and making it necessary for management to make an immediate response to their demands. Victories in the largest steel and automobile companies in 1937 were finally crowned by the capitulation of the Ford Motor Company in 1941.

World War II and after. The new competition induced the AFL in turn to engage in organizing drives. Total union membership grew from 2,900,000 in 1933 to 9,000,000 in 1939. By the end of World War II it had risen to 15,000,000, more than one-third of total nonagricultural employment. Just as had World War I, World War II induced union expansion, partly because of government assistance offered in return for a no-strike pledge.

The World War II era also witnessed a growth in anti-union sentiment, though not quite as sharp as that following the first World War. The passage of the Smith-Connally (War Labor Disputes) Act of 1943 over President Roosevelt's veto, and of the Taft-Hartley Act four years later over President Truman's veto, both of which introduced new government controls over unions, was evidence of a changed public opinion. In spite of the post-war economic expansion, union membership stagnated—partly because of the shift toward white-collar employment that the unions found difficult to cope with. Jurisdictional disputes between unions affiliated to the two competitive federations led to a waste of union resources.

Merger of the AFL and CIO

On the other hand, the departure of the union leaders who had initiated and continued the AFL-CIO split removed some of the difficulties in the way of merging the two organizations. The expulsion of the Communist-dominated unions by the CIO in 1949 and the withdrawal of the United Mine Workers from the AFL also facilitated unification. In 1955 the AFL and CIO merged under the combined leadership of George Meany and Walter Reuther. The new organization (AFL-CIO) represented some 15,000,000 members at its birth. Not included in the new federation were the mine workers, some railroad unions, and various smaller organizations expelled for Communist domination; others were expelled later for corrupt practices—including the International Brotherhood of Teamsters with its more than 1,500,000 members.

The AFL and CIO had been instrumental, even before the merger, in setting up the International Confederation of Free Trade Unions (ICFTU) to combat the Communist-dominated World Federation of Trade Unions (WFTU). However, issues of international policy soon caused sharp internal frictions between Meany and Reuther. While Meany criticized the ICFTU for failing to combat communism with enough vigour, Reuther criticized Meany for complacency, for excessive concentration on a negative anti-Communist policy out of line with the thinking of most European unions, and for lack of imaginative leadership for social change. Personal rivalries intensified the conflict. In 1968 Reuther's United Automobile Workers left the AFL-CIO and combined with the Teamsters and two smaller unions and formed the Alliance for Labor Action (ALA) to engage in new organizing drives and to operate on a wider range of issues. The ALA had barely established itself when it lost its main driving force with the death of Walter Reuther in an airplane accident in 1970.

Unions of public employees. By means of an executive order issued in 1962 the Kennedy administration gave federal public employees the right to organize and bargain collectively, though not to strike. That action, together with some legislation at the state level, led to a rapid expansion of unionism among public employees. The problem of how bargaining could be conducted effectively without the ultimate sanction of the strike remained to be solved.

A number of professional associations, though refusing to regard themselves as unions, began in the late 1960s to engage in collective bargaining and even, occasionally, to go out on strikes. The outstanding example was the influential National Education Association, several of whose affiliates carried out successful strikes. A whole class of quasi unions was thus developing outside the traditional labour movement.

WESTERN EUROPE

Unionism on the Continent shares some features with unionism in Britain: early persecution by government and courts, reliance upon government intervention and mediation in disputes; a tendency for unions to carry on collective bargaining with employers' associations rather than with individual companies; and a general acceptance of the terms of agreements as a floor upon which higher effective earnings are achieved by various devices. A distinguishing characteristic of European trade unionism has been, and to a large extent still is, its tendency toward ideologically oriented unions, most of which were established by political parties or religious organizations in the 19th century. After World War II a trend in the direction of unified trade unions, combining different ideological currents under one roof, manifested itself and has achieved considerable success.

Tendency toward ideological unions

Germany. Trade unionism began to gather momentum in Germany only after the fall of Otto von Bismarck in 1890. Under Socialist inspiration the bulk of the unions formed a national federation in that year. The Roman Catholic Centre Party followed in 1894 with the establishment of a much smaller federation of Christian unions. Nationalist- and, for a brief period, Communist-led unions, as well as separate white-collar unions, came into being at various times before World War I. There were some successes in collective bargaining, but unions remained fairly insignificant during this period.

After World War I the Socialist-inspired unions organized in three main federations: the General Confederation of German Trade Unions (Allgemeiner Deutscher Gewerkschaftsbund; ADGB) for the manual workers and two confederations for white-collar employees and civil servants. Collective agreements for an industry or trade set minimum wages, which in turn were supplemented by the efforts of workers' councils in the plants. The council members were elected by all workers and white-collar employees, whether union members or not. Legally, and often in fact, councils were independent of the unions. Unionism flowered during the short period of prosperity in the 1920s but was destroyed by the Nazi takeover in 1933. The Labour Front, established and totally controlled by the National Socialist (Nazi) Party, was not a union by any definition of the term. Immediately after World War II a powerful trade union movement came into being in West Germany. It was based on a merger of 16 Socialist and Christian industrial unions in a unified German Federation of Trade Unions (Deutscher Gewerkschaftsbund; DGB). Its membership, whose turnover is high, has leveled off at 6,500,000 workers in 1970 after a rapid rise. In addition, there are union federations for white-collar workers and civil servants. A small separate Roman Catholic federation has had little success.

Union delegates on management boards

Under the slogan of "codetermination" the unions have acquired direct representation in management. Legislation enacted in 1951 provides that in the coal and steel industries stockholders shall elect five members of an 11-member supervisory board of the company, and workers another five. Two of the workers' representatives come from the enterprise, a third is appointed by the union, and two more by the union federation, usually the DGB. The 11th member is jointly elected by the 10 others. In addition, one of the three members of the *Vorstand* (management), which actually runs the enterprise, must be

selected with the approval of the workers' representatives; that member is in charge of labour relations. In other large firms, outside the coal and steel industries, labour holds one-third of the seats on the supervisory boards. The DGB has been campaigning for an extension of the coal and steel system to all large firms. A committee of social scientists appointed by the government to study the question in a 1970 report supported the DGB demand, though with some significant modification.

France. Ever since the merger in 1895 between syndicalist and Marxist trade unions, the French labour movement has suffered persistently from internal cleavages and organizational divisions. The main union federation, the General Confederation of Labour (Confédération Générale du Travail; CGT), was set up in that year. It represented a coalition of anarchists and syndicalists, who put their trust in a revolutionary general strike to destroy capitalism and the state, together with a group of Marxists, who aimed at obtaining control of the state by a revolution. The syndicalists organized local union federations, mostly of an industrial nature.

Syndicalism broke down in the World War I wave of patriotism that engulfed French workers; after 1918 the Communists inherited a large part of the syndicalist following. For most of the time ever since, the French labour movement has been divided into three main (and several smaller) ideologically inspired organizations: the CGT under Communist control, the mostly Socialist Workers' Strength (Force Ouvrière; FO), and the formerly Christian French Democratic Confederation of Labour (Confédération Française Démocratique du Travail; CFDT), which has been trying to overcome its religious origins and the limits those origins set to its expansion among the main body of French workers.

Although most French workers are not organized, and those who are union members pay their dues only irregularly, the elections of workers' councils and shop stewards (*délégués du personnel*) roughly indicate worker allegiances. It would thus appear that a small majority supports the CGT and somewhat fewer the CFDT and FO. White-collar workers are organized in the General Confederation of Executive Staff (Confédération Générale des Cadres; CGC). Other unions remain outside all confederations.

Collective bargaining: a late and minor factor

In view of the fragmentation in the French labour movement, it is not surprising that collective bargaining developed late—after 1936. It is still less significant than labour legislation and administrative action in determining wages and working conditions. Workers' councils, to provide worker participation in management, and the shop steward system, to handle grievances in the plant, were both established by law. By the early 1970s, however, dwindling union strength had weakened both institutions to the point that they were not functioning in large numbers of enterprises. The great strike movement of the late 1960s, which was organized more for political reasons than as a collective bargaining tactic, developed outside the unions. Later on, however, the unions finally succeeded in controlling it.

Italy. In Italy, much the same as in France, the unions have remained weakened and divided by ideologies and religion. The late economic and political development of Italy directed labour's attention toward radical and political solutions. Soon after the resurgence of unions that followed the downfall of Mussolini's Fascist regime, the old tendency toward multiple and changing divisions within the movement reasserted itself.

Three major confederations emerged after World War II: the Italian General Confederation of Labour (Confederazione Generale Italiana del Lavoro; CGIL), led by a coalition of Communists and left-wing Socialists; the Italian Confederation of Workers' Unions (Confederazione Italiana Sindacati Lavoratori; CISL), representing the Roman Catholic labour movement but with some Socialists as well; and the Italian Labour Union (Unione Italiana del Lavoro; UIL) under Social Democratic leadership. In the wide-scale strikes of 1969–70 the unions persuaded far more than their members to join the so-called rolling strikes—short walkouts, one after the other, in different industries.

When labour agreements are reached, they usually are national in scope, and some cover benefits and conditions for all industries. In 1962 the unions obtained the additional right to conclude supplementary agreements at the plant level.

Scandinavia. The consolidation of local unions into national organizations paralleled the advance of industrialization in Norway, Sweden, and Denmark during the second half of the 19th century. That trend toward centralized unionism received support at the Scandinavian Labour Congress of 1897; and, by 1900, all three countries had small but firmly rooted federations of national trade unions. Employers' associations followed similar development from a local to national scale.

Agreement on basic principles

By the late 1930s the federations of workers and employers in the three countries had agreed on certain basic principles in the conduct of labour relations. Agreements between the labour and management federations, besides placing limitations on sympathetic lockouts and strikes, provide for orderly negotiations by upholding the right to organize for collective bargaining and by laying down procedures for discussion. The broad measure of understanding between labour and management and their common desire to keep government intervention at a minimum has facilitated responsible collective bargaining. Strikes and lockouts have not been entirely eliminated, but their occurrence has become less frequent, although a reversal of this trend manifested itself beginning in 1968. Full collective bargaining rights, including the right to strike, have been extended in Sweden to government employees of all kinds.

In Scandinavia collective bargaining is highly centralized in comparison with the United States or Great Britain. The influence of the federations of labour in conducting negotiations or determining national policy far surpasses that exerted by the AFL-CIO or the British TUC. The federations have the power to ensure that their basic decisions are implemented in the industry agreements. They have acquired an important role in political affairs through their close alliance with the Labour and Socialist parties. In Sweden the federation has pursued a policy of establishing skill and industry differentials—the so-called solidarity policy—but effective wages have often departed from contract rates, thus counteracting the solidarity policy.

Norwegian and Swedish unions are organized chiefly on an industrial basis, whereas the Danish unions—comparable to the British in this respect—are both general and industrial. Although it is not compulsory, union membership is widespread, even among white-collar workers.

COMMUNIST NATIONS

Soviet Union. The Russian labour movement owes its origin to the strikes of the 1890s, in which revolutionary students and workers cooperated against the tsarist regime and capitalist business owners. Hampered by a hostile government, the trade union movement slowly developed on the general pattern of the German labour organizations. With the Russian Revolution in 1917, however, the entire basis of the developing labour movement changed. Under Lenin's strategy the unions became one of the links connecting the Communist Party—a small, close-knit elite of professional revolutionaries—with the large mass of industrial workers on whose behalf the party would operate.

A key question arose after the Bolshevik takeover. What role would unions perform in a socialist system in which capitalism had ceased to operate? A long and involved debate followed. The unions themselves endeavoured to fill the traditional role of trade unions: to maintain or improve the living standards of their members. Leon Trotsky argued, however, that the unions should be transformed into "production agents" so that they would serve the workers' interests by fostering increased productivity of labour, promoting economic reconstruction, and providing increased quantities of consumer goods. Unions were thus to be instruments of economic administration, with only a subsidiary role in setting wages and working conditions. Another more radical role was proposed by the Workers' Opposition, which urged that the trade unions should dominate the state and manage the economy.

The answer to the question of what role unions were to play in the new Communist state was, in fact, dictated by an overall economic policy that called for the commitment of all possible resources to industrialization. The destruction of much of the productive equipment of the country in World War I, the civil war, and the ensuing foreign intervention made it impossible for the shattered economy to satisfy even the most elementary needs of the population. Moreover, the failure of Germany to abolish capitalism, as the Bolsheviks had hoped, required a complete revision of Bolshevik strategy. In the absence of victorious revolutions in the industrially advanced nations, the "dictatorship of the proletariat" in the Soviet Union was doomed unless that country could be rapidly industrialized. Thus, the growth of a working class to build an industrial base in the Soviet Union became the main task of both party and unions.

Such a program, however, required that real wages be kept to a minimum so that all possible resources could be devoted to capital formation. Consequently, unions could not be permitted to assist workers in raising their living standards. Instead, they became instruments of the state to increase output and speed industrialization. That production-oriented role gave rise to a trade union structure appropriate to the machinery of Soviet economic administration and adaptable to its changes.

Unions as instruments of the state

Soviet trade unions are strictly industrial in nature. All persons employed in any one factory, state farm, or other undertaking belong to the same union; and that same union comprises all of the employees of similar factories, farms, or other enterprises everywhere else in the Soviet Union. Horizontal coordination occurs at the republic, regional, and local levels; but just as the economic administration has been closely centralized, so the national organs of the trade unions exercise effective authority over all lower levels of the union. More than 90 percent of industrial workers belong to unions, mainly because of the considerable advantages, such as social insurance benefits, available to members.

Soviet trade unions do not limit their activities to industry and plant matters alone. Their other functions include legislative initiative in labour matters; participation in national economic planning and wages policy; the administration of social insurance; the organization of cultural and recreational facilities; and cooperation in the education and vocational training of workers. Moves in the 1960s toward decentralization in economic planning and administration, the use of incentives, and the assignment of greater authority to local managers have given rise to an increase in union influence at the lower levels. Among the unions' new roles are greater participation in the settling of grievances and in application of the wage system at the plant level.

Some observers have discerned a tendency for the Soviet industrial relations system to evolve in the direction of those of other advanced industrial nations. Nevertheless, like all other organizations in the Soviet Union, the trade unions remain in all essentials under the complete control of the Communist Party.

Poland. The Communist view of unions as organs of the state was dramatized in Poland, where an independent union, Solidarity (Solidarność), arose in 1980 amid serious economic difficulties, was recognized by the government as a matter of tactical accommodation, and was then suppressed a little more than a year later under a decree of martial law.

Yugoslavia. After its break with the Soviet Union in 1948, Yugoslavia commenced a program of economic reforms combining the profit motive, public ownership of most means of production, and a gradual decentralization of economic decision making.

The system of centrally directed state-managed enterprises has been progressively replaced by a workers' self-management system under which relatively free economic decisions are made. The concept of workers' self-management, recognizing the system's potential conflict of interest between management and workers, stresses the role of workers as coproducers in the enterprise. Management is exercised by a director, who is under the supervision of a workers' council elected by the employees of the enterprise.

Co-producer role for workers

To the extent that workers' councils actually represent the workers' interests, they have superseded the role of the trade union within the enterprise. Union functions have consequently been reduced to such activities as assisting individual workers in personal matters, obtaining better training and education for workers, and exercising a modest participatory role in various enterprise deliberations.

The unions appear to be attempting to offset the decline in their traditional functions within an enterprise by enhancing their influence at provincial and national levels. There, the trade unions endeavour to participate in decisions on social and economic issues within the framework set by the political system.

JAPAN

Japanese labour unions are mainly the product of the post-World War II period. With the support of the U.S. occupation forces, the Japanese government adopted a drastic labour-protection policy immediately after the war. Its constitution was amended in 1946 so that workers' rights "to organize and to bargain and act collectively" were guaranteed. The phenomenal postwar economic growth of Japan has resulted in a great increase in employment.

Unions enterprise-oriented. Union structure in Japan differs sharply from the Western pattern. The enterprise—not a craft or industry—provides the base of union organization. One of the main features of enterprise unionism is "lifetime" employment. Employers do not, as a rule, lay off workers either permanently or temporarily; on the other hand, workers are not expected to move from one company to another. Typically, the enterprise recruits and trains young workers immediately after their graduation from school and continues to employ them until they reach retirement age, usually 55. Educational background and length of service in a given company are the key factors in wage determination. Under such a system, collective bargaining naturally focuses on particular enterprises or companies.

Lifetime jobs for workers

National confederations. Most of the enterprise unions are combined in national unions or in federations that are affiliated with national confederations. Major confederations are General Council of Trade Unions of Japan (Sōhyō); Japanese Confederation of Labour (Dōmei); and Liaison Council of Neutral Labour Unions (Chūritsurōren). The national confederations participate very little in bargaining with employers; their major function is the guidance and coordination of their affiliated enterprise unions. Each spring the national confederations form a special committee for Shunto (the spring offensive). The committee plans the amount of the wage increase to be demanded, the period of negotiations, and the dates of industrywide strikes.

National unions and confederations play an even more important role in politics by allying themselves with one of the parties—Socialist (Nippon Shakaitō), Democratic Socialist (Minshatō), and Communist (Kyosanto).

The new mobility. Some changes in the system seem to be resulting from Japan's high economic growth rates. Because the labour market, especially for young workers, has become extremely tight, and because labour mobility has increased, the lifetime employment system has begun to be eroded. Rapid technological changes have made obsolete the traditional system of basing wages on length of service. Young workers with a modern school education are more suitable for many of the new jobs. Partly because of the high rate of economic growth and partly because of the Shunto bargaining targets, the unions have been successful in obtaining high wage increases. Thus the main emphasis of the Japanese labour movement has begun to shift from politics to economics.

AUSTRALIA AND NEW ZEALAND

Early settlers, who had brought the ideal of unionism along with them from England, established the first workingmen's organizations in Australia and New Zealand in the 1830s in the form of small trade societies. The foundations of modern unionism in Australia were not laid until

two or three decades later, however, when urban craft unions for skilled workers began to emerge in the 1850s and 1860s. During that period and until 1890, "mass" or "industrial" unions were formed, first by the coal miners and subsequently by others, including gold miners, seamen, and wool shearers. Economic growth and a labour shortage favoured union expansion. With suffrage already a reality, the union movement devoted its attention to such essentially economic issues as the eight-hour day.

Miners among the first "mass" unions

Two distinctive features of present-day trade unionism, namely, political participation and compulsory arbitration, can be traced back to the serious strike setbacks and the economic depression of the 1890s. The Australian Labor Party, created and largely financed by the unions, has been a significant force on the Australian political scene since it first gained office in 1910. The Labour Party of New Zealand also derives much of its strength from trade union support.

Compulsory arbitration. Before 1890 unions had opposed the idea of compulsory arbitration. After losing a number of major strikes in which union recognition had been the main issue, the unions adopted a more favourable attitude toward arbitration, which would force employers to recognize them. Despite the retention of state arbitration systems in some form since passage of the first Commonwealth Conciliation and Arbitration Act of 1904, the focus of attention and activity has gradually shifted to the federal sphere. By the 1970s, industrial relations were dominated by compulsory arbitration, yet collective bargaining continued, generally under close union control. Nor was strike action precluded by compulsory arbitration. Indeed, by international standards, the frequency of strikes in Australia has been relatively high, although with a short average duration.

In pre-Commonwealth days the need for close coordination for the successful pursuit of labour's objectives was expressed in successive Intercolonial Trade Union congresses, beginning in 1879. Their direct descendant is the Australian Council of Trade Unions (ACTU), founded in 1927.

Later developments in New Zealand. Trade unions developed in Australia and New Zealand along broadly similar lines. In New Zealand, however, 1936 legislation provided for compulsory unionism. This took the form of providing that every employer bound by an award of the Court of Arbitration, or by any industrial agreement, could not employ a worker subject to the award or agreement unless that worker was a member of the union also bound by the award or agreement. This resulted in a considerable expansion and strengthening of the union movement. In 1961 the 1936 legislation was amended. Compulsory membership was made subject to agreement between employer and union or to a demand by more than 50 percent of the employees in an industry for compulsory unionism. Provision was made also to allow any worker conscientiously objecting to union membership to apply for a certificate of exemption. In the absence of a compulsory membership agreement, provision was made for the preferential hiring of unionists over nonunionists.

Development of trade unions in emerging nations

THE INDIAN SUBCONTINENT

India. According to a 1961 census, India's wage earners totaled 24,000,000, or 13 percent of its labour force, most of which was engaged in subsistence agriculture. Of the 24,000,000 wage earners, about 4,000,000 (17 percent) were estimated to be organized in unions.

The weakness of Indian trade unionism is related to the country's slow industrialization. Historically, Indian unions were closely linked to the independence movement. The Indian National Congress Party took the initiative for the formation in 1920 of the first national body of trade unions, the All-India Trade Union Congress (AITUC), but this soon split into several factions.

The history of the trade union movement in India is one of conflicts and splits resulting mainly from ideological rivalries. Interunion rivalry has hindered its development. Collective bargaining is ineffective in India because the unions are not strong enough to force settlements; union rivalry makes the situation worse. As a result, unions tend to turn to the government for intervention. The Industrial Disputes Act of 1947 empowered the central or provincial governments to refer an industrial dispute for adjudication. In addition, wage boards have been set up for the determination of wages.

Pakistan. Although the nonagricultural sector of the Pakistan economy has expanded considerably since independence was achieved in 1947, some 20 years later (prior to separation from Bangladesh) approximately 68 percent of the total labour force of 40,000,000 was still engaged in agriculture.

The All-Pakistan Confederation of Labour, established in 1950, has been recognized by the government as the sole representative of the workers in Pakistan. Many other trade union federations, however, have sprung up. Just as its Indian counterpart, the Pakistan labour movement is divided, and inter-union rivalry is a major handicap to its development. Collective bargaining is extremely difficult; when agreements are concluded they usually result from voluntary concessions made by progressive employers. Unilateral employer's decisions and government intervention play a dominant role in industrial relations.

LATIN AMERICA

The highest rate of population increase of any area in the world not only has put a heavy strain on the educational systems of most Latin-American countries but also has produced excess supplies of common labour. As a result, advances in wages and working conditions have been exceedingly difficult. Accordingly, one of the main objectives of Latin-American trade unions has been to protect and insulate the employed labour force from the consequences of an excess supply of unemployed labour. Various job-protection devices have been developed, mainly based upon a system of severance pay proportionate to the length of an employee's service. Such a system makes dismissal expensive for the employer.

Ideological orientations. Labour movements in Latin America consistently have taken on an ideological colouring. In the early 1970s there were three main groupings of unions: those affiliated with the mildly progressive Inter-american Regional Organization of Labour (Organización Regional Interamericana de Trabajadores; ORIT), which in turn belongs to the International Confederation of Free Trade Unions (see below); the Latin American Confederation of Workers (Central Latinoamericana de Trabajadores; CLAT), a radical Christian movement strongly opposed to what is described as the American dominance of ORIT; and, finally, the largely underground organizations, dominated by communist parties or communist sympathizers.

Mexico. Among the Latin-American labour movements Mexico's is the most successful, partly as a result of the high growth rates of the Mexican economy during and after World War II. Although collective bargaining has been practiced for many years and is fairly well developed, the trade unions have succeeded in welding a close relationship to Mexico's governing party; the unions, in fact, constitute one of the three main divisions of the party organization. Thus, the success of the Mexican labour movement since about 1920 has depended mainly on relationship to the government party.

Decisive role of labour courts

Elaborate labour legislation based on Article 123 of the Mexican Constitution gives labour courts powerful means to intervene in disputes. They may declare a strike lawful or unlawful, and by their decision the outcome of a strike is more or less settled. Unlawful strikes have few chances of success, but a lawful strike is almost certain to bring advances to the union.

AFRICA

During most of the colonial period in sub-Saharan Africa, unionism, when it existed at all, was limited to white workers. Attempts of black workers to form unions—usually for the purpose of obtaining wages and working conditions similar to those of white workers—were frequently

treated as rebellious political movements and suppressed. Black workers were thus compelled to engage in strikes and demonstrations without having a stable organization at their disposal. Violence often accompanied these short-lived movements.

In the late 1930s and especially during and after World War II, colonial governments altered their policies and introduced trade unionism in their territories. Usually these unions were patterned after the model of the mother country, and in many cases they were simply branches of metropolitan unions. In some areas—as for instance Guinea and Kenya—close cooperation developed in due course between the unions and the national liberation movements. Political leaders were also often union leaders or closely associated with them. In other parts of Africa—in Rhodesia, for example—the national liberation movements ignored the unions, the membership of which was rather small and which were therefore not in a position to offer financial or mass support to political movements.

After independence was achieved, unions consistently abandoned the patterns inherited from the former metropolitan territories. Western-style unionism was clearly inappropriate in sub-Saharan Africa with its low level of literacy, a labour force with a large migratory component, few educated working-class leaders, low living standards, and a poorly developed labour market. The unions, however, were rarely given an opportunity to develop new autonomous forms of organization and action. Almost everywhere the government party, by a combination of rewards and pressures, took over the unions under the slogan of "African socialism" and turned them into instruments to enhance production and restrict the wage demands of their members.

Only sporadically, mostly in North Africa, did autonomous unions survive. In most cases they formed the nucleus of the opposition to the government, as is the case in Morocco.

International organizations

The large trade union movements of various countries for many years have maintained loose alliances by joining international organizations of labour; federations of unions, rather than individual unions, usually hold membership. In 1901 the International Federation of Trade Unions was established, chiefly under the guidance of German unions. It proved to be ineffective and disappeared during World War I. In 1919 it was revived at Amsterdam, but immediately came into collision with the Red International of Labour Unions, established by the new government of the Soviet Union. The Communist organization had a brief period of expansion but soon dwindled away and had disappeared before 1939.

Ideological cleavage of the international labour movement

Unity and division. At a 1945 conference in London a new organization came into being, the World Federation of Trade Unions (WFTU). Its principal organizers were the British Trades Union Congress, the U.S. Congress of Industrial Organizations, and the Soviet Russian Labour Federation. It made a vigorous but unsuccessful attempt to reconcile the differences between the communist and noncommunist factions. The U.S. announcement of the Marshall Plan for economic aid to Europe, in which the Soviet Union and other communist states declined to participate, provided the occasion for the withdrawal of the noncommunist faction.

In 1949 in London the noncommunist groups created the International Confederation of Free Trade Unions (ICFTU). At the first session 70 labour organizations of 53 nations, claiming to represent 50,000,000 workers, became "founding members." The ICFTU was readily accepted by nearly all of the noncommunist-led labour organizations of the world; and after 10 years it had grown to include 138 organizations located in 103 different countries, representing more than 57,000,000 workers.

The announced purposes of the ICFTU include striving for world peace, spreading democratic institutions, increasing the standard of living for workers everywhere, worldwide strengthening of free trade unions, and support to colonial people in their struggle for freedom. The ICFTU consistently opposed fascist as well as communist dictatorships and implemented that policy by giving such aid as was possible to free labour in Spain and certain Latin-American countries. It also furnished direct financial assistance to workers in Hungary and Tibet and campaigned against racialist policies in South Africa.

Failures and successes of the ICFTU. Lack of homogeneity among affiliates hindered the activity of the ICFTU in many fields, chiefly because of differences among its affiliates in the approach to unions in communist-controlled countries. It found its work to be most effective in the area of international education. By 1960 it had created an International Solidarity Fund of $2,000,000 to aid workers who became victims of oppression and to promote democratic trade unionism in economically developing countries. Problems of union organization were discussed at ICFTU seminars in various parts of the world, with experienced labour leaders and labour spokesmen from the less industrialized countries participating.

Worldwide program of training industrial workers

To facilitate the functioning of its widespread activities, the ICFTU established headquarters in Brussels with regional or subregional offices in many other countries. From one or more of those centres it conducted numerous educational conferences, maintained a residential trade union training college in Calcutta, and assisted in founding an African Labour College in Kampala, Uganda. It provided assistance to inexperienced workers in areas in the first stages of industrialization and sent organizers to Lebanon, Japan, Cyprus, Cameroon, India, Indonesia, Nigeria, and elsewhere.

It has been the consistent policy of the ICFTU to cooperate with the United Nations Educational, Scientific, and Cultural Organization and with the International Labour Office in Geneva. It is wholly financed by contributions from its affiliates.

Communist and Roman Catholic federations. The WFTU, meanwhile, has continued to be completely dominated by the Communists. It claimed to represent more than 90,000,000 workers by 1960, more than one-half in the Soviet Union and a large part of the remainder in Soviet satellite nations. It also has substantial affiliates in France and Italy. The WFTU has adhered strictly to the Communist Party line and has not cooperated with other organizations, serving more as a political instrument than as an economically oriented federation of unions.

A much smaller international organization, the International Federation of Christian Trade Unions (IFCTU), now called the World Confederation of Labour (WCL), is made up largely of Roman Catholic labour unions in France, Italy, and Latin America. The ICFTU, at its founding congress in 1949, invited the affiliates of the IFCTU to join, but the invitation was rejected. On the international scene the WCL has been a comparatively ineffective organization, its influence limited to a few countries in Europe and Latin America. (A.F.St.)

BIBLIOGRAPHY

Organization of work: ÉMILE DURKHEIM, *Émile Durkheim on the Division of Labor in Society,* trans. by W.D. HALLS (1984; originally published in French, 1893), started serious consideration of the organization of work in society. MELVIN KRANZBERG and JOSEPH GIES, *By the Sweat of Thy Brow: Work in the Western World* (1975, reprinted 1986), is a later popular survey. Organization of work in the context of economic history of various countries is dealt with in chapters of the multivolume *Cambridge Economic History of Europe* under the general editorship of M.M. POSTAN.

For the organization of work from prehistoric to classical times, see ROBERT J. BRAIDWOOD, *Prehistoric Men,* 8th ed. (1975); and ROBERT F.G. SPIER, *From the Hand of Man: Primitive and Preindustrial Technologies* (1970). The evolution of the organization of work in ancient civilizations is treated in KARL A. WITTFOGEL, *Oriental Despotism: A Comparative Study of Total Power* (1957, reprinted 1981); AHMED FAKHRY, *The Pyramids,* 2nd ed. (1969, reprinted 1974); ROBERT MCCORMICK ADAMS, *The Evolution of Urban Society: Early Mesopotamia and Prehispanic Mexico* (1966); CARL ROEBUCK (ed.), *The Muses at Work: Arts, Crafts, and Professions in Ancient Greece and Rome* (1969); and J.G. LANDELS, *Engineering in the Ancient World* (1978, reprinted 1981). The story of Middle Eastern developments is continued in HANS E. WULFF, *The Traditional Crafts of Persia: Their Development, Technology, and Influence*

on Eastern and Western Civilizations (1966); and AHMED Y. AL-HASSAN and DONALD R. HILL, *Islamic Technology: An Illustrated History* (1986); while the history of Chinese work is brought out in various segments of JOSEPH NEEDHAM, *Science and Civilization in China* (1954–), an ongoing work of which six multipart volumes had been published by 1988.

Medieval and early modern developments in the Western world are treated in JEAN GIMPEL, *The Medieval Machine: The Industrial Revolution of the Middle Ages,* 2nd ed. (1988; originally published in French, 1975); and, most fully, in FERNAND BRAUDEL, *Civilization and Capitalism, 15th–18th Century,* 3 vol. (1981–84; originally published in French, 1979). Changes in the organization of work with the development of the factory system in the early Industrial Revolution, especially in the textile industries, are covered in several standard works: HERBERT HEATON, *The Yorkshire Woollen and Worsted Industries: From the Earliest Times Up to the Industrial Revolution,* 2nd ed. (1965); J.L. HAMMOND and BARBARA HAMMOND, *The Rise of Modern Industry* (1925, reissued 1969); and PAUL MANTOUX, *The Industrial Revolution in the Eighteenth Century: An Outline of the Beginnings of the Modern Factory System in England,* rev. ed. (1961, reissued with a new foreword, 1983; originally published in French, 1905). The Industrial Revolution itself is the subject of many studies, the most notable in relation to the organization of work being E.P. THOMPSON, *The Making of the English Working Class,* new ed. (1968, reprinted 1980); JENNIFER TANN, *The Development of the Factory* (1970); WILLIAM H. SEWELL, JR., *Work and Revolution in France: The Language of Labor from the Old Regime to 1848* (1980); DAVID A. HOUNSHELL, *From the American System to Mass Production, 1800–1932: The Development of Manufacturing Technology in the United States* (1984); NEIL J. SMELSER, *Social Change in the Industrial Revolution: An Application of Theory to the British Cotton Industry* (1959); and GEORGES FRIEDMANN, *The Anatomy of Work: Labor, Leisure, and the Implications of Automation* (1962, reprinted 1978; originally published in French, 1956).

For scientific management see FREDERICK WINSLOW TAYLOR, *The Principles of Scientific Management* (1911, reprinted 1985); DANIEL NELSON, *Frederick W. Taylor and the Rise of Scientific Management* (1980); HUGH G.J. AITKEN, *Taylorism at Watertown Arsenal: Scientific Management in Action, 1908–1915* (1960, reprinted as *Scientific Management in Action,* 1985); and SAMUEL HABER, *Efficiency and Uplift: Scientific Management in the Progressive Era, 1890–1920* (1964, reprinted 1973). ELTON MAYO, *The Human Problems of an Industrial Civilization,* 2nd ed. (1946), and *The Social Problems of an Industrial Civilization* (1945, reprinted 1977), are two major summations by a pioneer industrial sociologist. Other major accounts of the modern factory system before the advent of automation are W. LLOYD WARNER and J.O. LOW, *The Social System of the Modern Factory: The Strike: A Social Analysis* (1947, reprinted 1976); and CHARLES R. WALKER and ROBERT H. GUEST, *The Man on the Assembly Line* (1952, reprinted 1979).

The development of automation and its effect on the organization of work are treated in WALTER BUCKINGHAM, *Automation: Its Impact on Business and People* (1961, reprinted 1982); JAMES R. BRIGHT, *Automation and Management* (1958, reprinted 1970); DAVID F. NOBLE, *Forces of Production: A Social History of Industrial Automation* (1984); MARVIN MINSKY (ed.), *Robotics* (1985); HARLEY SHAIKEN, *Work Transformed: Automation and Labor in the Computer Age* (1985); DANIEL B. CORNFIELD (ed.), *Workers, Managers, and Technological Change: Emerging Patterns of Labor Relations* (1987); ELI GINZBERG, THIERRY J. NOYELLE, and THOMAS M. STANBACK, JR., *Technology and Employment: Concepts and Clarifications* (1986); TOM FORESTER (ed.), *The Microelectronics Revolution: The Complete Guide to the New Technology and Its Impact on Society* (1980); E. FOSSUM (ed.), *Computerization of Working Life,* trans. from Norwegian (1983); LAURI PERMAN (ed.), *Work in Modern Society: A Sociology Reader* (1986); BRUCE R. GUILE (ed.), *Information Technologies and Social Transformation* (1985); CHARLES EDQUIST and STAFFAN JACOBSSON, *Flexible Automation: The Global Diffusion of New Technology in the Engineering Industry* (1988); and SHOSHANA ZUBOFF, *In the Age of the Smart Machine: The Future of Work and Power* (1988). The effects of automation on white-collar clerical tasks are treated in IDA R. HOOS, *Automation in the Office* (1961); and J. DAVID ROESSNER *et al., The Impact of Office Automation on Clerical Employment, 1985–2000: Forecasting Techniques and Plausible Futures in Banking and Insurance* (1985).

Most notable among the growing number of studies of the working role of women are BARBARA A. HANAWALT (ed.), *Women and Work in Preindustrial Europe* (1986); LINDSEY CHARLES and LORNA DUFFIN, *Women and Work in Pre-industrial England* (1985); ROSALYN BAXANDALL, LINDA GORDON, and SUSAN REVERBY (eds.), *America's Working Women* (1976); ANN OAKLEY, *Housewife* (1974; U.S. title, *Woman's Work: The Housewife, Past and Present,* 1976); ALICE KESSLER-HARRIS, *Out to Work: A History of Wage-Earning Women in the United States* (1982); and RUTH SCHWARTZ COWAN, *More Work for Mother: The Ironies of Household Technology from the Open Hearth to the Microwave* (1983).

The role of migrant workers is treated in JAN LUCASSEN, *Migrant Labour in Europe, 1600–1900: The Drift to the North Sea* (1987; originally published in Dutch, 1984); JONATHAN POWER, *Migrant Workers in Western Europe and the United States* (1979); HANS CHRISTIAN BUECHLER and JUDITH-MARIA BUECHLER (eds.), *Migrants in Europe: The Role of Family, Labor, and Politics* (1987); ROBIN COHEN, *The New Helots: Migrants in the International Division of Labour* (1987); THOMAS R. BAILEY, *Immigrant and Native Workers: Contrasts and Competition* (1987); and PHILIP L. MARTIN, *Harvest of Confusion: Migrant Workers in U.S. Agriculture* (1988). (M.Kr.)

Industrial and organizational relations: Important works in the field of industrial relations include SIDNEY WEBB and BEATRICE WEBB, *Industrial Democracy,* 2 vol. (1897, reissued in 1 vol., 1965), a classic theory in the British socialist tradition; JOHN R. COMMONS, *Institutional Economics: Its Place in Political Economy* (1934, reissued 1961), a summary of the theoretical ideas of the father of American industrial relations research; JOHN T. DUNLOP, *Industrial Relations Systems* (1958, reprinted 1977), a theoretical model that shaped the study of industrial relations between 1960 and 1980; JACK BARBASH, *The Elements of Industrial Relations* (1984), a modern statement on the nature of the employment relationship; RICHARD HYMAN, *Industrial Relations: A Marxist Introduction* (1975), an interpretation by a leading contemporary Marxist; THOMAS A. KOCHAN, HARRY C. KATZ, and ROBERT B. MCKERSIE, *The Transformation of American Industrial Relations* (1986), a study offering a strategic-choice model for interpretation of industrial relations; and RICHARD E. WALTON and ROBERT B. MCKERSIE, *A Behavioral Theory of Labor Negotiations: An Analysis of a Social Interaction System* (1965), a classic theory of the dynamics of negotiations. TAISHIRO SHIRAI (ed.), *Contemporary Industrial Relations in Japan* (1983), is a survey of the evolution of industrial relations in the country from World War II to the end of the 1970s; and WOLFGANG STREEK, *Industrial Relations in West Germany: A Case Study of the Car Industry* (1984), is an in-depth look at the German industrial system.

Leading works in the area of organizational behaviour include CHESTER I. BARNARD, *The Functions of the Executive* (1938, reissued 1979), a classic statement on the nature of organizational relations by a prominent executive of his time; MICHEL CROZIER, *The Bureaucratic Phenomenon* (1964; originally published in French, 1964), an insightful study of the organizational behaviour in France; JAMES G. MARCH and HERBERT A. SIMON, *Organizations* (1958), a theoretical analysis based on a decision-making approach; DOUGLAS MCGREGOR, *The Human Side of Enterprise* (1960, reprinted 1985), a study of two contrasting sets of assumptions managers make toward employees' motivations and of the effects of these assumptions on managerial behaviour; *Industrial Democracy in Europe* (1981), a comparative empirical study of worker participation conducted by an interdisciplinary team of researchers from 12 European countries; MICHAEL J. PIORE and CHARLES F. SABEL, *The Second Industrial Divide: Possibilities for Prosperity* (1984), a discussion of the changes in work organization and relations in response to changing technologies and market conditions; EDWARD B. ROBERTS (ed.), *Generating Technological Innovation* (1987), a collection of articles on the movement of new products and technologies from the laboratory to the market; and ANDREW M. PETTIGREW, *The Awakening Giant: Continuity and Change in Imperial Chemical Industries* (1985), a detailed study of the process of organizational change based on an example of a large British firm.

For studies of the changing work force, see JEYLAN T. MORTIMER and KATHRYN M. BORMAN (eds.), *Work Experience and Psychological Development Through the Life Span* (1988); GORDON E. O'BRIEN, *Psychology of Work and Unemployment* (1986); CHARLES H. SAVAGE and GEORGE F.F. LOMBARD, *Sons of the Machine: Case Studies of Social Change in the Workplace* (1986); MARC L. SILVER, *Under Construction: Work and Alienation in the Building Trades* (1986); SYLVIA ANN HEWLETT, ALICE S. ILCHMAN, and JOHN J. SWEENEY (eds.), *Family and Work: Bridging the Gap* (1986); JON MILLER, *Pathways in the Workplace: The Effects of Gender and Race on Access to Organizational Resources* (1986); CLARK KERR and PAUL D. STAUDOHAR (eds.), *Industrial Relations in a New Age* (1986); SHEILA B. KAMERMAN and ALFRED J. KAHN, *The Responsive Workplace: Employers and a Changing Labor Force* (1987); MARY ANN VON GLINOW, *The New Professionals: Managing Today's High-Tech Employees* (1988); LOIS FARRER COPPERMAN and FREDERICK D. KEAST, *Adjusting to an Older Work Force* (1983); STEVEN H. SANDELL (ed.), *The Problem Isn't Age: Work and Older Americans* (1987); and HELEN DENNIS (ed.), *Fourteen Steps in Managing an Aging Work Force* (1988). (T.A.K.)

Labour economics: Surveys of the field of labour economics include HENRY PHELPS BROWN, *The Economics of Labor* (1962); MALCOLM R. FISHER, *The Economic Analysis of Labour* (1971); JOHN T. ADDISON and W. STANLEY SIEBERT, *The Market for Labor: An Analytical Treatment* (1979); INGRID H. RIMA, *Labor Markets, Wages, and Employment* (1981); LLOYD G. REYNOLDS, STANLEY H. MASTERS, and COLLETTA H. MOSER, *Labor Economics and Labor Relations,* 9th ed. (1986); DANIEL S. HAMERMESH and ALBERT REES, *The Economics of Work and Pay,* 4th ed. (1988); HOWARD M. WACHTEL, *Labor and the Economy,* 2nd ed. (1988); and CAMPBELL R. MCCONNELL and STANLEY L. BRUE, *Contemporary Labor Economics,* 2nd ed. (1989).

CLARENCE D. LONG, *The Labor Force Under Changing Income and Employment* (1958), outlines the relation of the labour force to the population of a number of countries. Education is treated as a form of investment in GARY S. BECKER, *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education,* 2nd ed. (1975, reprinted 1983). Racial, sexual, and age discrimination in employment is explored in ROBERT C. SMITH, *Equal Employment Opportunity* (1982); E. ROBERT LIVERNASH (ed.), *Comparable Worth: Issues and Alternatives,* 2nd ed. (1984); and HARISH C. JAIN and PETER J. SLOANE, *Equal Employment Issues: Race and Sex Discrimination in the United States, Canada, and Britain* (1981).

For studies of the impact of collective bargaining, see LLOYD G. REYNOLDS and CYNTHIA H. TAFT, *The Evolution of Wage Structure* (1956, reprinted 1970); and JOHN T. DUNLOP (ed.), *The Theory of Wage Determination* (1957, reprinted 1986). Theories of collective bargaining are surveyed in ALAN CODDINGTON, *Theories of the Bargaining Process* (1968); see also RICHARD E. WALTON and ROBERT B. MCKERSIE, *A Behavioral Theory of Labor Negotiations: An Analysis of a Social Interaction System* (1965); HAROLD M. LEVINSON, *Determining Forces in Collective Wage Bargaining* (1966); JOHN G. KILGOUR, *Preventive Labor Relations* (1981); BARRY T. HIRSCH and JOHN T. ADDISON, *The Economic Analysis of Unions: New Approaches and Evidence* (1986); and RICHARD B. FREEMAN and JAMES L. MEDOFF, *What Do Unions Do?* (1984).

For studies of wage structures and their determinants, see E.M. HUGH-JONES (ed.), *Wage-Structure in Theory and Practice* (1966); CLARK KERR, *Labor Markets and Wage Determination* (1977); GUY ROUTH, *Occupation and Pay in Great Britain, 1906–79,* 2nd ed. (1980); HENRY PHELPS BROWN, *The Inequality of Pay* (1977), and *Egalitarianism and the Generation of Inequality* (1988). Theories of differences in the rates of pay for different jobs are developed in BARBARA WOOTTON, *The Social Foundations of Wage Policy: A Study of Contemporary British Wage and Salary Structure,* 2nd ed. (1962); and HAROLD LYDALL, *The Structure of Earnings* (1968). The movement of the general level of pay, in money and real terms, is recorded and analyzed in HENRY PHELPS BROWN and MARGARET H. BROWNE, *A Century of Pay: The Course of Pay and Production in France, Germany, Sweden, the United Kingdom, and the United States of America, 1860–1960* (1968). Work and its compensation are discussed also in ELLIOTT JAQUES, *Free Enterprise, Fair Employment* (1982); JOHN W. WRIGHT, *The American Almanac of Jobs and Salaries,* 3rd ed. (1987); FELICITY SKIDMORE (ed.), *Social Security Financing* (1981); SIMON ROTTENBERG (ed.), *The Economics of Legal Minimum Wages* (1981); and JAMES E. MEADE, *Wage-Fixing* (1982). (E.H.P.B./W.A.Br.)

Trade unionism: The classic works on this subject are: SIDNEY WEBB and BEATRICE WEBB, *History of Trade Unionism,* rev. ed. (1920, reprinted 1975); JOSEPH A. SCHUMPETER, *Capitalism, Socialism, and Democracy,* 6th ed. (1987); CLARK KERR *et al., Industrialism and Industrial Man,* 2nd ed. (1964, reissued 1973); SELIG PERLMAN, *A Theory of the Labor Movement* (1928, reprinted 1979); and JOHN T. DUNLOP, *Industrial Relations Systems* (1958, reprinted 1977). LEWIS L. LORWIN, *The International Labor Movement: History, Policies, Outlook* (1953, reprinted 1973), provides a historical outline. Modern theoretical studies include MICHAEL POOLE, *Theories of Trade Unionism: A Sociology of Industrial Relations* (1981); and SIMEON LARSON and BRUCE NISSEN (eds.), *Theories of the Labor Movement* (1987). JACK JONES and MAX MORRIS, *A–Z of Trade Unionism and Industrial Relations* (1982); and F. JOHN HARPER (ed.), *Trade Unions of the World* (1987), are informative reference sources.

British trade unions are discussed in INTERNATIONAL LABOUR OFFICE, *The Trade Union Situation in the United Kingdom* (1961); ALLAN D. FLANDERS, *Trade Unions,* 7th rev. ed. (1968); ALAN EREIRA, *The People's England* (1981); and DENIS BARNES and EILEEN REID, *Governments and Trade Unions: The British Experience, 1964–79* (1980). Trade union history in other western European countries is discussed in VAL R. LORWIN, *The French Labor Movement* (1954, reprinted 1966); JOHN P. WINDMULLER, *Labor Relations in the Netherlands* (1969); DANIEL L. HOROWITZ, *The Italian Labor Movement* (1963); LEIF LEWIN, *Governing Trade Unions in Sweden* (1980; originally published in Swedish, 1977); RICHARD J. EVANS (ed.), *The German Working Class, 1888–1933: The Politics of Everyday Life* (1982); and ANDREI S. MARKOVITS, *The Politics of the West German Trade Unions: Strategies of Class and Interest Representation in Growth and Crisis* (1986).

Comparative analyses include EVERETT M. KASSALOW, *Trade Unions and Industrial Relations: An International Comparison* (1969); JOHN CLARKE ADAMS *et al., Comparative Labor Movements* (1952, reissued 1968); DICK GEARY, *European Labor Protest, 1848–1939* (1981); THOMAS KENNEDY, *European Labor Relations* (1980); PETER LANGE, GEORGE ROSS, and MAURIZIO VANNICELLI, *Unions, Change, and Crisis: French and Italian Union Strategy and the Political Economy, 1945–1980* (1982); E. OWEN SMITH (ed.), *Trade Unions in the Developed Economies* (1981); ALBERT A. BLUM (ed.), *International Handbook of Industrial Relations: Contemporary Developments and Research* (1981); CHARLES HANSON, SHEILA JACKSON, and DOUGLAS MILLER, *The Closed Shop: A Comparative Study in Public Policy and Trade Union Security in Britain, the USA, and West Germany* (1981); and WILLIAM BRIERLEY (ed.), *Trade Unions and the Economic Crisis of the 1980s* (1987).

Organized labour in the United States, with discussion of Canadian developments, is explored in PHILIP TAFT, *Organizing Dixie: Alabama Workers in the Industrial Era,* rev. and ed. by GARY M. FINK (1981); MEREDITH TAX, *The Rising of the Women: Feminist Solidarity and Class Conflict, 1880–1917* (1980); DAVID MONTGOMERY, *The Fall of the House of Labor: The Workplace, the State, and American Labor Activism, 1865–1925* (1987); IRVING BERNSTEIN, *The Lean Years: A History of the American Worker, 1920–1933* (1960, reprinted 1983), and *Turbulent Years: A History of the American Worker, 1933–1941* (1970); CLETUS E. DANIEL, *Bitter Harvest: A History of California Farmworkers, 1870–1941* (1981); STEPHEN MEYER III, *The Five Dollar Day: Labor Management and Social Control in the Ford Motor Company, 1908–1921* (1981); JOHN BODNAR, *Workers' World: Kinship, Community, and Protest in an Industrial Society, 1900–1940* (1982); CLINTON C. BOURDON and RAYMOND E. LEVITT, *Union and Open-Shop Construction: Compensation, Work Practices, and Labor Markets* (1980); DAVID ALAN CORBIN, *Life, Work, and Rebellion in the Coal Fields: The Southern West Virginia Miners, 1880–1922* (1981); WAYNE J. URBAN, *Why Teachers Organized* (1982); GEORGIA PANTER NIELSEN, *From Sky Girl to Flight Attendant: Women and the Making of a Union* (1982); NANCY SCHROM DYE, *As Equals and as Sisters: Feminism, the Labor Movement, and the Women's Trade Union League of New York* (1980); HARVEY A. LEVENSTEIN, *Communism, Anti-Communism, and the CIO* (1981); JOSEPH R. CONLIN (ed.), *At the Point of Production: The Local History of the I.W.W.* (1981); PATRICIA CAYO SEXTON, *The New Nightingales: Hospital Workers, Unions, New Women's Issues* (1982); ROSS M. MARTIN, *TUC, the Growth of a Pressure Group, 1868–1976* (1980); GILBERT J. GALL, *The Politics of Right to Work: The Labor Federations as Special Interests, 1943–1979* (1988); DONALD M. WELLS, *Empty Promises: Quality of Working Life Programs and the Labor Movement* (1987); CHARLES C. HECKSCHER, *The New Unionism: Employee Involvement in the Changing Corporation* (1988); MICHAEL GOLDFIELD, *The Decline of Organized Labor in the United States* (1987); H. GREGG LEWIS, *Union Relative Wage Effects: A Survey* (1986); GARY N. CHAISON, *When Unions Merge* (1986); and SALLY F. ZERKER, *The Rise and Fall of the Toronto Typographical Union, 1832–1972* (1981).

Trade unions in communist countries are studied in ALEX PRAVDA and BLAIR A. RUBLE (eds.), *Trade Unions in Communist States* (1986); EMILY C. BROWN, *Soviet Trade Unions and Labor Relations* (1966); LEONARD SHAPIRO and JOSEPH GODSON, *The Soviet Worker: From Lenin to Andropov,* 2nd ed. (1984); and STANISLAW STARSKI, *Class Struggle in Classless Poland* (1982). Other developed countries are profiled in STEPHEN S. LARGE, *Organized Workers and Socialist Politics in Interwar Japan* (1981); SOLOMON B. LEVINE, *Industrial Relations in Postwar Japan* (1958); and KENNETH F. WALKER, *Australian Industrial Relations Systems* (1970). For the developing countries, see BRUCE H. MILLEN, *The Political Role of Labor in Developing Countries* (1963, reprinted 1980); EVERETT M. KASSALOW and UKANDI G. DAMACHI (eds.), *The Role of Trade Unions in Developing Societies* (1978); ROGER SOUTHALL (ed.), *Trade Unions and the New Industrialization of the Third World* (1988); ROBIN COHEN, PETER C.W. GUTKIND, and PHYLLIS BRAZIER, *Peasants and Proletarians: The Struggles of Third World Workers* (1979); and ROSALIND E. BOYD, ROBIN COHEN, and PETER C.W. GUTKIND (eds.), *International Labour and the Third World: The Making of a New Working Class* (1987). (A.F.St./Ed.)

The World Wars

World War I (1914–18) and World War II (1939–45) are so called because they were wars in which virtually all of the most powerful nations of the world were involved, either on one side or the other, and because they were fought over large areas of the globe, on both land and sea. World War I pitted Germany, Austria-Hungary, and Turkey against France, Great Britain, Russia, the United States, China, and a host of smaller nations. In World War II, Japan and the Fascist nations of Germany and Italy confronted the Soviet Union, the United States, Great Britain, France, China, and many smaller nations. World War II was a truly global conflict that was fought not only throughout the European continent but also in the islands of the Pacific Ocean, in China and Southeast Asia, in North Africa, and throughout the world's oceans.

The world wars are the great watersheds of 20th-century geopolitical history. World War I led to the fall of four great imperial dynasties (in Germany, Russia, Austria-Hungary, and Turkey), resulted in the Bolshevik Revolution in Russia, and, in its destabilization of European society, laid the groundwork for World War II. World War II resulted in the extension of the Soviet Union's power to the nations of eastern Europe, enabled a Communist movement to eventually achieve power in China, and marked the decisive shift of power in the world away from the states of western Europe and toward the United States and the Soviet Union. The world wars were almost unprecedented in their slaughter, carnage, and destruction, as mechanized armies of millions of soldiers destroyed both each other and the territories they were fighting over. The 40,000,000–50,000,000 deaths incurred in World War II make it the bloodiest conflict as well as the largest war in history.

This article treats the military course and the diplomatic history of World Wars I and II. Other aspects of the world wars are treated in several articles in the *Macropædia.* The wars' causes and consequences, as well as the political and other events leading up to them and resulting from them, are treated in INTERNATIONAL RELATIONS, 20TH-CENTURY. The strategic and tactical considerations of the wars' participants are treated in WAR, THE THEORY AND CONDUCT OF, while the weapons and weapons systems used in the wars are treated in WAR, THE TECHNOLOGY OF.

The article is divided into the following sections:

World War I 961
The outbreak of war
The initial stages of the war 964
Initial strategies
The war in the west, 1914
The Eastern and other fronts, 1914
The years of stalemate 970
The Western and Eastern fronts, 1915
Other fronts, 1915–16
Major developments in 1916
Developments in 1917
The last offensives and the Allies' victory 981
The Western Front, March–September 1918
Other developments in 1918
The final offensive on the Western Front
Killed, wounded, and missing
World War II 987
Axis initiative and Allied reaction 987
The war in Europe, 1939–41
Other fronts, 1940–41
Invasion of the Soviet Union, 1941
The war in the Pacific, 1938–41
Developments from autumn 1941 to spring 1942
The Allies' first decisive successes 1005
The Allied landings in Europe and the defeat of the Axis powers 1010
Developments from autumn 1943 to summer 1944
Developments from summer 1944 to autumn 1945
Costs of the war
Bibliography 1024

WORLD WAR I

THE OUTBREAK OF WAR

With Serbia already much aggrandized by the two Balkan Wars (1912–13, 1913), Serbian nationalists turned their attention back to the idea of "liberating" the South Slavs of Austria-Hungary. Colonel Dragutin Dimitrijević, head of Serbia's military intelligence, was also, under the alias "Apis," head of the secret society Union or Death, pledged to the pursuit of this pan-Serbian ambition. Believing that the Serbs' cause would be served by the death of the Austrian archduke Francis Ferdinand, heir presumptive to the Austrian emperor Francis Joseph, and learning that the Archduke was about to visit Bosnia on a tour of military inspection, Apis plotted his assassination. Nikola Pašić, the Serbian prime minister and an enemy of Apis, heard of the plot and warned the Austrian government of it, but his message was too cautiously worded to be understood.

At 11:15 AM, on June 28, 1914, in the Bosnian capital, Sarajevo, Francis Ferdinand and his morganatic wife, Sophie, duchess of Hohenberg, were shot dead by a Bosnian Serb, Gavrilo Princip.

The chief of the Austro-Hungarian general staff, Franz, Graf Conrad von Hötzendorf, and the foreign minister, Leopold, Graf von Berchtold, saw the crime as the occasion for measures to humiliate Serbia and so to enhance Austria-Hungary's prestige in the Balkans; and Conrad had already (October 1913) been assured by William II of Germany's support if Austria-Hungary should start a preventive war against Serbia. This assurance was confirmed in the week following the assassination, before William, on July 6, set off upon his annual cruise to the North Cape, off Norway.

The Austrian ultimatum

The Austrians decided to present an unacceptable ultimatum to Serbia and then to declare war, relying on Germany to deter Russia from intervention. Though the terms of the ultimatum were finally approved on July 19, its delivery was postponed to the evening of July 23, since by that time the French president, Raymond Poincaré, and his premier, René Viviani, who had set off on a state visit to Russia on July 15, would be on their way home and therefore unable to concert an immediate reaction with their Russian allies. When the delivery was announced, on July 24, Russia declared that Austria-Hungary must not be allowed to crush Serbia.

Serbia replied to the ultimatum on July 25, accepting most of its demands but protesting against two of them, namely, that Serbian officials (unnamed) should be dismissed at Austria-Hungary's behest and that Austro-Hungarian officials should take part, on Serbian soil, in proceedings against organizations hostile to Austria-Hungary. Though Serbia offered to submit the issue to international arbitration, Austria-Hungary promptly severed diplomatic relations and ordered partial mobilization.

Home from his cruise on July 27, William learned on July 28 how Serbia had replied to the ultimatum. At once he instructed the German Foreign Office to tell Austria-Hungary that there was no longer any justification for war and that it should content itself with a temporary occupation of Belgrade. But, meanwhile, the German Foreign Office had been giving such encouragement to Berchtold that already on July 27 he had persuaded Francis Joseph to authorize war against Serbia. War was, in fact, declared on July 28, and Austro-Hungarian artillery began to bombard Belgrade the next day. Russia then ordered partial mobilization against Austria-Hungary; and on July 30, when Austria-Hungary was riposting conventionally with an order of mobilization on its Russian frontier, Russia ordered general mobilization. Germany, which since July 28 had still been hoping, in disregard of earlier warning hints from Great Britain, that Austria-Hungary's war against Serbia could be "localized" to the Balkans, was now disillusioned insofar as eastern Europe was concerned. On July 31 Germany sent a 24-hour ultimatum requiring Russia to halt its mobilization and an 18-hour ultimatum requiring France to promise neutrality in the event of war between Russia and Germany.

Both Russia and France predictably ignored these demands. On August 1, Germany ordered general mobilization and declared war against Russia, and France likewise ordered general mobilization. The next day, Germany sent troops into Luxembourg and demanded from Belgium free passage for German troops across its neutral territory. On August 3 Germany declared war against France.

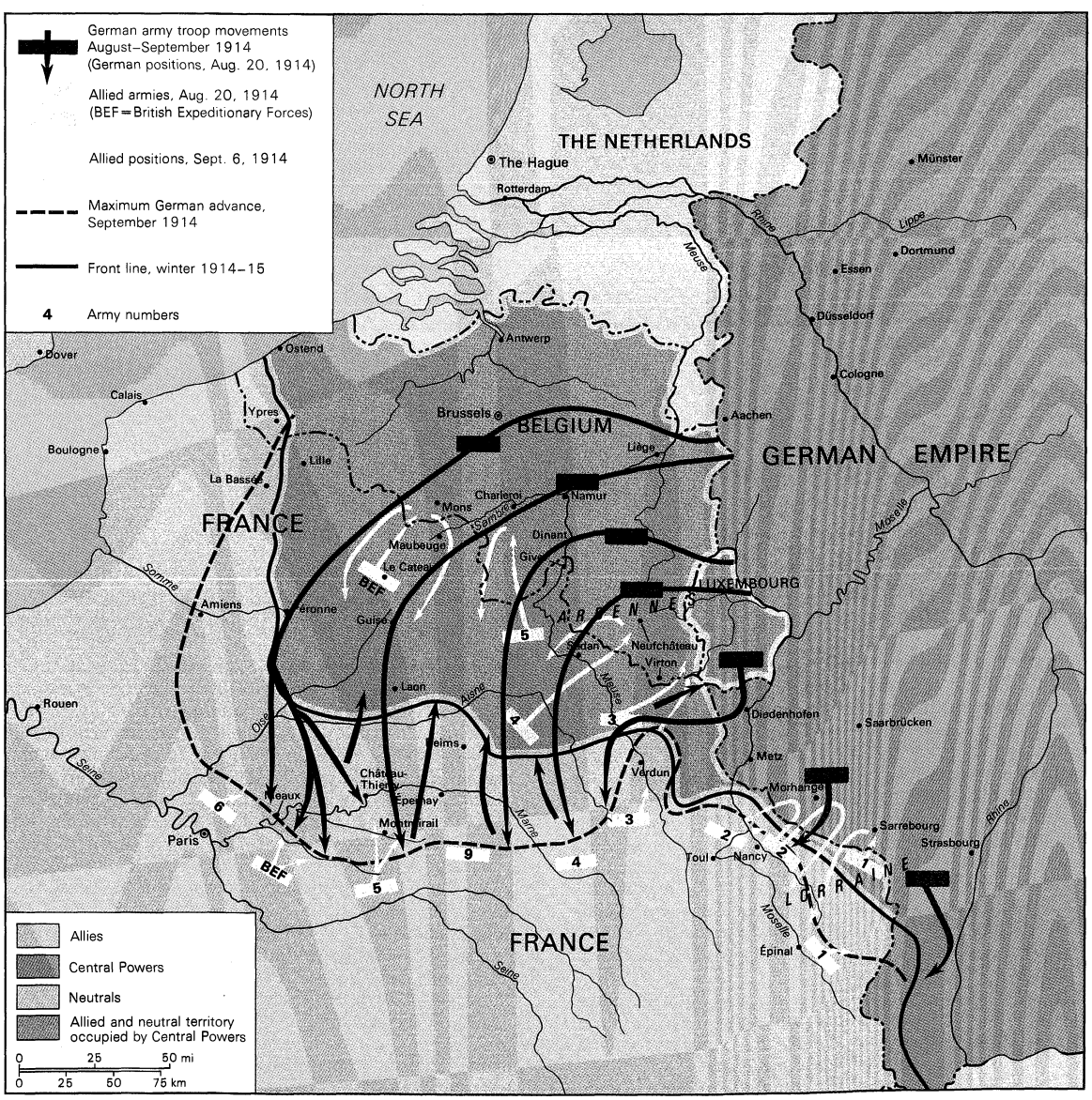
The German invasion of Belgium

In the night of August 3–4 German forces invaded Belgium. Thereupon, Great Britain, which had no concern with Serbia and no express obligation to fight either for Russia or for France but was expressly committed to defend Belgium, on August 4 declared war against Germany.

Austria-Hungary declared war against Russia on August 5; Serbia against Germany on August 6; Montenegro against Austria-Hungary on August 7 and against Germany on August 12; France and Great Britain against Austria-Hungary on August 10 and on August 12, respectively; Japan against Germany on August 23; Austria-Hungary against Japan on August 25 and against Belgium on August 28.

Romania had renewed its secret anti-Russian alliance of 1883 with the Central Powers on Feb. 26, 1914, but now chose to remain neutral. Italy had confirmed the Triple Alliance on Dec. 7, 1912, but could now propound formal arguments for disregarding it: first, Italy was not obliged to support its allies in a war of aggression; second, the original treaty of 1882 had stated expressly that the alliance was not against England.

On Sept. 5, 1914, Russia, France, and Great Britain concluded the Treaty of London, each promising not to make a separate peace with the Central Powers. Thenceforth, they could be called the Allied, or Entente, Powers, or simply the Allies.



German and Allied movements on the Western Front, August–September 1914.

The outbreak of war in August 1914 was generally greeted with confidence and jubilation by the peoples of Europe, among whom it inspired a wave of patriotic feeling and celebration. Few people imagined how long or how disastrous a war between the great nations of Europe could be, and most believed that their country's side would be victorious within a matter of months. The war was welcomed either patriotically, as a defensive one imposed by national necessity, or idealistically, as one for upholding right against might, the sanctity of treaties, and international morality.

Forces and resources of the combatant nations in 1914. When war broke out the Allied Powers possessed greater overall demographic, industrial, and military resources than the Central Powers and enjoyed easier access to the oceans for trade with neutral countries, particularly with the United States. Table 1 shows the population, steel production, and armed strengths of the two rival coalitions in 1914.

All the initial belligerents in World War I were self-sufficient in food except Great Britain and Germany. Great Britain's industrial establishment was slightly superior to Germany's (17 percent of world trade in 1913 as compared with 12 percent for Germany), but Germany's diversified chemical industry facilitated the production of ersatz, or substitute, materials, which compensated for the worst shortages ensuing from the British wartime blockade. The German chemist Fritz Haber was already developing a process for the fixation of nitrogen from air; this process made Germany self-sufficient in explosives and thus no longer dependent on imports of nitrates from Chile.

Table 1: Strength of the Belligerents, Aug. 4, 1914

resources	Central Powers	Allied Powers
Population (in millions)	115.2	265.5
Steel production (in millions of metric tons)	17.0	15.3
Army divisions available for mobilization in August 1914	146	212
Modern battleships	20	39

Of all the initial belligerent nations, only Great Britain had a volunteer army, and this was quite small at the start of the war. The other nations had much larger conscript armies that required three to four years of service from able-bodied males of military age, to be followed by several years in reserve formations. Military strength on land was counted in terms of divisions composed of 12,000–20,000 officers and men. Two or more divisions made up an army corps, and two or more corps made up an army. An army could thus comprise anywhere from 50,000 to 250,000 men.

The land forces of the belligerent nations at the outbreak of war in August 1914 are shown in Table 2.

Table 2: Land Forces of the Belligerents, Aug. 4, 1914

country	regular divisions (with number of field armies)	other land forces	total manpower
Central Powers			
Germany	98 (8)	27 *Landwehr* brigades	1,900,000
Austria-Hungary	48 (6)		450,000
Allied Powers			
Russia	102* (6)		1,400,000
France	72 (5)		1,290,000
Serbia	11 (3)		190,000
Belgium	7 (1)	69,000 fortress troops	186,000
Great Britain	6 (1)	14 territorial divisions†	120,000

*Russia planned to have 114 divisions in the field three months after mobilization.
†Restricted in 1914 to service at home.

The higher state of discipline, training, leadership, and armament of the German Army reduced the importance of the initial numerical inferiority of the armies of the Central Powers. Because of the comparative slowness of mobilization, poor higher leadership, and lower scale of armament of the Russian armies, there was an approximate balance of forces between the Central Powers and the Allies in August 1914 that prevented either side from gaining a quick victory.

Germany and Austria also enjoyed the advantage of "interior lines of communication," which enabled them to send their forces to critical points on the battlefronts by the shortest route. According to one estimate, Germany's railway network made it possible to move eight divisions simultaneously from the Western Front to the Eastern Front in four and a half days.

Even greater in importance was the advantage that Germany derived from its strong military traditions and its cadre of highly efficient and disciplined regular officers. Skilled in directing a war of movement and quick to exploit the advantages of flank attacks, German senior officers were to prove generally more capable than their Allied counterparts at directing the operations of large troop formations.

Table 3: British and German Naval Strength, August 1914

type	British	German
Dreadnought battleships	20	14
Battle cruisers	9	4
Pre-dreadnought battleships	39	22
Armoured cruisers	34	9
Cruisers	64	41
Destroyers	301*	144
Submarines	65	28

*Including Canadian, Australian, and New Zealand destroyers of all classes.

Sea power was largely reckoned in terms of capital ships, or dreadnought battleships and battle cruisers having extremely large guns. Despite intensive competition from the Germans, the British had maintained their superiority in numbers, with the result that, in capital ships, the Allies had an almost two-to-one advantage over the Central Powers. Table 3 compares the strength of the two principal rivals at sea, Great Britain and Germany.

The numerical superiority of the British Navy, however, was offset by the technological lead of the German Navy in many categories, such as range-finding equipment, magazine protection, searchlights, torpedoes, and mines. Great Britain relied on the Royal Navy not only to ensure necessary imports of food and other supplies in wartime but also to sever the Central Powers' access to the markets of the world. With superior numbers of warships, Great Britain could impose a blockade that gradually weakened Germany by preventing imports from overseas.

Technology of war in 1914. The planning and conduct of war in 1914 were crucially influenced by the invention of new weapons and the improvement of existing types since the Franco-German War of 1870–71. The chief developments of the intervening period had been the machine gun and the rapid-fire field artillery gun. The modern machine gun, which had been developed in the 1880s and '90s, was a reliable belt-fed gun capable of sustained rates of extremely rapid fire; it could fire 600 bullets per minute with a range of more than 1,000 yards (900 metres). In the realm of field artillery, the period leading up to the war saw the introduction of improved breech-loading mechanisms and brakes. Without a brake or recoil mechanism, a gun lurched out of position during firing and had to be re-aimed after each round. The new improvements were epitomized in the French 75-millimetre field gun; it remained motionless during firing, and it was not necessary to readjust the aim in order to bring sustained fire on a target. Machine guns and rapid-firing artillery, when used in combination with trenches and barbed-wire emplacements, gave a decided advantage to the defense, since these weapons' rapid and sustained firepower could decimate a frontal assault by either infantry or cavalry.

There was a considerable disparity in 1914 between the deadly effectiveness of modern armaments and the doctrinal teachings of some armies. The South African War and the Russo-Japanese War had revealed the futility of frontal infantry or cavalry attacks on prepared positions when unaccompanied by surprise, but few military leaders

foresaw that the machine gun and the rapid-firing field gun would force armies into trenches in order to survive. Instead, war was looked upon by many leaders in 1914 as a contest of national wills, spirit, and courage. A prime example of this attitude was the French Army, which was dominated by the doctrine of the offensive. French military doctrine called for headlong bayonet charges of French infantrymen against the German rifles, machine guns, and artillery. German military thinking, under the influence of Alfred, Graf von Schlieffen, sought, unlike the French, to avoid frontal assaults but rather to achieve an early decision by deep flanking attacks; and at the same time to make use of reserve divisions alongside regular formations from the outset of war. The Germans paid greater attention to training their officers in defensive tactics using machine guns, barbed wire, and fortifications.

The initial stages of the war

INITIAL STRATEGIES

The Schlieffen Plan. Years before 1914, successive chiefs of the German general staff had been foreseeing Germany's having to fight a war on two fronts at the same time, against Russia in the east and France in the west, whose combined strength was numerically superior to the Central Powers'. The elder Helmuth von Moltke, chief of the German general staff from 1858 to 1888, decided that Germany should stay at first on the defensive in the west and deal a crippling blow to Russia's advanced forces before turning to counterattack the French advance. His immediate successor, Alfred von Waldersee, also believed in staying on the defensive in the west. Alfred, Graf von Schlieffen, who served as chief of the German general staff from 1891 to 1905, took a contrary view, and it was the plan he developed that was to guide Germany's initial wartime strategy. Schlieffen realized that on the outbreak of war Russia would need six full weeks to mobilize and assemble its vast armies, given the immense Russian countryside and population, the sparsity of the rail network, and the inefficiency of the government bureaucracy. Taking advantage of this fact, Schlieffen planned to initially adopt a purely defensive posture on the Eastern Front with a minimal number of troops facing Russia's slowly gathering armies. Germany would instead concentrate almost all of its troops in the west against France and would seek to bypass France's frontier fortifications by an offensive through neutral Belgium to the north. This offensive would sweep westward and then southward through the heart of northern France, capturing the capital and knocking that country out of the war within a few weeks. Having gained security in the west, Germany would then shift its troops to the east and destroy the Russian menace with a similar concentration of forces.

By the time of his retirement in 1905, Schlieffen had elaborated a plan for a great wheeling movement of the right (northern) wing of the German armies not only through central Belgium but also, in order to bypass the Belgian fortresses of Liège and Namur in the Meuse Valley, through the southernmost part of The Netherlands. With their right wing entering France near Lille, the Germans would continue to wheel westward until they were near the English Channel; they would then turn southward so as to sever the French armies' line of retreat from France's eastern frontier to the south; and the outermost arc of the wheel would sweep southward west of Paris, in order to avoid exposing the German right flank to a counterstroke launched from the city's outskirts. If the Schlieffen Plan succeeded, Germany's armies would simultaneously encircle the French Army from the north, overrun all of northeastern France, and capture Paris, thus forcing France into a humiliating surrender. The large wheeling movement that the plan envisaged required correspondingly large forces for its execution, in view of the need to keep up the numerical strength of the long-stretched marching line and the need to leave adequate detachments on guard over the Belgian fortresses that had been bypassed. Accordingly, Schlieffen allocated nearly seven-eighths of Germany's available troop strength to the execution of the wheeling movement by the right and centre wings, leaving only one-eighth to face a possible French offensive on Germany's western frontier. Thus, the maximum of strength was allocated to the wheel's edge—that is, to the right. Schlieffen's plan was observed by the younger Helmuth von Moltke, who became chief of the general staff in 1906. Moltke was still in office when war broke out in 1914.

Moltke appointed chief of the general staff

Eastern Front strategy, 1914. Russian Poland, the westernmost part of the Russian Empire, was a thick tongue of land enclosed to the north by East Prussia, to the west by German Poland (Poznania) and by Silesia, and to the south by Austrian Poland (Galicia). It was thus obviously exposed to a two-pronged invasion by the Central Powers; but the Germans, apart from their grand strategy of crushing France before attempting anything against Russia, took note of the poverty of Russian Poland's transportation network and so were disinclined to overrun that vulnerable area prematurely. Austria-Hungary, however, whose frontier with Russia lay much farther east than Germany's and who was moreover afraid of disaffection among the Slav minorities, urged some immediate action to forestall a Russian offensive. Moltke therefore agreed to the Austrian general staff's suggestion for a northeastward thrust by the Austrian Army into Russian Poland—the more readily because it would occupy the Russians during the crisis in France.

The Russians, for their part, would have preferred to concentrate their immediately available forces against Austria and to leave Germany undisturbed until their mobilization should have been completed. The French were anxious to relieve the German pressure against themselves, however, and so they persuaded the Russians to undertake an offensive involving two armies against the Germans in East Prussia simultaneously with one involving four armies against the Austrians in Galicia. The Russian Army, whose proverbial slowness and unwieldy organization dictated a cautious strategy, thus undertook an extra offensive against East Prussia that only an army of high mobility and tight organization could have hoped to execute successfully.

The strategy of the Western Allies, 1914. For some 30 years after 1870, considering the likelihood of another German war, the French high command had subscribed to the strategy of an initial defensive to be followed by a counterstroke against the expected invasion: a great system of fortresses was created on the frontier, but gaps were left in order to "canalize" the German attack. France's alliance with Russia and its entente with Great Britain, however, encouraged a reversal of plan, and after the turn of the century a new school of military thinkers began to argue for an offensive strategy. The advocates of the offensive *à l'outrance* ("to the utmost") gained control of the French military machine, and in 1911 a spokesman of this school, General J.-J.-C. Joffre, was designated chief of the general staff. He sponsored the notorious Plan XVII, with which France went to war in 1914.

France's Plan XVII

Plan XVII gravely underestimated the strength that the Germans would deploy against France. Accepting the possibility that the Germans might employ their reserve troops along with regular troops at the outset, Plan XVII estimated the strength of the German Army in the west at a possible maximum of 68 infantry divisions. The Germans actually deployed the equivalent of 83½ divisions, counting *Landwehr* (reserve troops) and *Ersatz* (low-grade substitute troops) divisions. But French military opinion ignored or doubted this possibility; and during the war's crucial opening days, when the rival armies were concentrating and moving forward, the French Intelligence counted only Germany's regular divisions in its estimates of the enemy strength. This was a serious miscalculation. Plan XVII also miscalculated the direction and scope of the coming onslaught: though it foresaw an invasion through Belgium, it assumed that the Germans would take the route through the Ardennes, thereby exposing their communications to attack. Basing itself on the idea of an immediate and general offensive, Plan XVII called for a French thrust toward the Saar into Lorraine by the 1st and 2nd armies, while on the French left (the north) the 3rd and 5th armies, facing Metz and the Ardennes, respectively, stood ready either to launch an offensive between

Metz and Thionville or to strike from the north at the flank of any German drive through the Ardennes. When war broke out, it was taken for granted that the small British Expeditionary Force (BEF) under Sir John French should be used as an adjunct to France's forces, more or less as the French might see fit. It is clearly evident that the French were oblivious to the gigantic German offensive that was being aimed at their left (northern) wing.

THE WAR IN THE WEST, 1914

For the smooth working of their plan for the invasion of France, the Germans had preliminarily to reduce the ring fortress of Liège, which commanded the route prescribed for their 1st and 2nd armies and which was the foremost stronghold of the Belgian defenses. German troops crossed the frontier into Belgium on the morning of August 4. Thanks to the resolution of a middle-aged staff officer, Erich Ludendorff, a German brigade occupied the town of Liège itself in the night of August 5–6 and the citadel on August 7; but the surrounding forts held out stubbornly until the Germans brought their heavy howitzers into action against them on August 12. These 420-millimetre siege guns proved too formidable for the forts, which one by one succumbed. The vanguard of the German invasion was already pressing the Belgian field army between the Gete River and Brussels, when the last of the Liège forts fell on August 16. The Belgians then withdrew northward to the entrenched camp of Antwerp. On August 20 the German 1st Army entered Brussels while the 2nd Army appeared before Namur, the one remaining fortress barring the Meuse route into France.

The initial clashes between the French and German armies along the Franco-German and Franco-Belgian frontiers are collectively known as the Battle of the Frontiers. This group of engagements, which lasted from August 14 until the beginning of the First Battle of the Marne on September 6, was to be the largest battle of the war and was perhaps the largest battle in human history up to that time, given the fact that a total of more than 2,000,000 troops were involved. (See map.)

The planned French thrust into Lorraine, totaling 19 divisions, started on August 14 but was shattered by the German 6th and 7th armies in the Battle of Morhange-Sarrebourg (August 20–22). Yet this abortive French offensive had an indirect effect on the German plan. For when the French attack in Lorraine developed, Moltke was tempted momentarily to postpone the right-wing sweep and instead to seek a victory in Lorraine. This fleeting impulse led him to divert to Lorraine the six newly formed *Ersatz* divisions that had been intended to increase the weight of his right wing. This was the first of several impromptu decisions by Moltke that were to fatally impair the execution of the Schlieffen Plan.

Meanwhile, the German imperial princes who commanded armies on the Germans' left (southern) wing in Lorraine were proving unwilling to forfeit their opportunity for personal glory. Crown Prince Rupert of Bavaria on August 20 ordered his 6th Army to counterattack instead of continuing to fall back before the French advance as planned, and Crown Prince William of Germany ordered his 5th Army to do the same. The strategic result of these unplanned German offensives was merely to throw the French back onto a fortified barrier that both restored and augmented their power of resistance. Thus, the French were soon afterward enabled to dispatch troops to reinforce their left flank—a redistribution of strength that was to have far-reaching results in the decisive Battle of the Marne.

While this seesaw campaign in Lorraine was taking place, more decisive events were occurring to the northwest. The German attack on Liège had awakened Joffre to the reality of a German advance through Belgium, but not to its strength or to the wideness of its sweep. In preparing a counterattack against the German advance through Belgium, Joffre envisaged a pincer movement, with the French 3rd and 4th armies on the right and the 5th, supported by the BEF, on the left, to trap the Germans in the Meuse–Ardennes area south of Liège. The fundamental flaw in this new French plan was that the Germans had deployed about 50 percent more troops than the French had estimated, and for a vaster enveloping movement. Consequently, while the right-hand claw of the French pincer (23 divisions) collided with the German 5th and 4th armies (20 divisions) in the Ardennes and was thrown back, the left-hand claw (13 French and four British divisions) found itself nearly trapped between the German 1st and 2nd armies, with a total of 30 divisions, on the one hand, and the 3rd, on the other. As the French 5th Army, under General Charles Lanrezac, was checked in its offensive south of the Sambre River by a German attack on August 21, the British, who reached Mons on August 22, at first agreed to stand there to cover Lanrezac's left; but on August 23 news of the fall of Namur and of the German 3rd Army's presence near Dinant induced Lanrezac to wisely order a general retreat; and on August 24 the British began their retreat from Mons, just in time to escape envelopment by the German 1st Army's westward march around their unprotected left flank.

At last Joffre realized the truth and the utter collapse of Plan XVII. Resolution was his greatest asset, and with imperturbable coolness he formed a new plan out of the wreckage. Joffre decided to swing the Allied centre and left back southwestward from the Belgian frontier to a line pivoted on the French fortress of Verdun and at the same time to withdraw some strength from the right wing so as to be able to station a newly created 6th Army on the extreme left, north of Paris. This plan might, in turn, have collapsed if the Germans had not themselves departed from Schlieffen's original plan due to a combination of Moltke's indecisiveness, poor communications between his headquarters and the field army commanders of the German right wing, and Moltke's resulting confusion about the developing tactical situation. In the first place, the German right wing was weakened by the subtraction of 11 divisions; four were detached to watch Antwerp and to invest French fortresses near the Belgian frontier, instead of using reserve and *Ersatz* troops for this as earlier intended, and seven more regular divisions were transferred to check the Russian advance into East Prussia (see below). In the second place, Alexander von Kluck, in command of the 1st Army, did, in fact, wheel inward north of Paris rather than southwest of the city.

Kluck's change of direction meant the inevitable abandonment of the original wide sweep around the far (western) side of Paris. Now the flank of this wheeling German line would pass the near side of Paris and across the face of the Paris defenses into the valley of the Marne River. The premature inward wheel of Kluck's 1st Army before Paris had been reached thus exposed the German extreme right wing to a flank attack and a possible counter-envelopment. On September 4 Moltke decided to abandon the original Schlieffen Plan and substituted a new one: the German 4th and 5th armies should drive southeastward from the Ardennes into French Lorraine west of Verdun and then converge with the southwestward advance of the 6th and 7th armies from Alsace against the Toul–Épinal line of fortifications, so as to envelop the whole French right wing; the 1st and 2nd armies, in the Marne Valley, should stand guard, meanwhile, against any French countermove from the vicinity of Paris. But such an Allied countermove had already begun before the new German plan could be put into effect.

The First Battle of the Marne

Already on September 3, General J.-S. Gallieni, the military governor of Paris, had guessed the significance of the German 1st Army's swing inward to the Marne east of Paris. On September 4 Joffre, convinced by Gallieni's arguments, decisively ordered his whole left wing to turn about from their retreat and to begin a general offensive against the Germans' exposed right flank on September 6. The French 6th Army, under M.-J. Maunoury, forewarned by Gallieni, had actually begun attacking on September 5; and its pressure caused Kluck finally to engage the whole 1st Army in support of his right flank when he was still no farther up the Marne Valley than Meaux, with nothing but a cavalry screen stretched across the 30 miles between him and Karl von Bülow's 2nd Army (at Montmirail). While the French 5th Army was turning to attack Bülow, the BEF (between the 5th and the 6th armies) was still

continuing its retreat for another day; but on September 9 Bülow learned that the British too had turned and were advancing into the gap between him and Kluck. He therefore ordered the 2nd Army to retreat, thus obliging Kluck to do likewise with the 1st. The counterattack of the French 5th and 6th armies and the BEF developed into a general counterattack by the entire left and centre of the French Army. This counterattack is known as the First Battle of the Marne. By September 11 the German retreat extended to all the German armies.

There were several reasons for this extraordinary turn of events. Chief among them was the utter exhaustion of the German soldiery of the right wing, some of whom had marched more than 150 miles (240 kilometres) under conditions of frequent battle. Their fatigue was ultimately a by-product of the Schlieffen Plan itself, for while the retreating French had been able to move troops by rail to various points within the circle formed by the front, the German troops had found their advance hampered by demolished bridges and destroyed rail lines. Their food and ammunition supply was consequently restricted, and the troops also had to make their advance by foot. Moreover, the Germans had underestimated the resilient spirit of the French troops, who had maintained their courage and morale and their confidence in their commanders. This fact was strikingly evidenced by the comparatively small number of prisoners taken by the Germans in the course of what was undeniably a precipitous French retreat.

Meanwhile, the assault by the German 6th and 7th armies on the defenses of the French eastern frontier had already proved a predictably expensive failure, and the German attempt at a partial envelopment pivoted on Verdun was abandoned. The German right wing withdrew northward from the Marne and made a firm stand along the Lower Aisne River and the Chemin des Dames ridge. Along the Aisne the preponderant power of the defense over the offense was reemphasized as the Germans repelled successive Allied attacks from the shelter of trenches. The First Battle of the Aisne marked the real beginning of trench warfare on the Western Front. Both sides were in the process of discovering that, in lieu of frontal assaults for which neither had the manpower readily available, the only alternative was to try to overlap and envelop the other's flank, in this case the one on the side pointing toward the North Sea and the English Channel. Thus began the "Race to the Sea," in which the developing trench networks of both sides were quickly extended northwestward until they reached the Atlantic at a point just inside coastal Belgium, west of Ostend.

The First Battle of the Marne succeeded in pushing the Germans back for a distance of 40 to 50 miles and thus saved the capital city of Paris from capture. In this respect it was a great strategic victory, since it enabled the French to renew their confidence and to continue the war. But the great German offensive, though unsuccessful in its object of knocking France out of the war, had enabled the Germans to capture a large portion of northeastern France. The loss of this heavily industrialized region, which contained much of the country's coal, iron, and steel production, was a serious blow to the continuation of the French war effort.

The Belgian Army, meanwhile, had fallen back to the fortress city of Antwerp, which ended up behind the German lines. The Germans began a heavy bombardment of Antwerp on September 28, and Antwerp surrendered to the Germans on October 10.

After the failure of his first two attempts to turn the Germans' western flank (one on the Somme, the other near Arras), Joffre obstinately decided to try again yet farther north with the BEF—which in any case was being moved northward from the Aisne. The BEF, accordingly, was deployed between La Bassée and Ypres, while on the left the Belgians—who had wisely declined to participate in the projected attack—continued the front along the Yser down to the Channel. Erich von Falkenhayn, however, who on September 14 had succeeded Moltke as chief of the German general staff, had foreseen what was coming and had prepared a counterplan: one of his armies, transferred from Lorraine, was to check the expected offensive, while another was to sweep down the coast and crush the attackers' left flank. The British attack was launched from Ypres on October 19, the German thrust the next day. Though the Belgians of the Yser had been under increasing pressure for two days already, both Sir John French and Ferdinand Foch, Joffre's deputy in the north, were slow to appreciate what was happening to their "offensive"; but in the night of October 29–30 the Belgians had to open the sluices on the Yser River to save themselves by flooding the Germans' path down the coast. The Battle of Ypres had its worst crises on October 31 and November 11 and did not die down into trench warfare until November 22.

The First Battle of Ypres

By the end of 1914 the casualties the French had so far sustained in the war totaled about 380,000 killed and 600,000 wounded; the Germans had lost a slightly smaller number. With the repulse of the German attempt to break through at the Battle of Ypres, the strained and exhausted armies of both sides settled down into trench warfare. The trench barrier was consolidated from the Swiss frontier to the Atlantic; the power of modern defense had triumphed over the attack, and stalemate ensued. The military history of the Western Front during the next three years was to be a story of the Allies' attempts to break this deadlock.

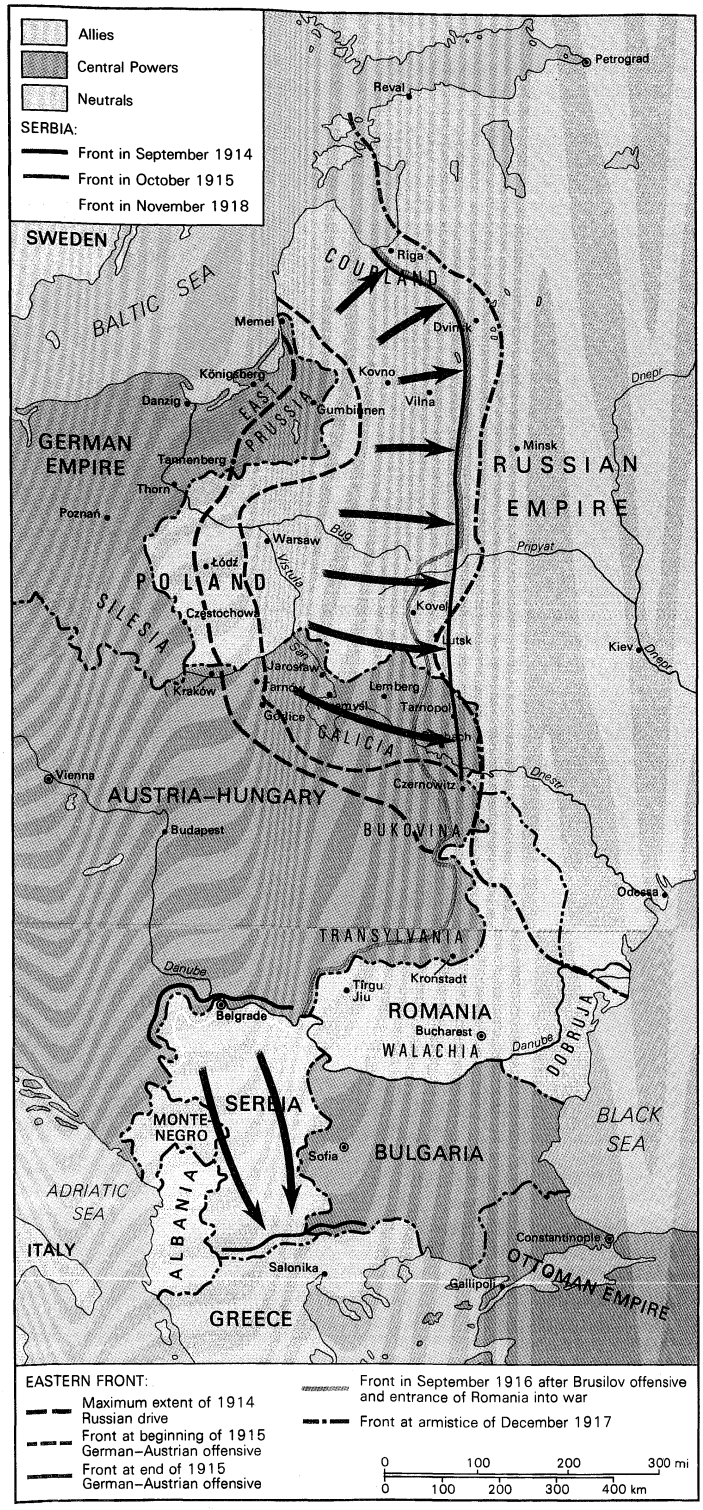
THE EASTERN AND OTHER FRONTS, 1914

The war in the east, 1914. On the Eastern Front, greater distances and quite considerable differences between the equipment and quality of the opposing armies ensured a fluidity of the front that was lacking in the west. Trench lines might form, but to break them was not difficult, particularly for the German army, and then mobile operations of the old style could be undertaken.

Urged by the French to take offensive action against the Germans, the Russian commander in chief, Grand Duke Nicholas, took it loyally but prematurely, before the cumbrous Russian war machine was ready, by launching a pincer movement against East Prussia. Under the higher control of General Ya.G. Zhilinsky, two armies, the 1st, or Vilna, Army under P.K. Rennenkampf and the 2nd, or Warsaw, Army under A.V. Samsonov, were to converge, with a two-to-one superiority in numbers, on the German 8th Army in East Prussia from the east and the south, respectively. Rennenkampf's left flank would be separated by 50 miles from Samsonov's right flank.

Max von Prittwitz und Gaffron, commander of the 8th Army, with his headquarters at Neidenburg (Nidzica), had seven divisions and one cavalry division on his eastern front but only the three divisions of Friedrich von Scholtz's XX Corps on his southern. He was therefore dismayed to learn, on August 20, when the bulk of his forces had been repulsed at Gumbinnen (August 19–20) by Rennenkampf's attack from the east, that Samsonov's 13 divisions had crossed the southern frontier of East Prussia and were thus threatening his rear. He initially considered a general retreat, but when his staff objected to this, he approved their counterproposal of an attack on Samsonov's left flank, for which purpose three divisions were to be switched in haste by rail from the Gumbinnen front to reinforce Scholtz (the rest of the Gumbinnen troops could make their retreat by road). The principal exponent of this counterproposal was Lieutenant Colonel Max Hoffmann. Prittwitz, having moved his headquarters northward to Mühlhausen (Młynary), was surprised on August 22 by a telegram announcing that General Paul von Hindenburg, with Ludendorff as his chief of staff, was coming to supersede him in command. Arriving the next day, Ludendorff readily confirmed Hoffmann's dispositions for the blow at Samsonov's left.

Meanwhile, Zhilinsky was not only giving Rennenkampf time to reorganize after Gumbinnen but even instructing him to invest Königsberg instead of pressing on to the west. When the Germans on August 25 learned from an intercepted Russian wireless message (the Russians habitually transmitted combat directives "in clear," not in code) that Rennenkampf was in no hurry to advance, Ludendorff saw a new opportunity. Developing the plan put forward by Hoffmann, Ludendorff concentrated about six divisions against Samsonov's left wing. This force, inferior in strength, could not have been decisive, but



The Eastern Front, 1914–17.

Ludendorff then took the calculated risk of withdrawing the rest of the German troops, except for a cavalry screen, from their confrontation with Rennenkampf and rushing them southwestward against Samsonov's right wing. Thus, August von Mackensen's XVII Corps was taken from near Gumbinnen and moved southward to duplicate the planned German attack on Samsonov's left with an attack on his right, thus completely enveloping the Russian 2nd Army. This daring move was made possible by the notable absence of communication between the two Russian field commanders, whom Hoffmann knew to personally dislike each other. Under the Germans' converging blows Samsonov's flanks were crushed and his centre surrounded during August 26–31. The outcome of this military masterpiece, called the Battle of Tannenberg, was the destruction or capture of almost the whole of Samsonov's army. The history of imperial Russia's unfortunate participation in World War I is epitomized in the ignominious outcome of the Battle of Tannenberg.

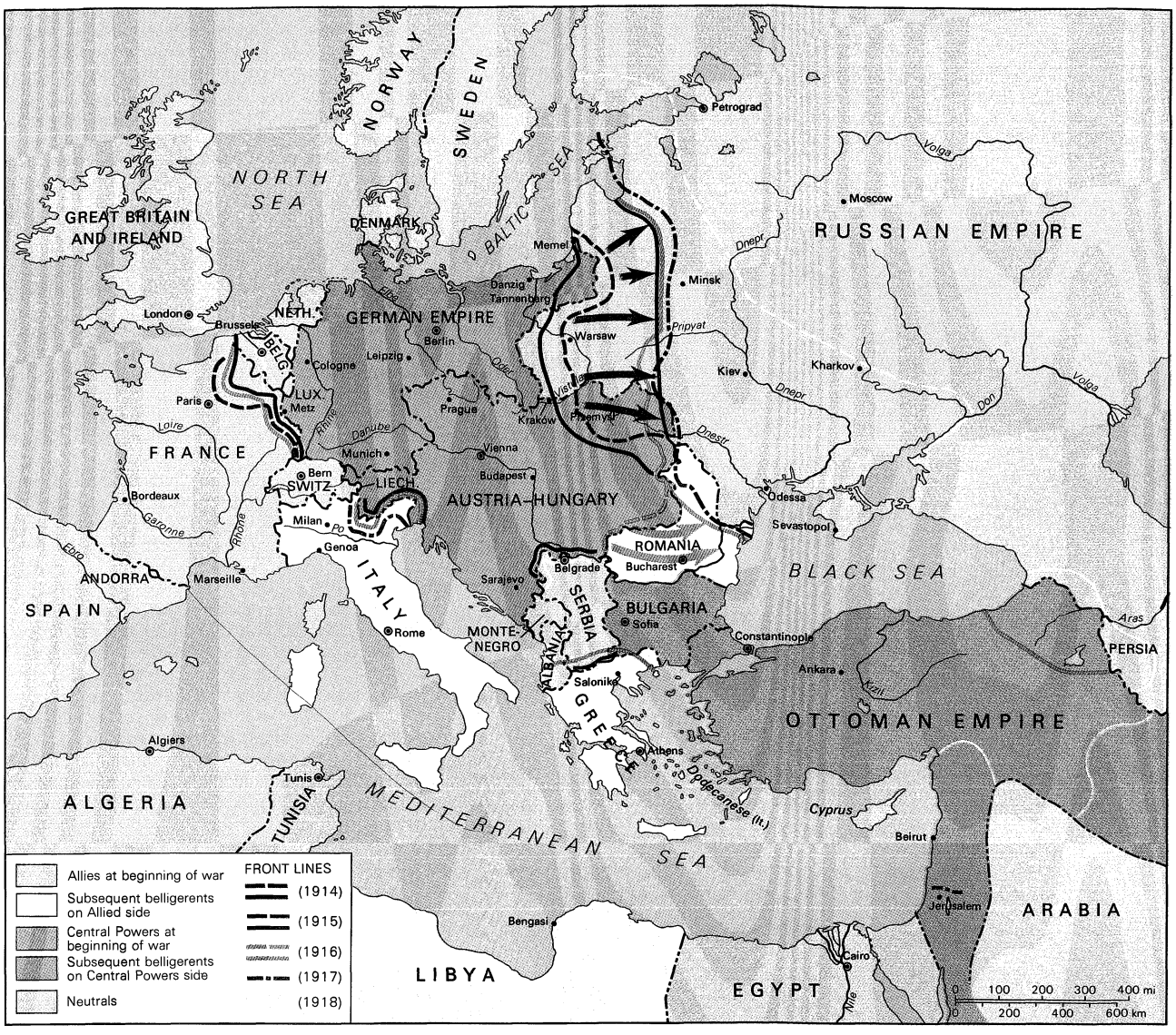
The Battle of Tannenberg

The progress of the battle was as follows. Samsonov, his forces spread out along a front 60 miles long, was gradually pushing Scholtz back toward the Allenstein–Osterode (Olsztyn–Ostróda) line when, on August 26, Ludendorff ordered General Hermann von François, with the I Corps on Scholtz's right, to attack Samsonov's left wing near Usdau (Uzdowo). There, on August 27, German artillery bombardments threw the hungry and weary Russians into precipitate flight. François started to pursue them toward Neidenburg, in the rear of the Russian centre, and then made a momentary diversion southward, to check a Russian counterattack from Soldau (Działdowo). Two of the Russian 2nd Army's six army corps managed to escape southeastward at this point, and François then resumed his pursuit to the east. By nightfall on August 29 his troops were in control of the road leading from Neidenburg eastward to Willenberg (Wielbark). The Russian centre, amounting to three army corps, was now caught in the maze of forest between Allenstein and the frontier of Russian Poland. It had no line of retreat, was surrounded by the Germans, and soon dissolved into mobs of hungry and exhausted men who beat feebly against the encircling German ring and then allowed themselves to be taken prisoner by the thousands. Samsonov shot himself in despair on August 29. By the end of August the Germans had taken 92,000 prisoners and annihilated half of the Russian 2nd Army. Ludendorff's bold recall of the last German forces facing Rennenkampf's army was wholly justified in the event, since Rennenkampf remained utterly passive while Samsonov's army was surrounded.

Having received two fresh army corps (seven divisions) from the Western Front, the Germans now turned on the slowly advancing 1st Army under Rennenkampf. The latter was attacked on a line extending from east of Königsberg to the southern end of the chain of the Masurian Lakes during September 1–15 and was driven from East Prussia. As a result of these East Prussian battles Russia had lost about 250,000 men and, what could be afforded still less, much war matériel. But the invasion of East Prussia had at least helped to make possible the French comeback on the Marne by causing the dispatch of two German army corps from the Western Front.

Having ended the Russian threat to East Prussia, the Germans could afford to switch the bulk of their forces from that area to the Częstochowa–Kraków front in southwestern Poland, where the Austrian offensive, launched on August 20, had been rolled back by Russian counterattacks. A new plan for simultaneous thrusts by the Germans toward Warsaw and by the Austrians toward Przemyśl was brought to nothing by the end of October, as the Russians could now mount counterattacks in overwhelming strength, their mobilization being at last nearly completed. The Russians then mounted a powerful effort to invade Prussian Silesia with a huge phalanx of seven armies. Allied hopes rose high as the much-heralded "Russian steamroller" (as the huge Russian Army was called) began its ponderous advance. The Russian armies were advancing toward Silesia when Hindenburg and Ludendorff, in November, exploited the superiority of the German railway network: when the retreating German forces had crossed the frontier back into Prussian Silesia, they were promptly moved northward into Prussian Poland and thence sent southeastward to drive a wedge between the two armies of the Russian right flank. The massive Russian operation against Silesia was disorganized, and within a week four new German army corps had arrived from the Western Front. Ludendorff was able to use them to press the Russians back by mid-December to the Bzura–Rawka (rivers) line in front of Warsaw, and the depletion of their munition supplies compelled the Russians to also fall back in Galicia to trench lines along the Nida and Dunajec rivers.

The Serbian campaign, 1914. The first Austrian invasion of Serbia was launched with numerical inferiority



Europe and the Mediterranean during World War I.

(part of one of the armies originally destined for the Balkan front having been diverted to the Eastern Front on August 18); and the able Serbian commander, Radomir Putnik, brought the invasion to an early end by his victories on the Cer Mountain (August 15–20) and at Šabac (August 21–24). In early September, however, Putnik's subsequent northward offensive on the Sava River, in the north, had to be broken off when the Austrians began a second offensive, against the Serbs' western front on the Drina River. After some weeks of deadlock, the Austrians began a third offensive, which had some success in the Battle of the Kolubara, and forced the Serbs to evacuate Belgrade on November 30; but by December 15 a Serbian counterattack had retaken Belgrade and forced the Austrians to retreat. Mud and exhaustion kept the Serbs from turning the Austrian retreat into a rout, but the victory sufficed to allow Serbia a long spell of freedom from further Austrian advances.

The Turkish entry. The entry of Turkey (or the Ottoman Empire, as it was then called) into the war as a German ally was the one great success of German wartime diplomacy. Since 1909 Turkey had been under the control of the Young Turks, over whom Germany had skillfully gained a dominating influence. German military instructors permeated the Turkish Army, and Enver Paşa, the leader of the Young Turks, saw alliance with Germany as the best way of serving Turkey's interests, in particular for protection against the Russian threat to the Straits. He therefore persuaded the grand vizier, Said Halim Paşa, to make a secret treaty (negotiated late in July, signed on August 2) pledging Turkey to the German side if Germany should have to take Austria-Hungary's side against Russia. The unforeseen entry of Great Britain into the war against Germany alarmed the Turks, but the timely arrival of two German warships, the *Goeben* and the *Breslau,* in the Dardanelles on August 10 turned the scales in favour of Enver's policy. The ships were ostensibly sold to Turkey, but they retained their German crews. The Turks began detaining British ships, and more anti-British provocations followed, both in the Straits and on the Egyptian frontier. Finally the *Goeben* led the Turkish fleet across the Black Sea to bombard Odessa and other Russian ports (October 29–30). Russia declared war against Turkey on November 1; and the western Allies, after an ineffective bombardment of the outer forts of the Dardanelles on November 3, declared war likewise on November 5. A British force from India occupied Basra, on the Persian Gulf, on November 21. In the winter of 1914–15 Turkish offensives in the Caucasus and in the Sinai Desert, albeit abortive, served German strategy well by tying Russian and British forces down in those peripheral areas.

The *Goeben* and the *Breslau* in the Dardanelles

The war at sea, 1914–15. In August 1914 Great Britain, with 29 capital ships ready and 13 under construction, and Germany, with 18 and nine, were the two great rival sea powers. Neither of them at first wanted a direct confrontation: the British were chiefly concerned with the protection of their trade routes; the Germans hoped that mines and submarine attacks would gradually destroy Great Britain's

numerical superiority, so that confrontation could eventually take place on equal terms.

The first significant encounter between the two navies was that of the Helgoland Bight, on Aug. 28, 1914, when a British force under Admiral Sir David Beatty, having entered German home waters, sank or damaged several German light cruisers and killed or captured 1,000 men at a cost of one British ship damaged and 35 deaths. For the following months the Germans in European or British waters confined themselves to submarine warfare—not without some notable successes: on September 22 a single German submarine, or U-boat, sank three British cruisers within an hour; on October 7 a U-boat made its way into the anchorage of Loch Ewe, on the west coast of Scotland; on October 15 the British cruiser *Hawke* was torpedoed; and on October 27 the British battleship *Audacious* was sunk by a mine.

On December 15 battle cruisers of the German High Seas Fleet set off on a sortie across the North Sea, under the command of Admiral Franz von Hipper: they bombarded several British towns and then made their way home safely. Hipper's next sortie, however, was intercepted on its way out: on Jan. 24, 1915, in the Battle of the Dogger Bank, the German cruiser *Blücher* was sunk and two other cruisers damaged before the Germans could make their escape.

Abroad on the high seas, the Germans' most powerful surface force was the East Asiatic squadron of fast cruisers, including the *Scharnhorst,* the *Gneisenau,* and the *Nürnberg,* under Admiral Graf Maximilian von Spee. For four months this fleet ranged almost unhindered over the Pacific Ocean, while the *Emden,* having joined the squadron in August 1914, was detached for service in the Indian Ocean. The Germans could thus threaten not only merchant shipping on the British trade routes but also troopships on their way to Europe or the Middle East from India, New Zealand, or Australia. The *Emden* sank merchant ships in the Bay of Bengal, bombarded Madras (September 22), haunted the approaches to Ceylon, and had destroyed 15 Allied ships in all before it was caught and sunk off the Cocos Islands on November 9 by the Australian cruiser *Sydney.*

The battles of Coronel and the Falkland Islands

Meanwhile, Admiral von Spee's main squadron since August had been threading a devious course in the Pacific from the Caroline Islands toward the Chilean coast and had been joined by two more cruisers, the *Leipzig* and the *Dresden.* On November 1, in the Battle of Coronel, it inflicted a sensational defeat on a British force, under Sir Christopher Cradock, which had sailed from the Atlantic to hunt it down: without losing a single ship, it sank Cradock's two major cruisers, Cradock himself being killed. But the fortunes of the war on the high seas were reversed when, on December 8, the German squadron attacked the Falkland Islands (in the South Atlantic), probably unaware of the naval strength that the British, since Coronel, had been concentrating there under Admiral Sir Doveton Sturdee: two battle cruisers (the *Invincible* and *Inflexible,* each equipped with eight 12-inch guns) and six other cruisers. The German ships were suffering from wear and tear after their long cruise in the Pacific and were no match for the newer, faster British ships, which soon overtook them. The *Scharnhorst,* with Admiral von Spee aboard, was the first ship to be sunk, then the *Gneisenau,* followed by the *Nürnberg* and the *Leipzig.* The British ships, which had fought at long range so as to render useless the smaller guns of the Germans, sustained only 25 casualties in this engagement. When the German light cruiser *Dresden* was caught and sunk off the Juan Fernández Islands on March 14, 1915, commerce raiding by German surface ships on the high seas was at an end. It was just beginning by German submarines, however.

The belligerent navies were employed as much in interfering with commerce as in fighting each other. Immediately after the outbreak of war, the British had instituted an economic blockade of Germany, with the aim of preventing all supplies reaching that country from the outside world. The two routes by which supplies could reach German ports were: (1) through the English Channel and the Dover Straits, and (2) around the north of Scotland. A minefield laid in the Dover Straits with a narrow free lane made it fairly easy to intercept and search ships using the Channel. To the north of Scotland, however, there was an area of more than 200,000 square miles (520,000 square kilometres) to be patrolled, and the task was assigned to a squadron of armed merchant cruisers. During the early months of the war, only absolute contraband such as guns and ammunition was restricted, but the list was gradually extended to include almost all material that might be of use to the enemy.

The prevention of the free passage of trading ships led to considerable difficulties among the neutral nations, particularly with the United States, whose trading interests were hampered by British policy. Nevertheless, the British blockade was extremely effective, and during 1915 the British patrols stopped and inspected more than 3,000 vessels, of which 743 were sent into port for examination. Outward-bound trade from Germany was brought to a complete standstill.

The Germans similarly sought to attack Great Britain's economy with a campaign against its supply lines of merchant shipping. In 1915, however, with their surface commerce raiders eliminated from the conflict, they were forced to rely entirely on the submarine.

The Germans began their submarine campaign against commerce by sinking a British merchant steamship (*Glitra*), after evacuating the crew, on Oct. 20, 1914. A number of other sinkings followed, and the Germans soon became convinced that the submarine would be able to bring the British to an early peace where the commerce raiders on the high seas had failed. On Jan. 30, 1915, Germany carried the campaign a stage further by torpedoing two Japanese liners (*Tokomaru* and *Ikaria*) without warning. They next announced, on February 4, that from February 18 they would treat the waters around the British Isles as a war zone in which all Allied merchant ships were to be destroyed, and in which no ship, whether enemy or not, would be immune.

Yet, whereas the Allied blockade was preventing almost all trade for Germany from reaching that nation's ports, the German submarine campaign yielded less satisfactory results. During the first week of the campaign seven Allied or Allied-bound ships were sunk out of 11 attacked, but 1,370 others sailed without being harassed by the German submarines. In the whole of March 1915, during which 6,000 sailings were recorded, only 21 ships were sunk, and in April only 23 ships from a similar number. Apart from its lack of positive success, the U-boat arm was continuously harried by Great Britain's extensive antisubmarine measures, which included nets, specially armed merchant ships, hydrophones for locating the noise of a submarine's engines, and depth bombs for destroying it underwater.

For the Germans, a worse result than any of the British countermeasures imposed on them was the long-term growth of hostility on the part of the neutral countries. Certainly the neutrals were far from happy with the British blockade, but the German declaration of the war zone and subsequent events turned them progressively away from their attitude of sympathy for Germany. The hardening of their outlook began in February 1915, when the Norwegian steamship *Belridge,* carrying oil from New Orleans to Amsterdam, was torpedoed and sunk in the English Channel. The Germans continued to sink neutral ships occasionally, and undecided countries soon began to adopt a hostile outlook toward this activity when the safety of their own shipping was threatened.

The *Lusitania* sunk

Much more serious was an action that confirmed the inability of the German command to perceive that a minor tactical success could constitute a strategic blunder of the most extreme magnitude. This was the sinking by a German submarine on May 7, 1915, of the British liner *Lusitania,* which was on its way from New York to Liverpool: though the ship was, in fact, carrying 173 tons of ammunition, it had nearly 2,000 civilian passengers, and the 1,198 people who were drowned included 128 U.S. citizens. The loss of the liner and so many of its passengers, including the Americans, aroused a wave of indignation in the United States, and it was fully expected that a declaration of war might follow. But the U.S. government

clung to its policy of neutrality and contented itself with sending several notes of protest to Germany. Despite this, the Germans persisted in their intention and, on August 17, sank the *Arabic,* which also had U.S. and other neutral passengers. Following a new U.S. protest, the Germans undertook to ensure the safety of passengers before sinking liners henceforth; but only after the torpedoing of yet another liner, the *Hesperia,* did Germany, on September 18, decide to suspend its submarine campaign in the English Channel and west of the British Isles, for fear of provoking the United States further. The German civilian statesmen had temporarily prevailed over the naval high command, which advocated "unrestricted" submarine warfare.

The loss of the German colonies. Germany's overseas colonies, virtually without hope of reinforcement from Europe, defended themselves with varying degrees of success against Allied attack.

Togoland was conquered by British forces from the Gold Coast (now Ghana) and by French forces from Dahomey (now Benin) in the first month of the war. In the Cameroons (German: Kamerun), invaded by Allied forces from the south, the east, and the northwest in August 1914 and attacked from the sea in the west, the Germans put up a more effective resistance, and the last German stronghold there, Mora, held out until Feb. 18, 1916.

Operations by South African forces in huge numerical superiority were launched against German South West Africa (Namibia) in September 1914 but were held up by the pro-German rebellion of certain South African officers who had fought against the British in the South African War of 1899–1902. The rebellion died out in February 1915, but the Germans in South West Africa nevertheless did not capitulate until July 9.

In Kiaochow, a small German enclave on the Chinese coast, the port of Tsingtao was the object of Japanese attack from September 1914. With some help from British troops and from Allied warships, the Japanese captured it on November 7. In October, meanwhile, the Japanese had occupied the Marianas, the Caroline Islands, and the Marshalls in the North Pacific, these islands being defenseless since the departure of Admiral von Spee's naval squadron.

In the South Pacific, Western Samoa fell without blood at the end of August 1914 to a New Zealand force supported by Australian, British, and French warships. In September an Australian invasion of Neu-Pommern (New Britain) won the surrender of the whole colony of German New Guinea within a few weeks.

Lettow-Vorbeck in East Africa

The story of German East Africa (comprising present-day Rwanda, Burundi, and continental Tanzania) was very different, thanks to the quality of the local askaris (European-trained African troops) and to the military genius of the German commander Paul von Lettow-Vorbeck. A landing of troops from India was repelled with ignominy by the Germans in November 1914. A massive invasion from the north, comprising British and colonial troops under the South African J.C. Smuts, was launched in February 1916, to be coordinated with a Belgian invasion from the west and with an independent British one from Nyasaland in the south; but though Dar es Salaam fell to Smuts and Tabora to the Belgians in September, Lettow-Vorbeck maintained his small force in being. In November 1917 he began to move southward across Portuguese East Africa (Germany had declared war on Portugal in March 1916); and, after crossing back into German East Africa in September 1918, he turned southwestward to invade Northern Rhodesia in October. Having taken Kasama on November 9 (two days before the German armistice in Europe), he finally surrendered on November 25. With some 12,000 men at the outset, he eventually tied down 130,000 or more Allied troops.

The years of stalemate

Rival strategies and the Dardanelles campaign, 1915–16. By late 1914 the state of deadlock on the Western Front had become clear to the governments of the warring countries and even to many members of their general staffs. Each side sought a solution to this deadlock, and the solutions varied in form and manner.

Erich von Falkenhayn had succeeded the dispirited Moltke as chief of the German general staff in September 1914. By the end of 1914 Falkenhayn seems to have concluded that although the final decision would be reached in the West, Germany had no immediate prospect of success there, and that the only practicable theatre of operations in the near future was the Eastern Front, however inconclusive those operations might be. Falkenhayn was convinced of the strength of the Allied trench barrier in France, so he took the momentous decision to stand on the defensive in the West.

Falkenhayn saw that a long war was now inevitable and set to work to develop Germany's resources for such a warfare of attrition. Thus, the technique of field entrenchment was carried to a higher pitch by the Germans than by any other country; Germany's military railways were expanded for the lateral movement of reserves; and the problem of the supply of munitions and of the raw materials for their manufacture was tackled so energetically and comprehensively that an ample flow was ensured from the spring of 1915 onward—a time when the British were only awakening to the problem. Here were laid the foundations of that economic organization and utilization of resources that was to be the secret of Germany's power to resist the pressure of the British blockade.

The western Allies were divided into two camps about strategy. Joffre and most of the French general staff, backed by the British field marshal Sir John French, argued for continuing assaults on the Germans' entrenched line in France, despite the continued attrition of French forces that this strategy entailed. Apart from this, the French high command was singularly lacking in ideas to break the deadlock of trench warfare. While desire to hold on to territorial gains governed the German strategy, the desire to recover lost territory dominated the French.

British-inspired solutions to the deadlock crystallized into two main groups, one tactical, the other strategical. The first was to unlock the trench barrier by inventing a machine that would be invulnerable to machine guns and capable of crossing trenches and would thus restore the tactical balance upset by the new preponderance of defensive over offensive power. The idea of such a machine was conceived by Colonel Ernest Swinton in October 1914, was nourished and tended in infancy by Winston Churchill, then first lord of the Admiralty, and ultimately, after months of experiment hampered by official opposition, came to maturity in 1916 in the weapon known as the tank. Some of the British strategists, on the other hand, argued that instead of seeking a breakthrough on the Germans' impregnable Western Front, the Allies should turn the whole position of the Central Powers either by an offensive through the Balkans or even by a landing on Germany's Baltic coast. Joffre and his supporters won the argument, and the Balkan projects were relinquished in favour of a concentration of effort on the Western Front. But misgivings were not silenced, and a situation arose that revived the Middle Eastern scheme in a new if attenuated form.

Early in January 1915 the Russians, threatened by the Turks in the Caucasus, appealed to the British for some relieving action against Turkey. The British, after acrimonious argument among themselves, decided in favour of "a naval expedition in February to bombard and take the Gallipoli Peninsula (the western shore of the Dardanelles), with Constantinople as its objective." Though subsequently it was agreed that army troops might be provided to hold the shores if the fleet forced the Straits, the naval attack began on February 19 without army support. When at last Sir Ian Hamilton's troops from Egypt began to land on the Turkish shores, on April 25, the Turks and their German commander, Otto Liman von Sanders, had had ample time to prepare adequate fortifications, and the defending armies were now six times as large as when the campaign opened.

Against resolute opposition from the local Turkish commander (Mustafa Kemal, the future Atatürk), Australian and New Zealand troops won a bridgehead at "Anzac Cove," north of Kaba Tepe, on the Aegean side of the peninsula, with some 20,000 men landing in the first two

days. The British, meanwhile, tried to land at five points around Cape Helles but established footholds only at three of them and then asked for reinforcements. Thereafter little progress was made, and the Turks took advantage of the British halt to bring into the peninsula as many troops as possible. The standstill of the enterprise led to a political crisis in London between Churchill, the Liberal government's first lord of the Admiralty, who, after earlier doubts, had made himself the foremost spokesman of the Dardanelles operation, and John, Lord Fisher, the first sea lord, who had always expressed doubts about it. Fisher demanded on May 14 that the operation be discontinued and, when he was overruled, resigned the next day. The Liberal government was replaced by a coalition, but Churchill, though relieved of his former post, remained in the War Council of the Cabinet.

The Churchill–Lord Fisher controversy

In July the British began sending five more divisions to the peninsula, and a new plan was hatched. In the hope of cutting the Turks' north–south communications down the peninsula by seizing the Sari Bair heights, which commanded the Straits from the west, the British reinforced the bridgehead at "Anzac Cove" and, in the night of August 6–7, landed more troops at Suvla Bay (Anafarta Limanı), farther to the north. Within a few days, both the offensive from "Anzac" and the new landing had proved ineffectual. More argument ensued in the War Council, and only late in the year was it acknowledged that the initially promising but ill-conducted enterprise should be given up. The evacuation of the troops was carried out from Suvla Bay and from "Anzac Cove" under cover of darkness in December 1915, and from the Cape Helles beaches in January 1916. The Dardanelles campaign thus came to a frustrating end. Had it succeeded it might well have ended Turkey's participation in the war. In failing, it had cost about 214,000 casualties and achieved nothing.

THE WESTERN AND EASTERN FRONTS, 1915

The Western Front, 1915. Repeated French attacks in February–March 1915 on the Germans' trench barrier in Champagne won only 500 yards (460 metres) of ground at a cost of 50,000 men. For the British, Sir Douglas Haig's 1st Army, between Armentières and Lens, tried a new experiment at Neuve-Chapelle on March 10, when its artillery opened an intense bombardment on a 2,000-yard front and then, after 35 minutes, lengthened its range, so that the attacking British infantry, behind the second screen of shells, could overrun the trenches ravaged by the first. But the experiment's immediate result was merely loss of life, both because shortage of munitions made the second barrage inadequate and because there was a five-hour delay in launching the infantry assault, against which the Germans, having overcome their initial surprise, had time to rally their resistance. It was clear to the Allies that this small-scale tactical experiment had missed success only by a narrow margin and that there was scope for its development. But the Allied commands missed the true lesson, which was that a surprise attack could be successfully made immediately following a short bombardment that compensated for its brevity by its intensity. Instead, they drew the superficial deduction that mere volume of shellfire was the key to reducing a trench line prior to an assault. Not until 1917 did they revert to the Neuve-Chapelle method. It was left to the Germans to profit from the experiment. In the meantime, a French offensive in April against the Germans' Saint-Mihiel salient, southeast of Verdun, sacrificed 64,000 men to no effect.

The Germans, in accordance with Falkenhayn's strategy, remained generally on the defensive in the West. They did, however, launch an attack on the Allies' Ypres salient (where the French had in November 1914 taken the place of the British). There, on April 22, 1915, they used chlorine gas for the first time on the Western Front, but they made the mistake of discharging it from cylinders (which were dependent on a favourable wind) rather than lobbing it onto the enemy trenches in artillery shells. The gas did throw the agonized defenders into chaotic flight; but the German high command, having been disappointed by the new weapon's performance under adverse conditions in Poland earlier in the year, had failed to provide adequate reserves to exploit its unforeseen success. By the end of a month-long battle, the Allies' front was only slightly retracted.

Gas warfare at Ypres

On May 9, meanwhile, the Allies had launched yet another premature offensive, combining a major French onslaught between Lens and Arras with two thrusts by Haig's 1st Army, from Festubert and from Fromelles, against the Aubers Ridge north of Lens. The French prolonged their effort until June 18, losing 102,000 men without securing any gain; the British, still short of shells against the Germans' mass of machine guns, had suspended their attacks three weeks earlier.

An even worse military failure was the joint offensive launched by the Allies on Sept. 25, 1915. While 27 French divisions with 850 heavy guns attacked on a front 18 miles long in Champagne, north and east of Reims, simultaneous blows were delivered in distant Artois by 14 French divisions with 420 heavy guns on a 12-mile front south of Lens and by six British divisions with only 117 guns at Loos north of Lens. All of these attacks were disappointing failures, partly because they were preceded by prolonged bombardments that gave away any chance of surprise and allowed time for German reserves to be sent forward to close up the gaps that had been opened in the trench defenders' ranks by the artillery bombardment. At Loos the British use of chlorine gas was less effective than Haig had hoped, and his engagement of all his own available forces for his first assault came to nothing when his commander in chief, Sir John French, was too slow in sending up reserves; the French on both their fronts likewise lost, through lack of timely support, most of what they had won by their first attacks. In all, for a little ground, the Allies paid 242,000 men, against the defenders' loss of 141,000.

Having subsequently complained bitterly about Sir John French's management of operations, Haig was appointed British commander in chief in his place in December.

The Eastern Front, 1915. The Russians' plans for 1915 prescribed the strengthening of their flanks in the north and in Galicia before driving westward again toward Silesia. Their preparations for a blow at East Prussia's southern frontier were forestalled, as Ludendorff, striking suddenly eastward from East Prussia, enveloped four Russian divisions in the Augustów forests, east of the Masurian Lakes, in the second week of February; but in Galicia the winter's fighting culminated, on March 22, in the fall of Przemyśl to the Russians.

For the Central Powers, the Austrian spokesman, Conrad, primarily required some action to relieve the pressure on his Galician front, and Falkenhayn was willing to help him for that purpose without departing from his own general strategy of attrition—which was already coming into conflict with Ludendorff's desire for a sustained effort toward decisive victory over Russia. The plan finally adopted, with the aim of smashing the Russian centre in the Dunajec River sector of Galicia by an attack on the 18-mile front from Gorlice to Tuchów (south of Tarnów), was conceived with tactical originality: in order to maintain the momentum of advance, no daily objectives were to be set for individual corps or divisions; instead, each should make all possible progress before the Russians could bring their reserves up, on the assumption that the rapid advance of some attacking units would contagiously promote the subsequent advance of others that had at first met more resistance. Late in April, 14 divisions, with 1,500 guns, were quietly concentrated for the stroke against the six Russian divisions present. Mackensen was in command, with Hans von Seeckt, sponsor of the new tactic of infiltration, as his chief of staff.

The Dunajec–San offensive

The Gorlice attack was launched on May 2 and achieved success beyond all expectation. Routed on the Dunajec, the Russians tried to stand on the Wisłoka, then fell back again. By May 14, Mackensen's forces were on the San, 80 miles from their starting point, and at Jarosław they even forced a crossing of that river. Strengthened with more German troops from France, Mackensen then struck again, taking Przemyśl on June 3 and Lemberg (Lvov) on June 22. The Russian front was now bisected, but Falkenhayn and Conrad had foreseen no such result and

had made no preparations to exploit it promptly. Their consequent delays enabled the Russian armies to retreat without breaking up entirely.

Falkenhayn then decided to pursue a new offensive. Mackensen was instructed to veer northward, so as to catch the Russian armies in the Warsaw salient between his forces and Hindenburg's, which were to drive southeastward from East Prussia. Ludendorff disliked the plan as being too much of a frontal assault: the Russians might be squeezed by the closing-in of the two wings, but their retreat to the east would not be cut off. He once more urged his spring scheme for a wide enveloping maneuver through Kovno (Kaunas) on Vilna (Vilnius) and Minsk, in the north. Falkenhayn opposed this plan, fearing that it would mean more troops and a deeper commitment, and on July 2 the German emperor decided in favour of Falkenhayn's plan.

The results justified Ludendorff's reservations. The Russians held Mackensen at Brest-Litovsk and Hindenburg on the Narew River long enough to enable the main body of their troops to escape through the unclosed gap to the east. Though by the end of August all of Poland had been occupied and 750,000 Russians had been taken prisoner in four months of fighting, the Central Powers had missed their opportunity to break Russia's ability to carry on the war.

Too late, Falkenhayn in September allowed Ludendorff to try what he had been urging much earlier, a wider enveloping movement to the north on the Kovno–Dvinsk–Vilna triangle. The German cavalry, in fact, approached the Minsk railway, far beyond Vilna; but the Russians' power of resistance was too great for Ludendorff's slender forces, whose supplies moreover began to run out, and by the end of the month his operations were suspended. The crux of this situation was that the Russian armies had been allowed to draw back almost out of the net before the long-delayed Vilna maneuver was attempted. Meanwhile, an Austrian attack eastward from Lutsk (Luck), begun later in September and continued into October, incurred heavy losses for no advantage at all. By October 1915 the Russian retreat, after a nerve-wracking series of escapes from the salients the Germans had systematically created and then sought to cut off, had come to a definite halt along a line running from the Baltic Sea just west of Riga southward to Czernowitz (Chernovtsy) on the Romanian border.

OTHER FRONTS, 1915–16

The Caucasus, 1914–16. The Caucasian front between Russia and Turkey comprised two battlegrounds: Armenia in the west, Azerbaijan in the east. While the ultimate strategic objectives for the Turks were to capture the Baku oilfields in Azerbaijan and to penetrate Central Asia and Afghanistan in order to threaten British India, they needed first to capture the Armenian fortress of Kars, which, together with that of Ardahan, had been a Russian possession since 1878.

A Russian advance from Sarıkamış (Sarykamysh, south of Kars) toward Erzurum in Turkish Armenia in November 1914 was countered in December when the Turkish 3rd Army, under Enver himself, launched a three-pronged offensive against the Kars–Ardahan position. This offensive was catastrophically defeated in battles at Sarıkamış and at Ardahan in January 1915; but the Turks, ill-clad and ill-supplied in the Caucasian winter, lost many more men through exposure and exhaustion than in fighting (their 3rd Army was reduced in one month from 190,000 to 12,400 men, the battle casualties being 30,000). Turkish forces, which had meanwhile invaded neutral Persia's part of Azerbaijan and taken Tabriz on January 14, were expelled by a Russian counterinvasion in March.

Sarıkamış and Ardahan operations

During this campaign the Armenians had created disturbances behind the Turkish lines in support of the Russians and had threatened the already arduous Turkish communications. The Turkish government on June 11, 1915, decided to deport the Armenians. In the process of deportation, the Turkish authorities committed atrocities on a vast scale: Armenian deaths have been estimated at more than 1,000,000. Subsequently, the Armenians perpetrated similar atrocities against the Turkish population of the Armenian country, but perforce on a smaller scale.

Grand Duke Nicholas, who had hitherto been commander in chief of all Russia's armies, was superseded by Emperor Nicholas himself in September 1915; the Grand Duke was then sent to command in the Caucasus. He and General N.N. Yudenich, the victor of Sarıkamış, started a major assault on Turkish Armenia in January 1916; Erzurum was taken on February 16, Trabzon on April 18, Erzıncan on August 2; and a long-delayed Turkish counterattack was held at Oğnut. Stabilized to Russia's great advantage in the autumn, the new front in Armenia was thereafter affected less by Russo-Turkish warfare than by the consequences of revolution in Russia.

Mesopotamia, 1914–April 1916. The British occupation of Basra, Turkey's port at the head of the Persian Gulf, in November 1914 had been justifiable strategically because of the need to protect the oil wells of southern Persia and the Abadan refinery. The British advance of 46 miles northward from Basra to al-Qurnah in December and the further advance of 90 miles up the Tigris to al-ʿAmārah in May–June 1915 ought to have been reckoned enough for all practical purposes; but the advance was continued, in the direction of the fatally magnetic Baghdad, ancient capital of the Arab caliphs of Islām. Al-Kūt was occupied in September 1915, and the advance was pushed on until the British, under Major General Charles Townshend, were 500 miles away from their base at Basra. They fought a profitless battle at Ctesiphon, only 18 miles from Baghdad, on November 22 but then had to retreat to al-Kūt. There, from December 7, Townshend's 10,000 men were besieged by the Turks; and there, on April 29, 1916, they surrendered themselves into captivity.

The Egyptian frontiers, 1915–July 1917. Even after the evacuation from Gallipoli, the British maintained 250,000 troops in Egypt. A major source of worry to the British was the danger of a Turkish threat from Palestine across the Sinai Desert to the Suez Canal. That danger waned, however, when the initially unpromising rebellion of the Hāshimite amir Ḥusayn ibn ʿAlī against the Turks in the Hejaz was developed by the personal enterprise of an unprofessional soldier of genius, T.E. Lawrence, into a revolt infecting the whole Arabian hinterland of Palestine and Syria and threatening to sever the Turks' vital Hejaz Railway (Damascus–Amman–Maʿān–Medina). Sir Archibald Murray's British troops at last started a massive advance in December 1916 and captured some Turkish outposts on the northeastern edge of the Sinai Desert but made a pusillanimous withdrawal from Gaza in March 1917 at the very moment when the Turks were about to surrender the place to them; the attempt the next month to retrieve the mistake was repulsed with heavy losses. In June the command was transferred from Murray to Sir Edmund Allenby. In striking contrast to Murray's performance was Lawrence's capture of Aqaba (al-ʿAqabah) on July 6, 1917: his handful of Arabs got the better of 1,200 Turks there.

The revolt of the Hejaz

Italy and the Italian front, 1915–16. Great Britain, France, and Russia concluded on April 26, 1915, the secret Treaty of London with Italy, inducing the latter to discard the obligations of the Triple Alliance and to enter the war on the side of the Allies by the promise of territorial aggrandizement at Austria-Hungary's expense. Italy was offered not only the Italian-populated Trentino and Trieste but also South Tirol (to consolidate the Alpine frontier), Gorizia, Istria, and northern Dalmatia. On May 23, 1915, Italy accordingly declared war on Austria-Hungary.

The Italian commander, General Luigi Cadorna, decided to concentrate his effort on an offensive eastward from the province of Venetia across the comparatively low ground between the head of the Adriatic and the foothills of the Julian Alps; that is to say, across the lower valley of the Isonzo (Soča) River. Against the risk of an Austrian descent on his rear from the Trentino (which bordered Venetia to the northwest) or on his left flank from the Carnic Alps (to the north), he thought that limited advances would be precaution enough.

The Italians' initial advance eastward, begun late in May 1915, was soon halted, largely because of the flooding of the Isonzo, and trench warfare set in. Cadorna, however,

was determined to make progress and so embarked on a series of persistent renewals of the offensive, known as the Battles of the Isonzo. The first four of these (June 23–July 7; July 18–August 3; October 18–November 4; and November 10–December 2) achieved nothing worth the cost of 280,000 men; and the fifth (March 1916) was equally fruitless. The Austrians had shown on this front a fierce resolution that was often lacking when they faced the Russians. In mid-May 1916 Cadorna's program was interrupted by an Austrian offensive from the Trentino into the Asiago region of western Venetia. Though the danger of an Austrian breakthrough from the mountainous borderland into the Venetian plain in the rear of the Italians' Isonzo front was averted, the Italian counteroffensive in mid-June recovered only one-third of the territory overrun by the Austrians north and southwest of Asiago. The Sixth Battle of the Isonzo (August 6–17), however, did win Gorizia for the Italians. On August 28 Italy declared war on Germany. The next three months saw three more Italian offensives on the Isonzo, none of them really profitable. In the course of 1916 the Italians had sustained 500,000 casualties, twice as many as the Austrians, and were still on the Isonzo.

Serbia and the Salonika expedition, 1915–17. Austria's three attempted invasions of Serbia in 1914 had been brusquely repulsed by Serbian counterattacks. By the summer of 1915 the Central Powers were doubly concerned to close the account with Serbia, both for reasons of prestige and for the sake of establishing secure rail communications with Turkey across the Balkans. In August, Germany sent reinforcements to Austria's southern front; and, on Sept. 6, 1915, the Central Powers concluded a treaty with Bulgaria, whom they drew to their side by the offer of territory to be taken from Serbia. The Austro-German forces attacked southward from the Danube on October 6; and the Bulgars, undeterred by a Russian ultimatum, struck at eastern Serbia on October 11 and at Serbian Macedonia on October 14.

Bulgaria's entry into the war

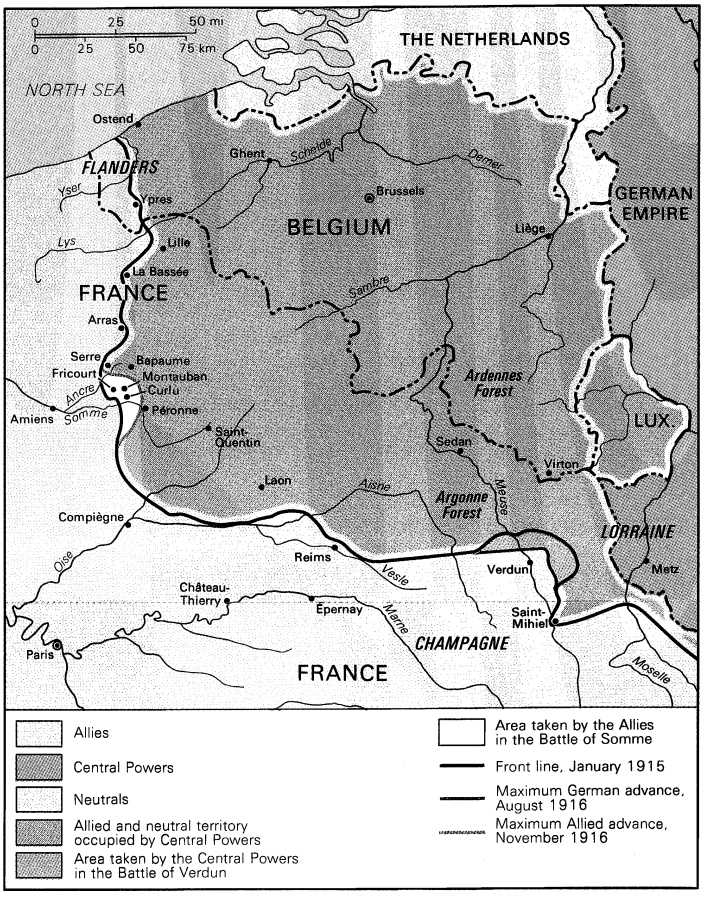
The western Allies, surprised in September by the prospect of a Bulgarian attack on Serbia, hastily decided to send help through neutral Greece's Macedonian port of Salonika, relying on the collusion of Greece's pro-Entente prime minister, Eleuthérios Venizélos. Troops from Gallipoli, under the French general Maurice Sarrail, reached Salonika on October 5, but on that day Venizélos fell from power. The Allies advanced northward up the Vardar into Serbian Macedonia but found themselves prevented from junction with the Serbs by the westward thrust of the Bulgars. Driven back over the Greek frontier, the Allies were merely occupying the Salonika region by mid-December. The Serbian Army, meanwhile, to avoid double envelopment, had begun an arduous winter retreat westward over the Albanian mountains to refuge on the island of Corfu.

In the spring of 1916 the Allies at Salonika were reinforced by the revived Serbs from Corfu as well as by French, British, and some Russian troops, and the bridgehead was expanded westward to Vodena (Edessa) and eastward to Kilkis; but the Bulgars, who in May obtained Fort Rupel (Klidhi, on the Struma) from the Greeks, in mid-August not only overran Greek Macedonia east of the Struma but also, from Monastir (Bitola), invaded the Florina region of Greek Macedonia, to the west of the Allies' Vodena wing. The Allied counteroffensive took Monastir from the Bulgars in November 1916, but more ambitious operations, from March to May 1917, proved abortive. The Salonika front was tying down some 500,000 Allied troops without troubling the Central Powers in any significant way.

MAJOR DEVELOPMENTS IN 1916

The Western Front, 1916. In 1914 the centre of gravity of World War I had been on the Western Front, in 1915 it shifted to the Eastern, and in 1916 it once more moved back to France. Though the western Allies had dissipated some of their strength in the Dardanelles, Salonika, and Mesopotamia, the rising tide of Britain's new armies and of its increased munition supplies promised the means for an offensive far larger in scale than any before to break the trench deadlock. Britain's armies in France had grown to 36 divisions by the end of 1915. By that time voluntary enlistments, though massive, had nevertheless proved to be inadequate to meet Britain's needs, so in January 1916, by the Military Service Act, voluntary service was replaced by conscription.

In December 1915 a conference of the leaders of the French, British, Belgian, and Italian armies, with representatives present from the Russian and Japanese armies, was held at Joffre's headquarters. They adopted the principle of a simultaneous general offensive in 1916 by France, Great Britain, Russia, and Italy. But military action by Germany was to dislocate this scheme, and only the British offensive came fully into operation.



Major campaigns on the Western Front, 1916.

By the winter of 1915–16, Falkenhayn regarded Russia as paralyzed and Italy as inconsiderable. He considered the time at last ripe for positive action against France, after whose collapse Great Britain would have no effective military ally on the European continent and would be brought to terms rather by submarine warfare than by land operations. For his offensive in the West, however, Falkenhayn clung always to his method of attrition. He believed that a mass breakthrough was unnecessary and that instead the Germans should aim to bleed France of its manpower by choosing a point of attack "for the retention of which the French Command would be compelled to throw in every man they have." The town of Verdun and its surrounding complex of forts was chosen, because it was a menace to the main German lines of communications, because it was within a French salient and thus cramped the defenders, and because of the certainty that the French would sacrifice any number of men to defend Verdun for reasons of patriotism associated with the town itself.

The battle for Verdun

The keynote of Falkenhayn's tactical plan was to place a dense semicircle of German heavy and medium artillery to the north and east of Verdun and its outlying fortresses and then to stage a continuous series of limited infantry advances upon the forts. These advances would draw the French infantry into defending or trying to retake the forts, in the process of which they would be pulverized by

German artillery fire. In addition, each German infantry advance would have its way smoothed by a brief but extremely intense artillery bombardment that would clear the targeted ground of defenders.

Although French Intelligence had given early warnings of the Germans' offensive preparations, the French high command was so preoccupied with its own projected offensive scheme that the warning fell on deaf ears. At 7:15 AM on Feb. 21, 1916, the heaviest German artillery bombardment yet seen in the war began on a front of eight miles around Verdun, and the French trenches and barbed wire fields there were flattened out or upheaved in a chaos of tumbled earth. At 4:45 PM the German infantry advanced—although for the first day only on a front of two and a half miles. From then until February 24 the French defenders' lines east of the Meuse River crumbled away. Fort-Douaumont, one of the most important fortresses, was occupied by the Germans on February 25. By March 6, when the Germans began to attack on the west bank of the Meuse as well as on the east bank, the French had come to see that something more than a feint was intended. To relieve the pressure on France, the Russians made a sacrificial attack on the Eastern Front at Lake Naroch (see below); the Italians began their fifth offensive on the Isonzo (see above); and the British took over the Arras sector of the Western Front, thus becoming responsible for the whole line from the Yser southward to the Somme. Meanwhile, General Philippe Pétain was entrusted with commanding the defense of Verdun. He organized repeated counterattacks that slowed the German advance, and, more importantly, he worked to keep open the one road leading into Verdun that had not been closed by German shelling. This was the Bar-le-Duc road, which became known as La Voie Sacrée (the "Sacred Way") because vital supplies and reinforcements continued to be sent to the Verdun front along it despite constant harassment from the German artillery.

Slowly but steadily the Germans moved forward on Verdun: they took Fort-Vaux, southeast of Fort-Douaumont, on June 7 and almost reached the Belleville heights, the last stronghold before Verdun itself, on June 23. Pétain was preparing to evacuate the east bank of the Meuse when the Allies' offensive on the Somme River was at last launched. Thereafter, the Germans assigned no more divisions to the Verdun attack.

Preceded by a week's bombardment, which gave ample warning of its advent, the Somme offensive was begun on July 1, 1916, by the 11 British divisions of Rawlinson's new 4th Army on a 15-mile front between Serre, north of the Ancre, and Curlu, north of the Somme, while five French divisions attacked at the same time on an eight-mile front mainly south of the Somme, between Curlu and Péronne. With incredibly misplaced optimism, Haig had convinced himself that the British infantry would be able to walk forward irresistibly over ground cleared of defenders by the artillery. But the unconcealed preparations for the assault and the long preliminary bombardment had given away any chance of surprise, and the German defenders were well prepared for what was to come. In the event, the 60,000 attacking British infantrymen moving forward in symmetrical alignment at a snail's pace enforced by each man's 66 pounds (30 kilograms) of cumbrous equipment were mowed down in masses by the German machine guns, and the day's casualties were the heaviest ever sustained by a British army. The French participants in the attack had twice as many guns as the British and did better against a weaker system of defenses, but almost nothing could be done to exploit this comparative success.

The First Battle of the Somme

Resigning himself now to limited advances, Haig concentrated his next effort on the southern sector of his Somme front. The Germans' second position there (Longueval, Bazentin, and Ovillers) fell on July 14, but again the opportunity of exploitation was missed. Thenceforward, at great cost in lives, a methodical advance was continued, gaining little ground but straining the German resistance. The first tanks to be used in the war, though in numbers far too small to be effective, were thrown into the battle by the British on September 15. In mid-November early rains halted operations. The four-month Battle of the Somme was a miserable failure except that it diverted German resources from the attack on Verdun. It cost the British 420,000 casualties, the French 195,000, and the Germans 650,000.

At Verdun, the summer slackening of German pressure enabled the French to organize counterattacks. Surprise attacks directed by General Robert-Georges Nivelle and launched by General Charles Mangin's army corps recovered Fort-Douaumont on October 24, Fort-Vaux on November 2, and places north of Douaumont in mid-December. Pétain's adroit defense of Verdun and these counterattacks had deprived Falkenhayn's offensive of its strategic fulfillment; but France had been so much weakened in the first half of 1916 that it could scarcely satisfy the Allies' expectations in the second. Verdun was one of the longest, bloodiest, and most ferocious battles of the war; French casualties amounted to about 400,000, German ones to about 350,000.

The Battle of Jutland. The summer of 1916 saw the long-deferred confrontation of Germany's High Seas Fleet and Great Britain's Grand Fleet in the Battle of Jutland—history's biggest naval battle, which both sides claimed as a victory.

Admiral Reinhard Scheer, who became commander in chief of the High Seas Fleet in January 1916, planned to contrive an encounter on the open sea between his fleet and some part of the British fleet in separation from the whole, so that the Germans could exploit their momentary superiority in numbers to achieve victory. Scheer's plan was to ensnare Admiral Beatty's squadron of battle cruisers at Rosyth, midway up Britain's eastern coast, by stratagem and destroy it before any reinforcements from the Grand Fleet's main base at Scapa Flow could reach it.

The German Trap

To set the trap, five battle cruisers of the German High Seas Fleet, together with four light cruisers, were to sail northward, under Hipper's command, from Wilhelmshaven, Ger., to a point off the southwestern coast of Norway. Scheer himself, with the battle squadrons of the High Seas Fleet, was to follow, 50 miles behind, to catch Beatty's forces in the gap once they had been lured eastward across the North Sea in pursuit of Hipper. But the signal for the German operation to begin, made in the afternoon of May 30, was intercepted and partially decoded by the British; and before midnight the whole British Grand Fleet was on its way to a rendezvous off Norway's southwestern coast and roughly across the planned route of the German fleet.

At 2:20 PM on May 31, when Admiral John Jellicoe's Grand Fleet squadrons from Scapa Flow were still 65 miles away to the north, Beatty's advance guard of light cruisers—five miles ahead of his heavier ships—and Hipper's scouting group learned quite accidentally of one another's proximity. An hour later the two lines were drawn up for battle, and in the next 50 minutes the British suffered severely, and the *Indefatigable* was sunk. When Beatty's battle cruisers came up, however, the German cruisers, in their turn, sustained such damage that Hipper sent a protective screen of German destroyers in to launch a torpedo attack. The British had lost another battle cruiser, the *Queen Mary,* before the German High Seas Fleet was sighted by a British patrol to the south, at 4:35 PM. On this report Beatty ordered his ships northward, to lure the Germans toward the Grand Fleet under Jellicoe's command.

Not until 6:14 PM, after Jellicoe's squadrons and Beatty's had been within sight of one another for nearly a quarter of an hour, was the German fleet precisely located—only just in time for Jellicoe to deploy his ships to the best advantage. Jellicoe arrayed the Grand Fleet end-to-end in a line so that their combined broadsides could be brought to bear on the approaching German ships, who could in turn reply only with the forward guns of their leading ships. The British ships in effect formed the horizontal stroke and the German ships the vertical stroke of the letter "T," with the British having deployed into line at a right angle to the German ships' forward progress. This maneuver was in fact known as "crossing the enemy's T" and was the ideal situation dreamed of by the tacticians of both navies, since by "crossing the T" one's forces temporarily gained an overwhelming superiority of firepower.

For the Germans this was a moment of unparalleled risk. Three factors helped prevent the destruction of the German ships in this trap: their own excellent construction, the steadiness and discipline of their crews, and the poor quality of the British shells. The *Lützow,* the *Derfflinger,* and the battleship *König* led the line and were under broadside fire from some 10 British battleships, yet their main guns remained undamaged and they fought back to such effect that one of their salvoes fell full on the *Invincible* and blew it up. This success, however, did little to relieve the intense bombardment from the other British ships, and the German fleet was still pressing forward into the steel trap of the Grand Fleet.

Relying on the magnificent seamanship of the German crews, Scheer extricated his fleet from the appalling danger into which it had run by a simple but, in practice, extremely difficult maneuver. At 6:30 PM he ordered a turn of 180° for all his ships at once; it was executed without collision; and the German battleships reversed course in unison and steamed out of the jaws of the trap, while German destroyers spread a smoke screen across their rear. The smoke and worsening visibility left Jellicoe in doubt about what had happened, and the British had lost contact with the Germans by 6:45 PM.

Yet the British Grand Fleet had maneuvered in such a way that it ended up between the German High Seas Fleet and the German ports, and this was the situation Scheer most dreaded, so at 6:55 PM Scheer ordered another reverse turn, perhaps hoping to pass around the rear of the British fleet. But the result for him was a worse position than that from which he had just escaped: his battle line had become compressed, and his leading ships found themselves again under intense bombardment from the broadside array of the British ships. Jellicoe had succeeded in crossing the Germans' "T" again. The *Lützow* now received irreparable damage, and many other German ships were damaged at this point. At 7:15 PM, therefore, to cause a diversion and win time, Scheer ordered his battle cruisers and destroyers ahead to virtually immolate themselves in a massed charge against the British ships.

This was the crisis of the Battle of Jutland. As the German battle cruisers and destroyers steamed forward, the German battleships astern became confused and disorganized in trying to execute their reverse turn. Had Jellicoe ordered the Grand Fleet forward through the screen of charging German battle cruisers at that moment, the fate of the German High Seas Fleet would likely have been sealed. As it was, fearing and overestimating the danger of torpedo attacks from the approaching destroyers, he ordered his fleet to turn away, and the two lines of battleships steamed apart at a speed of more than 20 knots. They did not meet again, and when darkness fell, Jellicoe could not be sure of the route of the German retreat. By 3:00 AM on June 1 the Germans had safely eluded their pursuers.

The British had sustained greater losses than the Germans in both ships and men. In all, the British lost three battle cruisers, three cruisers, eight destroyers, and 6,274 officers and men in the Battle of Jutland. The Germans lost one battleship, one battle cruiser, four light cruisers, five destroyers, and 2,545 officers and men. The losses inflicted on the British, however, were not enough to affect the numerical superiority of their fleet over the German in the North Sea, where their domination remained practically unchallengeable during the course of the war. Henceforth, the German High Seas Fleet chose not to venture out from the safety of its home ports.

The Eastern Front, 1916. In the hope of diverting German strength from the attack at Verdun on the Western Front, the Russians gallantly but prematurely opened an offensive north and south of Lake Naroch (Narocz, east of Vilna) on March 18, 1916, and continued it until March 27, though they won very little ground at great cost and only for a short time. They then reverted to preparations for a major offensive in July. The main blow, it was planned, should be delivered by A.E. Evert's central group of armies, assisted by an inward movement of A.N. Kuropatkin's army in the northern sector of the front. But at the same time, A.A. Brusilov's southwestern army group was authorized to make a supposedly diversionary attack in its own sectors. In the event, Brusilov's attack became by far the more important operation of the offensive.

Brusilov's breakthrough

Surprised by the Austrians' Asiago offensive in May, Italy promptly appealed to the Russians for action to draw the enemy's reserves away from the Italian fronts, and the Russians responded by advancing their timetable again. Brusilov undertook to start his attack on June 4, on the understanding that Evert's should be launched 10 days later.

Thus began an offensive on the Eastern Front that was to be imperial Russia's last really effective military effort. Popularly known as Brusilov's offensive, it had such an astonishing initial success as to revive Allied dreams about the irresistible Russian "steamroller." Instead, its ultimate achievement was to sound the death knell of the Russian monarchy. Brusilov's four armies were distributed along a very wide front, with Lutsk at the northern end, Tarnopol and Buchach (Buczacz) in the central sector, and Czernowitz at the southern end. Having struck first in the Tarnopol and Czernowitz sectors on June 4, Brusilov on June 5 took the Austrians wholly by surprise when he launched A.M. Kaledin's army toward Lutsk: the defenses crumbled at once, and the attackers pushed their way between two Austrian armies. As the offensive was developed, the Russians were equally successful in the Buchach sector and in their thrust into Bukovina, which culminated in the capture of Czernowitz. By June 20, Brusilov's forces had captured 200,000 prisoners.

Evert and Kuropatkin, however, instead of striking in accordance with the agreed plan, found excuses for procrastination. The Russian chief of general staff, M.V. Alekseyev, therefore tried to transfer this inert couple's reserves to Brusilov, but the Russians' lateral communications were so poor that the Germans had time to reinforce the Austrians before Brusilov was strong enough to make the most of his victory. Though his forces in Bukovina advanced as far as the Carpathian Mountains, a counterstroke by Alexander von Linsingen's Germans in the Lutsk sector checked Russian progress at the decisive point. Further Russian drives from the centre of Brusilov's front were launched in July; but by early September the opportunity of exploiting the summer's victory was lost. Brusilov had driven the Austrians from Bukovina and from much of eastern Galicia and had inflicted huge losses of men and equipment on them, but he had depleted Russia's armies by about 1,000,000 men in doing so. (A large portion of this number consisted of deserters or prisoners.) This loss seriously undermined both the morale and the material strength of Russia. Brusilov's offensive also had indirect results of great consequence. First, it had compelled the Germans to withdraw at least seven divisions from the Western Front, where they could ill be spared from the Verdun and Somme battles. Second, it hastened Romania's unfortunate entry into the war.

Disregarding Romania's military backwardness, the Romanian government of Ionel Brătianu declared war against Austria-Hungary on Aug. 27, 1916. In entering the war, Romania succumbed to the Allies' offers of Austro-Hungarian territory and to the belief that the Central Powers would be too much preoccupied with other fronts to mount any serious riposte against a Romanian offensive. Some 12 of Romania's 23 divisions, in three columns, thus began on August 28 a slow westward advance across Transylvania, where at first there were only five Austro-Hungarian divisions to oppose them.

Romania's entry into the war

The riposte of the Central Powers was swifter than the progress of the invasion: Germany, Turkey, and Bulgaria declared war against Romania on August 28, August 30, and September 1, respectively; and Falkenhayn had plans already prepared. Though the miscarriage of his overall program for the year led to his being replaced by Hindenburg as chief of the German general staff on August 29, Falkenhayn's recommendation that Mackensen should direct a Bulgarian attack on southern Romania was approved; and Falkenhayn himself went to command on the Transylvanian front, for which five German as well as two more Austrian divisions were found available as reinforcements.

Mackensen's forces from Bulgaria stormed the Turtucaia (Tutrakan) bridgehead on the Danube southeast of Bucharest on September 5. His subsequent advance eastward into the Dobruja caused the Romanians to switch their reserves to that quarter instead of reinforcing their Transylvanian enterprise, which thereupon came to a halt. Falkenhayn soon attacked: first at the southern end of the 200-mile front, where he threw one of the Romanian columns back into the Roter Turm (Turnu Roşu) Pass, then in the centre, where by October 9 he had defeated another at Kronstadt (Braşov). For a month, however, the Romanians withstood Falkenhayn's attempts to drive them out of the Vulcan and Szurduk (Surduc) passes into Walachia. But just before winter snows blocked the way, the Germans took the two passes and advanced southward to Tîrgu Jiu, where they won another victory. Then Mackensen, having turned westward from the Dobruja, crossed the Danube near Bucharest, on which his and Falkenhayn's armies converged. Bucharest fell on December 6, and the Romanian Army, a crippled force, could only fall back northeastward into Moldavia, where it had the belated support of Russian troops. The Central Powers had access to Romania's wheat fields and oil wells, and the Russians had 300 more miles of front to defend.

German strategy and the submarine war, 1916–January 1917. Both Admiral Scheer and General Falkenhayn doubted whether the German submarines could do any decisive damage to Great Britain so long as their warfare was restricted in deference to the protests of the United States; and, after a tentative reopening of the submarine campaign on Feb. 4, 1916, the German naval authorities in March gave the U-boats permission to sink without warning all ships except passenger vessels. The German civilian statesmen, however, who paid due attention to their diplomats' warnings about U.S. opinion, were soon able to prevail over the generals and the admirals: on May 4 the scope of the submarine campaign was again severely restricted.

The controversy between the statesmen and the advocates of unrestricted warfare was not dead yet. Hindenburg, chief of the general staff from August 29, had Ludendorff as his quartermaster general, and Ludendorff was quickly won over to supporting the chief of the Admiralty staff, Henning von Holtzendorff, in his arguments against the German chancellor, Theobald von Bethmann Hollweg, and the foreign minister, Gottlieb von Jagow. Whereas Bethmann and some other statesmen were hoping for a negotiated peace (see below), Hindenburg and Ludendorff were committed to a military victory. The British naval blockade, however, threatened to starve Germany into collapse before a military victory could be achieved, and soon Hindenburg and Ludendorff got their way: it was decided that, from Feb. 1, 1917, submarine warfare should be unrestricted and overtly so.

Declaration of unrestricted submarine warfare

Peace moves and U.S. policy to February 1917. There were few efforts by any of the Central or Allied Powers to achieve a negotiated peace in the first two years of the war. By 1916 the most promising signs for peace seemed to exist only in the intentions of two statesmen in power—the German chancellor Bethmann and the U.S. president Woodrow Wilson. Wilson, having proclaimed the neutrality of the United States in August 1914, strove for the next two years to maintain it. Early in 1916 he sent his confidant, Colonel Edward M. House, to sound London and Paris about the possibility of U.S. mediation between the belligerents. House's conversations with the British foreign secretary, Sir Edward Grey, resulted in the House–Grey Memorandum (Feb. 22, 1916), declaring that the United States might enter the war if Germany rejected Wilson's mediation but that Great Britain reserved the right to initiate U.S. mediatory action. By mid-1916, the imminent approach of the presidential election in the United States caused Wilson to suspend his moves for peace.

The Wilson–House overtures

In Germany, meanwhile, Bethmann had succeeded, with difficulty, in postponing the declaration of unrestricted submarine warfare. Wilson, though he was reelected president on Nov. 7, 1916, let another month pass without doing anything for peace, and during that period the German victory over Romania was taking place. Thus, while Bethmann lost patience with waiting for Wilson to act, the German military leaders came momentarily to think that Germany, from a position of strength, might now propose a peace acceptable to themselves. Having been constrained to agree with the militarists that, if his proposals were rejected by the Allies, unrestricted submarine warfare should be resumed, Bethmann was allowed to announce, on December 12, the terms of a German offer of peace—terms, however, that were militarily so far-reaching as to preclude the Allies' acceptance of them. The main stumbling block was Germany's insistence upon its annexation of Belgium and of the occupied portion of northeastern France.

On Dec. 18, 1916, Wilson invited both belligerent camps to state their "war aims." The Allies were secretly encouraged by the U.S. secretary of state to offer terms too sweeping for German acceptance; and the Germans, suspecting collusion between Wilson and the Allies, agreed in principle to the opening of negotiations but left their statement of December 12 practically unchanged and privately decided that Wilson should not actually take part in any negotiation that he might bring about. By mid-January 1917 the December overtures had ended.

Strangely enough, Wilson's next appeal, a speech of Jan. 22, 1917, preaching international conciliation and a "peace without victory," elicited a confidential response from the British expressing readiness to accept his mediation. In the opposite camp, Austria-Hungary would likewise have listened readily to peace proposals, but Germany had already decided, on January 9, to declare unrestricted submarine warfare. Bethmann's message restating Germany's peace terms and inviting Wilson to persevere in his efforts was delivered on January 31 but was paradoxically accompanied by the announcement that unrestricted submarine warfare would begin the next day.

Wilson severed diplomatic relations between the United States and Germany on Feb. 3, 1917, and asked Congress, on February 26, for power to arm merchantmen and to take all other measures to protect U.S. commerce. But American opinion was still not ready for war, and the Germans wisely abstained from attacks on U.S. shipping. What changed the tenor of public feeling was the publication of the Zimmermann Telegram.

The Zimmermann Telegram

Arthur Zimmermann had succeeded Jagow as Germany's secretary of state for foreign affairs in November 1916; and in that same month the Mexican president, Venustiano Carranza, whose country's relations with the United States had been critical since March, had virtually offered bases on the Mexican coast to the Germans for their submarines. Zimmermann on Jan. 16, 1917, sent a coded telegram to his ambassador in Mexico instructing him to propose to the Mexican government that, if the United States should enter the war against Germany, Mexico should become Germany's ally with a view to recovering Texas, New Mexico, and Arizona from the United States. Intercepted and decoded by the British Admiralty Intelligence, this message was communicated to Wilson on February 24. It was published in the U.S. press on March 1, and it immediately set off a nationwide demand for war against Germany.

DEVELOPMENTS IN 1917

The Western Front, January–May 1917. The western Allies had good reason to be profoundly dissatisfied with the poor results of their enterprises of 1916, and this dissatisfaction was signalized by two major changes made at the end of the year. In Great Britain, the government of H.H. Asquith, already turned into a coalition in May 1915, was replaced in December 1916 by a coalition under David Lloyd George; and that same month in France the post of commander in chief of the army was transferred from Joffre to General R.-G. Nivelle.

As for the military situation, the fighting strength of the British Army on the Western Front had grown to about 1,200,000 men and was still growing. That of the French Army had been increased by the incorporation of colonial troops to some 2,600,000, so that, including the Belgians, the Allies disposed an estimated 3,900,000 men against 2,500,000 Germans. To the Allies, these figures suggested an offensive on their part.

Nivelle, who owed his appointment to the contrast between the brilliant success of his recent counterattacks at Verdun and the meagre results of Joffre's strategy of attrition, was deeply imbued with the optimism of which experience was by now curing Joffre. He also had ideas of national glory and, accordingly, modified plans made by Joffre in such a way as to assign to the French Army the determinant role in the offensive that, it was calculated, must decide the issue on the Western Front in 1917. Nivelle's plan in its final stage was that the British should make preparatory attacks not only north of the wilderness of the old Somme battlefields but also south of them (in the sector previously held by French troops); that these preparatory attacks should attract the German reserves; and, finally, that the French should launch the major offensive in Champagne (their forces in that sector having been strengthened both by new troops from the overseas colonies and by those transferred from the Somme). The tactics Nivelle planned to use were based on those he had employed so successfully at Verdun. But he placed an optimistic overreliance on his theory of combining "great violence with great mass," which basically consisted of intense artillery bombardments followed by massive frontal attacks.

Meanwhile, Ludendorff had foreseen a renewal of the Allied offensive on the Somme, and he used his time to frustrate Nivelle's plans and to strengthen the German front in two different ways. First, the hitherto rather shallow defenses in Champagne were by mid-February reinforced with a third line, out of range of the French artillery. Second, Ludendorff decided to anticipate the attack by falling back to a new and immensely strong line of defense. This new line, called the Siegfriedstellung, or "Hindenburg Line," was rapidly constructed across the base of the great salient formed by the German lines between Arras and Reims. From the German position east of Arras, the line ran southeastward and southward, passing west of Cambrai and Saint-Quentin to rejoin the old German line at Anizy (between Soissons and Laon). After a preliminary step backward on February 23, a massive withdrawal of all German troops from the westernmost bulges of the great salient to the new and shorter line was smoothly and quickly made on March 16. The major towns within the areas evacuated by the Germans (*i.e.*, Bapaume, Péronne, Roye, Noyon, Chauny, and Coucy) were abandoned to the Allies, but the area was left as a desert, with roads mined, trees cut down, wells fouled, and houses demolished, the ruins being strewn with explosive booby traps.

This baffling and unexpected German withdrawal dislocated Nivelle's plan, but, unperturbed by warnings from all quarters about the changed situation, Nivelle insisted on carrying it out. The Battle of Arras, with which the British started the offensive on April 9, 1917, began well enough for the attackers, thanks to much-improved artillery methods and to a new poison gas shell that paralyzed the hostile artillery. Vimy Ridge, at the northern end of the 15-mile battlefront, fell to the Canadian Corps, but the exploitation of this success was frustrated by the congestion of traffic in the British rear, and though the attack was continued until May 5, stiffer German resistance prevented exploitation of the advances made in the first five days.

Nivelle's offensive: Arras and Champagne

Nivelle's own offensive in Champagne, launched on April 16 on the Aisne front from Vailly eastward toward Craonne and Reims, proved to be a fiasco. The attacking troops were trapped in a web of machine-gun fire, and by nightfall the French had advanced about 600 yards instead of the six miles anticipated in Nivelle's program. Only on the wings was any appreciable progress achieved. The results compared favourably with Joffre's offensives, as some 28,000 German prisoners were taken at a cost to the French of just under 120,000 casualties. But the effect on French morale was worse, because Nivelle's fantastic predictions of the offensive's success were more widely known than Joffre's had ever been. With the collapse of Nivelle's plan, his fortunes were buried in the ruins, and after some face-saving delay he was superseded as commander in chief by Pétain on May 15, 1917.

This change was made too late to avert a more harmful sequel, for in late April a mutiny broke out among the French infantry and spread until 16 French army corps were affected. The authorities chose to ascribe it to seditious propaganda, but the mutinous outbreaks always occurred when exhausted troops were ordered back into the line, and they signaled their grievances by such significant cries as: "We'll defend the trenches, but we won't attack." Pétain restored tranquillity by meeting the just grievances of the troops; his reputation for sober judgment restored the troops' confidence in their leaders, and he made it clear that he would avoid future reckless attacks on the German lines. But the military strength of France could never be fully restored during the war.

Pétain insisted that the only rational strategy was to keep to the defensive until new factors had changed the conditions sufficiently to justify taking the offensive with a reasonable hope of success. His constant advice was: "We must wait for the Americans and the tanks." Tanks were now being belatedly built in large numbers, and this emphasis on them showed a dawning recognition that machine warfare had superseded mass infantry warfare.

The U.S. entry into the war. After the rupture of diplomatic relations with Germany on Feb. 3, 1917, events pushed the United States inexorably along the road to war. Using his authority as commander in chief, Wilson on March 9 ordered the arming of American merchant ships so that they could defend themselves against U-boat attacks. German submarines sank three U.S. merchant ships during March 16–18 with heavy loss of life. Supported by his Cabinet, by most newspapers, and by a large segment of public opinion, Wilson made the decision on March 20 for the United States to declare war on Germany, and on March 21 he called Congress to meet in special session on April 2. He delivered a ringing war message to that body, and the war resolution was approved by the Senate on April 3 and by the House of Representatives on April 6. The presidential declaration of war followed immediately.

The entry of the United States was the turning point of the war, because it made the eventual defeat of Germany possible. It had been foreseen in 1916 that if the United States went to war, the Allies' military effort against Germany would be upheld by U.S. supplies and by enormous extensions of credit. These expectations were amply and decisively fulfilled. The United States' production of armaments was to meet not only its own needs but also France's and Great Britain's. In this sense, the American economic contribution alone was decisive. By April 1, 1917, the Allies had exhausted their means of paying for essential supplies from the United States, and it is difficult to see how they could have maintained the war effort if the United States had remained neutral. American loans to the Allies worth $7,000,000,000 between 1917 and the end of the war maintained the flow of U.S. arms and food across the Atlantic.

The American military contribution was as important as the economic one. A system of conscription was introduced by the Selective Service Act of May 18, 1917, but many months were required for the raising, training, and dispatch to Europe of an expeditionary force. There were still only 85,000 U.S. troops in France when the Germans launched their last great offensive in March 1918; but there were 1,200,000 there by the following September. The U.S. commander in Europe was General John J. Pershing.

The U.S. Navy was the second largest in the world when America entered the war in 1917. The Navy soon abandoned its plans for the construction of battleships and instead concentrated on building the destroyers and submarine chasers so desperately needed to protect Allied shipping from the U-boats. By July 1917 there were already 35 U.S. destroyers stationed at Queenstown (Cobh) on the coast of Ireland—enough to supplement British destroyers for a really effective transatlantic convoy system. By the end of the war there were more than 380 U.S. craft stationed overseas.

The U.S. declaration of war also set an example to other states in the Western Hemisphere. Cuba, Panama, Haiti, Brazil, Guatemala, Nicaragua, Costa Rica, and

Latin-American repercussions

Honduras were all at war with Germany by the end of July 1918, while the Dominican Republic, Peru, Uruguay, and Ecuador contented themselves with the severance of relations.

The Russian revolutions and the Eastern Front, March 1917–March 1918. The Russian Revolution of March (February, old style) 1917 put an end to the autocratic monarchy of imperial Russia and replaced it with a provisional government. But the latter's authority was at once contested by soviets, or "councils of workers' and soldiers' deputies," who claimed to represent the masses of the people and so to be the rightful conductors of the revolution. The March Revolution was an event of tremendous magnitude. Militarily it appeared to the western Allies as a disaster and to the Central Powers as a golden opportunity. The Russian Army remained in the field against the Central Powers, but its spirit was broken, and the Russian people were utterly tired of a war that the imperial regime for its own reasons had undertaken without being morally or materially prepared for it. The Russian Army had been poorly armed, poorly supplied, poorly trained, and poorly commanded and had suffered a long series of defeats. The soviets' propaganda—including the notorious Order No. 1 of the Petrograd Soviet (March 14, 1917), which called for committees of soldiers and sailors to take control of their units' arms and to ignore any opposition from their officers—served to subvert the remnants of discipline in troops who were already deeply demoralized.

But the leaders of the provisional government foresaw that a German victory in the war would bode ill for Russia in the future, and they were also conscious of their nation's obligations toward the western Allies. A.F. Kerensky, minister of war from May 1917, thought that a victorious offensive would enhance the new government's authority, besides relieving pressure on the Western Front. The offensive, however, which General L.G. Kornilov launched against the Austrians in eastern Galicia on July 1, 1917, was brought to a sudden halt by German reinforcements after 10 days of spectacular advances, and it turned into a catastrophic rout in the next three weeks. By October the advancing Germans had won control of most of Latvia and of the approaches to the Gulf of Finland.

Meanwhile, anarchy was spreading over Russia. The numerous non-Russian peoples of the former empire were one after another claiming autonomy or independence from Russia—whether spontaneously or at the prompting of the Germans in occupation of their countries. Finns, Estonians, Latvians, Lithuanians, and Poles were, by the end of 1917, all in various stages of the dissidence from which the independent states of the postwar period were to emerge; and, at the same time, Ukrainians, Georgians, Armenians, and Azerbaijanis were no less active in their own nationalist movements.

The provisional government's authority and influence were rapidly fading away in Russia proper during the late summer and autumn of 1917. The Bolshevik Revolution of November (October, O.S.) 1917 overthrew the provisional government and brought to power the Marxist Bolsheviks under the leadership of Vladimir I. Lenin. The Bolshevik Revolution spelled the end of Russia's participation in the war. Lenin's decree on land, of November 8, undermined the Eastern Front by provoking a homeward rush of soldiers anxious to profit from the expropriation of their former landlords. On November 8, likewise, Lenin issued his decree on peace, which offered negotiations to all belligerents but precluded annexations and indemnities and stipulated a right of self-determination for all peoples concerned. Finally, on November 26, the new Bolshevik government unilaterally ordered a cessation of hostilities both against the Central Powers and against the Turks.

Soviet Russia's separate peace

An armistice between Lenin's Russia and the Central Powers was signed at Brest-Litovsk on Dec. 15, 1917. The ensuing peace negotiations were complicated: on the one hand, Germany wanted peace in the east in order to be free to transfer troops thence to the Western Front, but Germany was at the same time concerned to exploit the principle of national self-determination in order to transfer as much territory as possible into its own safe orbit from that of revolutionary Russia. On the other hand, the Bolsheviks wanted peace in order to be free to consolidate their regime in the east with a view to being able to extend it westward as soon as the time should be ripe. When the Germans, despite the armistice, invaded the Ukraine to cooperate with the Ukrainian nationalists against the Bolsheviks there and furthermore resumed their advance in the Baltic countries and in Belorussia, Lenin rejected his colleague Leon Trotsky's stopgap policy ("neither peace nor war") and accepted Germany's terms in order to save the Bolshevik Revolution. By the Treaty of Brest-Litovsk (March 3, 1918), Soviet Russia recognized Finland and the Ukraine as independent; renounced control over Estonia, Latvia, Lithuania, Poland, and most of Belorussia; and ceded Kars, Ardahan, and Batumi to Turkey.

The conflict between Constantine and Venizélos

Greek affairs. Greece's attitude toward the war was long uncertain: whereas King Constantine I and the general staff stood for neutrality, Eleuthérios Venizélos, leader of the Liberal Party, favoured the Allied cause. As prime minister from 1910, Venizélos wanted Greece to participate in the Allies' Dardanelles enterprise against Turkey in 1915, but his arguments were overruled by the general staff. The Allies occupied Lemnos and Lesbos regardless of Greece's neutrality. Constantine dismissed Venizélos from office twice in 1915, but Venizélos still commanded a majority in Parliament. The Bulgarians' occupation of Greek Macedonia in summer 1916 provoked another political crisis. Venizélos left Athens for Crete late in September, set up a government of his own there, and transferred it early in October to Salonika. On November 27 it declared war on Germany and Bulgaria. Finally, the Allies, on June 11, 1917, deposed King Constantine. Venizélos then returned to Athens to head a reunified Greek government, which on June 27 declared war on the Central Powers.

Caporetto. On the Italian front, Cadorna's 10th Battle of the Isonzo in May–June 1917 won very little ground; but his 11th, from August 17 to September 12, during which General Luigi Capello's 2nd Army captured much of the Bainsizza Plateau (Banjška Planota), north of Gorizia, strained Austrian resistance very severely. To avert an Austrian collapse, Ludendorff decided that the Austrians must take the offensive against Italy and that he could, with difficulty, lend them six German divisions for that purpose.

The offensive was boldly planned, very ably organized, and well executed. While two Austrian armies, under General Svetozar Borojević von Bojna, attacked the eastern end of the Italians' Venetian salient on the Bainsizza Plateau and on the low ground near the Adriatic shore, the German 14th Army, comprising the six German divisions and nine Austrian ones under Otto von Below, with Konrad Krafft von Dellmensingen as his chief of staff, on Oct. 24, 1917, began to force its way over the barrier of the Julian Alps at the northeastern corner of the Venetian salient, with Caporetto approximately opposite the middle point of the line. The Italians, completely surprised by this thrust, which threatened their forces both to the north and to the south, fell back in confusion: Below's van reached Udine, the former site of the Italian general headquarters, by October 28 and was on the Tagliamento River by October 31. Below's success had far exceeded the hopes of the planners of the offensive, and the Germans could not exploit their speedy advance as effectively as they wished. Cadorna, with his centre shattered, managed by precipitate retreat to save the wings of his army and was able, by November 9, to rally his remaining 300,000 troops behind the Piave River, north of Venice. The Italians had sustained about 500,000 casualties, and 250,000 more had been taken prisoner. General Armando Diaz was then appointed commander in chief in Cadorna's place. The Italians managed to hold the Piave front against direct assaults and against attempts to turn its left flank by an advance from the Trentino. The Italians' defense was helped by British and French reinforcements that had been rushed to Italy when the collapse began. A conference of the military and political leaders of the Allies was held at Rapallo in November, and out of this conference there sprang the joint Supreme War Council at Versailles, and ultimately a unified military command.

Mesopotamia, summer 1916–winter 1917. The British

forces in Mesopotamia, neglected hitherto and discouraged by the disaster at al-Kūt (see above), received better attention from London in the second half of 1916; and Sir Frederick Stanley Maude, who became commander in chief in August, did so much to restore their morale that by December he was ready to undertake the recapture of al-Kūt as a first step toward capturing Baghdad.

By a series of outflanking movements, the British made their way gradually and methodically up the Tigris, compelling the Turks to extend their defenses upstream. When the final blow at al-Kūt was delivered by a frontal attack on Feb. 22, 1917, British forces were already crossing the river from the west bank behind the town; but though al-Kūt fell two days later most of the Turkish garrison extricated itself from the threatened encirclement. Unable to hold a new line on the Diyālā River, the Turkish commander, Kâzim Karabekir, evacuated Baghdad, which the British entered on March 11. In September the British position in Baghdad was definitively secured by the capture of ar-Ramādī, on the Euphrates about 60 miles to the west; and early in November the main Turkish force in Mesopotamia was driven from Tikrīt, on the Tigris midway between Baghdad and Mosul.

The capture of Baghdad

Maude, having within a year changed the Mesopotamian scene from one of despair to one of victory, died of cholera on Nov. 18, 1917. His successor in command was Sir William Marshall.

Palestine, autumn 1917. Having assumed command in Egypt (see above), Allenby transferred his headquarters from Cairo to the Palestinian front and devoted the summer of 1917 to preparing a serious offensive against the Turks. On the Turkish side, Falkenhayn, now in command at Aleppo, was at this time himself planning a drive into the Sinai Peninsula for the autumn, but the British were able to strike first.

The Turkish front in southern Palestine extended from Gaza, on the coast, southeastward to Abu Hureira (Tel Haror) and thence to the stronghold of Beersheba. To disguise his real intention of achieving a breakthrough at Abu Hureira, for which, however, the capture of Beersheba was obviously prerequisite, Allenby began his operation with a heavy bombardment of Gaza from October 20 onward. When Beersheba had been seized by converging movements on October 31, a feint attack on Gaza was launched next day to draw the Turkish reserves thither. Then, the main attack, delivered on November 6, broke through the weakened defenses at Abu Hureira and into the plain of Philistia. Falkenhayn had attempted a counterstroke at Beersheba, but the collapse of the Turkish centre necessitated a general retreat. By November 14 the Turkish forces were split in two divergent groups, the port of Jaffa was taken, and Allenby wheeled his main force to the right for an advance inland on Jerusalem. On December 9 the British occupied Jerusalem.

The Western Front, June–December 1917. Pétain's decision to remain temporarily on the defensive after Nivelle's failure gave Haig the opportunity to fulfill his desire for a British offensive in Flanders. He took the first step on June 7, 1917, with a long-prepared attack on the Messines Ridge, north of Armentières, on the southern flank of his Ypres salient. This attack by General Sir Herbert Plumer's 2nd Army proved an almost complete success; it owed much to the surprise effect of 19 huge mines simultaneously fired after having been placed at the end of long tunnels under the German front lines. The capture of the ridge inflated Haig's confidence; and, though General Sir Hubert Gough, in command of the 5th Army, advocated a step-by-step method for the offensive, Haig committed himself to Plumer's view that they "go all out" for an early breakthrough. Haig disregarded the well-founded forecast that, from the beginning of August, rain would be turning the Flanders countryside into an almost impassable swamp. The Germans, meanwhile, were well aware that an offensive was coming from the Ypres salient: the flatness of the plain prevented any concealment of Haig's preparations, and a fortnight's intensive bombardment (4,500,000 shells from 3,000 guns) served to underline the obvious—without, however, destroying the German machine gunners' concrete pillboxes.

Thus, when the Third Battle of Ypres was begun, on July 31, only the left wing's objectives were achieved: on the crucial right wing the attack was a failure. Four days later, the ground was already swampy. When the attack was resumed on August 16, very little more was won, but Haig was still determined to persist in his offensive. Between September 20 and October 4, thanks to an improvement in the weather, the infantry was able to advance into positions cleared by bombardment, but no farther. Haig launched another futile attack on October 12, followed by three more attacks, scarcely more successful, in the last 10 days of October. At last, on November 6, when his troops advanced a very short distance and occupied the ruins of Passchendaele (Passendale), barely five miles beyond the starting point of his offensive, Haig felt that enough had been done. Having prophesied a decisive success without "heavy losses," he had lost 325,000 men and inflicted no comparable damage on the Germans.

The Third Battle of Ypres

Pétain, less pretentious and merely testing what might be done with his rehabilitated French Army, had at least as much to show for himself as Haig. In August the French 2nd Army under General M.-L.-A. Guillaumat fought the last battle of Verdun, winning back all the remainder of what had been lost to the Germans in 1916. In October General P.-A.-M. Maistre's 10th Army, in the Battle of Malmaison, took the ridge of the Chemin des Dames, north of the Aisne to the east of Soissons, where the front in Champagne joined the front in Picardy south of the Somme.

The British, at least, closed the year's campaign with an operation of some significance for the future. When the offensive from Ypres died out in the Flanders mud, they looked again at their tanks, of which they now had a considerable force but which they could hardly use profitably in the swamps. A Tank Corps officer, Colonel J.F.C. Fuller, had already suggested a large-scale raid on the front southwest of Cambrai, where a swarm of tanks, unannounced by any preparatory bombardment, could be released across the rolling downland against the German trenches. This comparatively modest scheme might have been wholly successful if left unchanged, but the British command transformed it: Sir Julian Byng's 3rd Army was to actually try to capture Cambrai and to push on toward Valenciennes. On November 20, therefore, the attack was launched, with 324 tanks leading Byng's six divisions. The first massed assault of tanks in history took the Germans wholly by surprise, and the British achieved a far deeper penetration and at less cost than in any of their past offensives. Unfortunately, however, all of Byng's troops and tanks had been thrown into the first blow, and, as he was not reinforced in time, the advance came to a halt several miles short of Cambrai. A German counterstroke, on November 30, broke through on the southern flank of the new British salient and threatened Byng's whole army with disaster before being checked by a further British counterattack. In the end, three-quarters of the ground that the British had won was reoccupied by the Germans. Even so, the Battle of Cambrai had proved that surprise and the tank in combination could unlock the trench barrier.

Cambrai

The Far East. China's entry into the war in 1917 on the side of the Allies was motivated not by any grievance against the Central Powers but by the Peking government's fear lest Japan, a belligerent since 1914, should monopolize the sympathies of the Allies and of the United States when Far Eastern affairs came up for settlement after the war. Accordingly, in March 1917 the Peking government severed its relations with Germany; and on August 14 China declared war not only on Germany but also on the western Allies' other enemy, Austria-Hungary. China's contribution to the Allied war effort was to prove negligible in practical effects, however.

Naval operations, 1917–18. Since Germany's previous restrictions of its submarine warfare had been motivated by fear of provoking the United States into war, the U.S. declaration of war in April 1917 removed any reason for the Germans to retreat from their already declared policy of unrestricted warfare. Consequently, the U-boats, having sunk 181 ships in January, 259 in February, and 325 in March, sank 430 in April. The April sinkings

represented 852,000 gross tons, to be compared both with the 600,000 postulated by the German strategists as their monthly target and with the 700,000 that the British in March had pessimistically foretold for June. The Germans had calculated that if the world's merchant shipping could be sunk at the monthly rate of 600,000 tons, the Allies, being unable to build new merchant ships fast enough to replace those lost, could not carry on the war for more than five months. At the same time, the Germans, who had 111 U-boats operational when the unrestricted campaign began, had embarked on an extensive building program that, when weighed against their current losses of one or two U-boats per month, promised a substantial net increase in the U-boats' numbers. During April, one in every four of the merchant ships that sailed from British ports was destined to be sunk, and by the end of May the quantity of shipping available to carry the vital foodstuffs and munitions to Great Britain had been reduced to only 6,000,000 tons.

The convoy system

The April total, however, proved to be a peak figure—primarily because the Allies at last adopted the convoy system for the protection of merchant ships. Previously, a ship bound for one of the Allies' ports had set sail by itself as soon as it was loaded. The sea was thus dotted with single and unprotected merchant ships, and a scouting U-boat could rely on several targets coming into its range in the course of a cruise. The convoy system remedied this by having groups of merchant ships sail within a protective ring of destroyers and other naval escorts. It was logistically possible and economically worthwhile to provide this kind of escort for a group of ships. Furthermore, the combination of convoy and escort would force the U-boat to risk the possibility of a counterattack in order to sink the merchant ships, thus giving the Allies a prospect of reducing the U-boats' numbers. Despite the manifest and seemingly overwhelming benefits of the convoy system, the idea was novel and, like any untried system, met with powerful opposition from within the military. It was only in the face of extreme necessity and under great pressure from Lloyd George that the system was tried, more or less as a last resort.

The first convoy sailed from Gibraltar to Great Britain on May 10, 1917; the first from the United States sailed later in May; ships using the South Atlantic sailed in convoy from July 22. During the later months of 1917 the use of convoys caused an abrupt fall in the sinkings by U-boats: 500,500 tons in May, 300,200 in September, and only about 200,600 in November. The convoy system was so quickly vindicated that in August it was extended to shipping outward-bound from Great Britain. The Germans themselves soon observed that the British had grasped the principles of antisubmarine warfare, and that sailing ships in convoys considerably reduced the opportunities for attack.

Apart from the convoys, the Allies improved their antisubmarine technology (hydrophones, depth charges, etc.) and extended their minefields. In 1918, moreover, Admiral Sir Roger Keyes, in command at Dover, set up a system whereby the English Channel was patrolled by surface craft with searchlights, so that U-boats passing through it had to submerge themselves to depths at which they were liable to strike the mines that had been laid for them. Subsequently, most of the U-boats renounced the Channel as a way into the Atlantic and instead took the passage north of Great Britain, thus losing precious fuel and time before reaching the heavily traveled sea lanes of the western approaches to Great Britain. In the summer of 1918, U.S. minelayers laid more than 60,000 mines (13,000 of them British) in a wide belt across 180 miles of the North Sea between Scotland and Norway, so as to obstruct the U-boats' only access from Germany to the Atlantic other than the closely guarded Channel.

The cumulative effect of all these measures was the gradual containment and ultimately the defeat of the U-boat campaign, which never again achieved the success of April 1917. While sinkings by submarines, after that month, steadily fell, the losses of U-boats showed a slow but steady rise, and more than 40 were destroyed in the first six months of 1918. At the same time the replacement of merchant vessels in the building program improved steadily, until it eventually far outstripped losses. In October 1918, for example, 511,000 tons of new Allied merchant ships were launched, while only 118,559 tons were lost.

Air warfare. At the start of the war the land and sea forces used the aircraft put at their disposal primarily for reconnaissance, and air fighting began as the exchange of shots from small arms between enemy airmen meeting one another in the course of reconnoitering. Fighter aircraft armed with machine guns, however, made their appearance in 1915. Tactical bombing and the bombing of enemy air bases were also gradually introduced at this time. Contact patrolling, with aircraft giving immediate support to infantry, was developed in 1916.

Strategic bombing

Strategic bombing, on the other hand, was initiated early enough: British aircraft from Dunkirk bombed Cologne, Düsseldorf, and Friedrichshafen in the autumn of 1914, their main objective being the sheds of the German dirigible airships, or Zeppelins; and raids by German airplanes or seaplanes on English towns in December 1914 heralded a great Zeppelin offensive sustained with increasing intensity from January 1915 to September 1916 (London was first bombed in the night of May 31–June 1, 1915). In October 1916 the British, in turn, began a more systematic offensive, from eastern France, against industrial targets in southwestern Germany.

While the British directed much of their new bombing strength to attacks on the bases of the U-boats, the Germans used theirs largely to continue the offensive against the towns of southeastern England. On June 13, 1917, in daylight, 14 German bombers dropped 118 high explosive bombs on London and returned home safely. This lesson and that of subsequent raids by the German Gotha bombers made the British think more seriously about strategic bombing and about the need for an air force independent of the other fighting services. The Royal Air Force (RAF), the world's first separate air service, was brought into active existence by a series of measures taken between October 1917 and June 1918.

Peace moves, March 1917–September 1918. Until the end of 1916, the pursuit of peace was confined to individuals and to small groups. In the following months it began to acquire a broad popular backing. Semi-starvation in towns, mutinies in the armies, and casualty lists that seemed to have no end made more and more people question the need and the wisdom of continuing the war.

Francis Joseph, Austria's venerable old emperor, died on Nov. 21, 1916. The new emperor, Charles I, and his foreign minister, Graf Ottokar Czernin, initiated peace moves in the spring of 1917 but unfortunately did not concert their diplomatic efforts, and the channels of negotiation they opened between Austria-Hungary and the Allies had dried up by that summer.

In Germany, Matthias Erzberger, a Roman Catholic member of the Reichstag, had, on July 6, 1917, proposed that territorial annexations be renounced in order to facilitate a negotiated peace. During the ensuing debates Bethmann Hollweg resigned the office of chancellor, and the emperor William II appointed the next chancellor, Ludendorff's nominee Georg Michaelis, without consulting the Reichstag. The Reichstag, offended, proceeded to pass its *Friedensresolution,* or "peace resolution," of July 19 by 212 votes. The peace resolution was a string of innocuous phrases expressing Germany's desire for peace but without a clear renunciation of annexations or indemnities. The Allies took almost no notice of it.

Erzberger's proposal of July 6 had been intended to pave the way for Pope Benedict XV's forthcoming note to the belligerents of both camps. Dated Aug. 1, 1917, this note advocated a German withdrawal from Belgium and from France, the Allies' withdrawal from the German colonies, and the restoration not only of Serbia, Montenegro, and Romania but also of Poland to independence. France and Great Britain declined to give an express reply pending Germany's statement of its attitude about Belgium, on which Germany avoided committing itself.

An unofficial peace move was made in London: on Nov. 29, 1917, the *Daily Telegraph* published a letter from Lord Lansdowne suggesting negotiations on the basis of

the status quo antebellum. Lloyd George rejected Lansdowne's theses on December 14.

Wilson's Fourteen Points

The U.S. president Woodrow Wilson made himself the chief formulator and spokesman of the war aims of the Allies and the United States. The first nine months of 1918 saw Wilson's famous series of pronouncements on his war aims: the Fourteen Points (January 8), the "Four Principles" (February 11), the "Four Ends" (July 4), and the "Five Particulars" (September 27). Most important, not least because of Germany's deluded reliance on them in its eventual suing for peace, were the Fourteen Points: (1) open covenants of peace and the renunciation of secret diplomacy, (2) freedom of navigation on the high seas in wartime as well as peace, (3) the maximum possible freedom of trade, (4) a guaranteed reduction of armaments, (5) an impartial colonial settlement accommodating not only the colonialist powers but also the peoples of the colonies, (6) the evacuation of all Russian territory and respect for Russia's right of self-determination, (7) the complete restoration of Belgium, (8) a complete German withdrawal from France and satisfaction for France about Alsace-Lorraine, (9) a readjustment of Italy's frontiers on an ethnic basis, (10) an open prospect of autonomy for the peoples of Austria-Hungary, (11) the restoration of Romania, Serbia, and Montenegro, with free access to the sea for Serbia and international guarantees of the Balkan states' independence and integrity, (12) the prospect of autonomy for non-Turkish peoples of the Ottoman Empire and the unrestricted opening of the Straits, but secure sovereignty for the Turks in their own areas, (13) an independent Poland with access to the sea and under international guarantee, and (14) "a general association of nations," to guarantee the independence and integrity of all states, great and small. The three subsequent groups of pronouncements mainly consisted of idealistic expansions of themes implicit in the Fourteen Points, with increasing emphasis on the wishes of subject populations; but the first of the "Four Ends" was that every arbitrary power capable by itself of disturbing world peace should be rendered innocuous.

Wilson's peace campaign was a significant factor in the collapse of the will to fight of the German people and the decision of the German government to sue for peace in October 1918. Indeed, the Germans conducted their preliminary peace talks exclusively with Wilson. And the Armistice, when it came on Nov. 11, 1918, was formally based upon the Fourteen Points and additional Wilsonian pronouncements, with two reservations by the British and French relating to freedom of the seas and reparations.

The last offensives and the Allies' victory

THE WESTERN FRONT, MARCH–SEPTEMBER 1918

As the German strength on the Western Front was being steadily increased by the transfer of divisions from the Eastern Front (where they were no longer needed since Russia had withdrawn from the war), the Allies' main problem was how to withstand an imminent German offensive pending the arrival of massive reinforcements from the United States. Eventually Pétain persuaded the reluctant Haig that the British with 60 divisions should extend their sector of the front from 100 to 125 miles as compared with the 325 miles to be held by the French with approximately 100 divisions. Haig thus devoted 46 of his divisions to the front from the Channel to Gouzeaucourt (southwest of German-held Cambrai) and 14 to the remaining third of the front from Gouzeaucourt past German-held Saint-Quentin to the Oise River.

On the German side, between Nov. 1, 1917, and March 21, 1918, the German divisions on the Western Front were increased from 146 to 192, the troops being drawn from Russia, Galicia, and Italy. By these means the German armies in the west were reinforced by a total of about 570,000 men. Ludendorff's interest was to strike from his temporary position of strength—before the arrival of the major U.S. contingents—and at the same time to ensure that his German offensive should not fail for the same reasons as the Allies' offensives of the past three years. Accordingly he formed an offensive strategy based on taking the tactical line of least resistance. The main German attacks would begin with brief but extremely intense artillery bombardments using a high proportion of poison gas and smoke shells. These would incapacitate the Allies' forward trenches and machine-gun emplacements and would obscure their observation posts. Then a second and lighter artillery barrage would begin to creep forward over the Allied trenches at a walking pace (in order to keep the enemy under fire), with the masses of German assault infantry advancing as closely as possible behind it. The key to the new tactics was that the assault infantry would bypass machine-gun nests and other points of strong resistance instead of waiting, as had been the previous practice on both sides, for reinforcements to mop up the obstructions before continuing the advance. The Germans would instead continue to advance in the direction of the least enemy resistance. The mobility of the German advance would thus be assured, and its deep infiltration would result in large amounts of territory being taken.

Such tactics demanded exceptionally fit and disciplined troops and a high level of training. Ludendorff accordingly drew the best troops from all the Western Front forces at his disposal and formed them into elite shock divisions. The troops were systematically trained in the new tactics, and every effort was also made to conceal the actual areas at which the German main attacks would be made.

Ludendorff's western offensive

Ludendorff's main attack was to be on the weakest sector of the Allies' front, the 47 miles between Arras and La Fère (on the Oise). Two German armies, the 17th and the 2nd, were to break through the front between Arras and Saint-Quentin, north of the Somme, and then wheel right so as to force most of the British back toward the Channel, while the 18th Army, between the Somme and the Oise, protected the left flank of the advance against counterattack from the south. Code-named "Michael," this offensive was to be supplemented by three other attacks: "St. George I" against the British on the Lys River south of Armentières; "St. George II" against the British again between Armentières and Ypres; and "Blücher" against the French in Champagne. It was finally decided to use 62 divisions in the main attack, "Michael."

Preceded by an artillery bombardment using 6,000 guns, "Michael" was launched on March 21, 1918, and was helped by an early morning fog that hid the German advance from the Allied observation posts. The attack, which is known as the Second Battle of the Somme or the Battle of Saint-Quentin, took the British altogether by surprise, but it did not develop as Ludendorff had foreseen. While the 18th Army under von Hutier achieved a complete breakthrough south of the Somme, the major attack to the north was held up, mainly by the British concentration of strength at Arras. For a whole week Ludendorff, in violation of his new tactical emphasis, vainly persisted in trying to carry out his original plan instead of exploiting the unexpected success of the 18th Army, though the latter had advanced more than 40 miles westward and had reached Montdidier by March 27. At last, however, the main effort of the Germans was converted into a drive toward Amiens, which began in force on March 30. By that time the Allies had recovered from their initial dismay, and French reserves were coming up to the British line. The German drive was halted east of Amiens and so too was a renewed attack on April 4. Ludendorff then suspended his Somme offensive. This offensive had yielded the largest territorial gains of any operation on the Western Front since the First Battle of the Marne in September 1914.

The Allies' cause at least derived one overdue benefit from the collapse of one-third of the British front: at Haig's own suggestion, Foch was on March 26 appointed to coordinate the Allies' military operations; and on April 14 he was named commander in chief of the Allied armies. Previously, Haig had resisted the idea of a generalissimo.

On April 9 the Germans began "St. George I" with an attack on the extreme northern front between Armentières and the canal of La Bassée, their aim being to advance across the Lys River toward Hazebrouck. Such was the initial success of this attack that "St. George II" was launched the next day, with the capture of Kemmel Hill (Kemmelberg), southwest of Ypres, as its first objective.

Armentières fell, and Ludendorff came to think for a time that this Battle of the Lys might be turned into a major effort. The British, however, after being driven back 10 miles, halted the Germans short of Hazebrouck. French reinforcements began to come up; and, when the Germans had taken Kemmel Hill (April 25), Ludendorff decided to suspend exploitation of the advance, for fear of a counterstroke against his front's new bulge.

Thus far Ludendorff had fallen short of strategic results, but he could claim huge tactical successes—the British casualties alone amounted to more than 300,000. Ten British divisions had to be broken up temporarily, while the German strength mounted to 208 divisions, of which 80 were still in reserve. A restoration of the balance, however, was now in sight. A dozen U.S. divisions had arrived in France, and great efforts were being made to swell the stream. Furthermore, Pershing, the U.S. commander, had placed his troops at Foch's disposal for use wherever required.

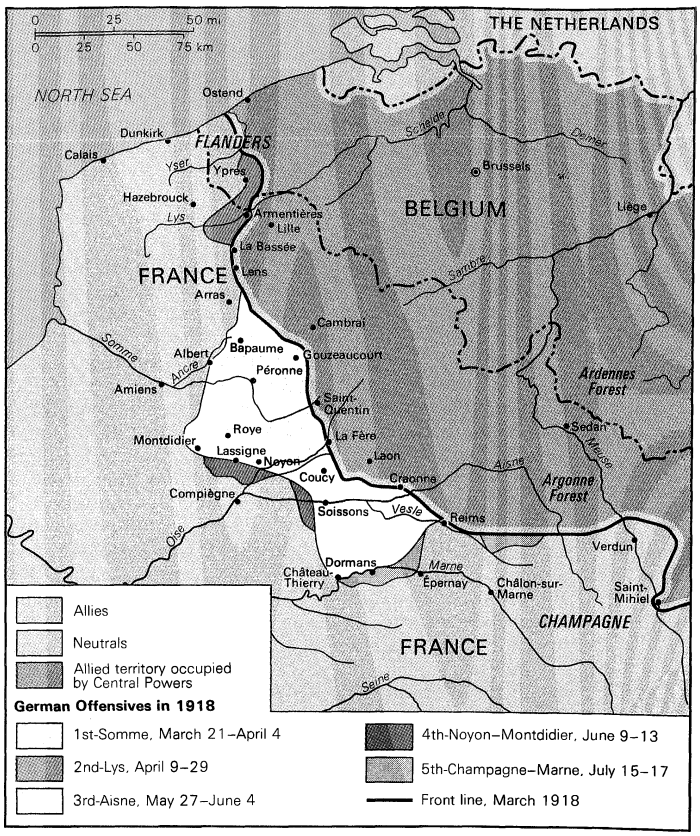
Ludendorff finally launched "Blücher" on May 27, on a front extending from Coucy, north of Soissons, eastward toward Reims. The Germans, with 15 divisions, suddenly attacked the seven French and British divisions opposing them, swarmed over the ridge of the Chemin des Dames and across the Aisne River, and, by May 30, were on the Marne, between Château-Thierry and Dormans. Once again the attack's initial success went far beyond Ludendorff's expectation or intention; and, when the Germans tried to push westward against the right flank of the Allies' Compiègne salient, which was sandwiched between the Germans' Amiens and Champagne bulges, they were checked by counterattacks, which included one sustained for a fortnight from June 6 by U.S. divisions at Belleau Wood (Bois de Belleau). An attack from Noyon, against the left flank of the Compiègne salient, came too late (June 9).

Overtaken by the inordinate fruition of his own offensives, Ludendorff paused for a month's recuperation. The tactical success of his own blows had been his undoing; yielding to their influence, he had pressed each too far and too long, using up his own reserves and causing an undue interval between blows. He had driven three great wedges into the Allied lines, but none had penetrated far enough to sever a vital rail artery, and this strategic failure left the Germans with a front whose several bulges invited flanking counterstrokes. Moreover, Ludendorff had used up many of his shock troops in the attacks, and the remaining troops, though strong in numbers, were relatively lower in quality. The Germans were to end up sustaining a total of 800,000 casualties in their great 1918 offensives. Meanwhile, the Allies were now receiving U.S. troops at the rate of 300,000 men per month.

The next German offensive, which opened the Second Battle of the Marne, was launched in Champagne on July 15. It came to nothing: a German thrust from the front east of Reims toward Châlons-sur-Marne was frustrated by the "elastic defense" that Pétain had recently been prescribing but that the local commanders had failed to practice against the offensive of May 27. A drive from Dormans, on the left flank of the Germans' huge Soissons–Reims bulge, across the Marne toward Épernay simply made the Germans' situation more precarious when Foch's long-prepared counterstroke was launched on July 18. In this great counterstroke one of Foch's armies assailed the Germans' Champagne bulge from the west, another from the southwest, one more from the south, and a fourth from the vicinity of Reims. Masses of light tanks—a weapon on which Ludendorff had placed little reliance, preferring gas instead in his plans for the year—played a vital part in forcing the Germans into a hasty retreat. By August 2 the French had pushed the Champagne front back to a line following the Vesle River from Reims and then along the Aisne to a point west of Soissons.

The Second Battle of the Marne

Having recovered the initiative, the Allies were determined not to lose it, and for their next blow they chose again the front north and south of the Somme. The British 4th Army, including Australian and Canadian forces, with 450 tanks, struck the Germans with maximum surprise on Aug. 8, 1918. Overwhelming the German forward divisions, who had failed to entrench themselves adequately since their recent occupation of the "Michael" bulge, the 4th Army advanced steadily for four days, taking 21,000 prisoners and inflicting as many or more casualties at the cost of only about 20,000 casualties to itself, and halting only when it reached the desolation of the old battlefields of 1916. Several German divisions simply collapsed in the face of the offensive, their troops either fleeing or surrendering. The Battle of Amiens was thus a striking material and moral success for the Allies. Ludendorff put it differently: "August 8 was the black day of the German Army in the history of the war . . . It put the decline of our fighting power beyond all doubt . . . The war must be ended." He informed Emperor William II and Germany's political chiefs that peace negotiations should be opened before the situation became worse, as it must. The conclusions reached at a German Crown Council held at Spa were that "We can no longer hope to break the war-will of our enemies by military operations," and "the objects of our strategy must be to paralyse the enemy's war-will gradually by a strategic defensive." In other words, the German high command had abandoned hope of victory or even of holding their gains and hoped only to avoid surrender.



German offensives on the Western Front, March–July 1918.

Meanwhile, the French had retaken Montdidier and were thrusting toward Lassigny (between Roye and Noyon); and on August 17 they began a new drive from the Compiègne salient south of Noyon. Then, in the fourth week of August, two more British armies went into action on the Arras–Albert sector of the front, the one advancing directly eastward on Bapaume, the other operating farther to the north. From then on Foch delivered a series of hammer blows along the length of the German front, launching a series of rapid attacks at different points, each broken off as soon as its initial impetus waned, and all close enough in time to attract German reserves, which consequently were unavailable to defend against the next Allied attack along a different part of the front. By the early days of September the Germans were back where they had been before March 1918—behind the Hindenburg Line.

The Allies' recovery was consummated by the first feat executed by Pershing's U.S. forces as an independent army

(hitherto the U.S. divisions in France had fought only in support of the major French or British units): the U.S. 1st Army on September 12 erased the triangular Saint-Mihiel salient that the Germans had been occupying since 1914 (between Verdun and Nancy).

The clear evidence of the Germans' decline decided Foch to seek victory in the coming autumn of 1918 instead of postponing the attempt until 1919. All the Allied armies in the west were to combine in a simultaneous offensive.

OTHER DEVELOPMENTS IN 1918

Czechs, Yugoslavs, and Poles. Something must now be said about the growth of the national movements, which, under the eventual protection of the Allies, were to result in the foundation of new states or the resurrection of long-defunct ones at the end of the war. There were three such movements: that of the Czechs, with the more backward Slovaks in tow; that of the South Slavs, or Yugoslavs (Serbs, Croats, and Slovenes); and that of the Poles. The Czech country, namely Bohemia and Moravia, belonged in 1914 to the Austrian half of the Habsburg monarchy, the Slovak to the Hungarian half. The Yugoslavs had already been represented in 1914 by two independent kingdoms, Serbia and Montenegro, but they were also predominantly numerous in territories still under Habsburg rule: Serbs in Bosnia and Hercegovina (an Austro-Hungarian condominium) and in Dalmatia (an Austrian possession); Croats in Croatia (Hungarian), in Istria (Austrian), and in Dalmatia; Slovenes in Istria and in Illyria (Austrian likewise). Poland was divided into three parts: Germany had the north and the west as provinces of the Kingdom of Prussia; Austria had Galicia (including an ethnically Ukrainian extension to the east); Russia had the rest.

The Czechs had long been restless under the Austrian regime, and one of their leading intellectual spokesmen, Tomáš Masaryk (in fact a Slovak), had already envisaged the carving of Czechoslovak and Yugoslav states out of Austria-Hungary in December 1914. In 1916 he and a fellow émigré, Edvard Beneš, based respectively in London and in Paris, organized a Czechoslovak National Council. The western Allies committed themselves to the Czechoslovak idea from 1917 onward, when Russia's imminent defection from the war made them ready to exploit any means at hand for the disabling of Austria-Hungary; and Wilson's sympathy was implicit in his successive peace pronouncements of 1918.

The exiles and the Allies

For the South Slavs of Austria-Hungary the Yugoslav Committee, with representatives in Paris and in London, was founded in April 1915. On July 20, 1917, this committee and the Serbian government in exile made the joint Corfu Declaration forecasting a South Slav state to comprise Serbs, Croats, and Slovenes.

The Polish nationalist leaders in the first years of the war were uncertain whether to rely on the Central Powers or on the Allies for a restoration of Poland's independence. So long as the western Allies hesitated to encourage Polish nationalism for fear of offending imperial Russia, the Central Powers seemed to be the most likely sponsors; and Austria at least allowed Józef Piłsudski, from 1914, to organize his volunteer Polish legions to serve with Austrian forces against the Russians. Austria's benevolence, however, was not reflected by Germany; and when the Two Emperors' Manifesto of Nov. 5, 1916, provided for the constitution of an independent Polish kingdom, it was clear that this kingdom would consist only of Polish territory conquered from Russia, not of any German or Austrian territory. When, after the March Revolution of 1917, the Russian provisional government had recognized Poland's right to independence, Roman Dmowski's Polish National Committee, which from 1914 had been functioning in a limited way under Russian protection, could at last count seriously on the sympathy of the western Allies. While Piłsudski declined to raise a Polish army to fight on against the new Russia, a Polish army was formed in France, as well as two army corps in Belorussia and in the Ukraine, to fight against the Central Powers. The Bolshevik Revolution and Wilson's Fourteen Points together consummated the alignment of the Poles on the side of the western powers.

Eastern Europe and the Russian periphery, March–November 1918. The Treaty of Brest-Litovsk (March 3, 1918) gave Germany a free hand to do what it liked with Russia's former possessions in eastern Europe. While they pursued their plan of 1916 for a kingdom of Poland, the Germans took new measures for the other countries. Lithuania, recognized as independent, was to be a kingdom under some German prince. Latvia and Estonia were to be merged into a grand duchy of the Baltikum under the hereditary rule of Prussia. An expeditionary force of 12,000 men, under General Graf Rüdiger von der Goltz, was sent to Finland to uphold the Finnish general C.G.E. Mannerheim's nationalist forces against the Red Guards, whom the Bolsheviks, despite their recognition of Finland's independence, were now promoting there. And finally, the Ukrainian nationalist government, which had already been challenged by a Communist one before its separate peace with the Central Powers (Brest-Litovsk, February 9), was promptly displaced by a new regime after the advance of German and Austro-Hungarian troops into its territory.

The Romanian armistice of December 1917 was converted into the Treaty of Bucharest on May 7, 1918. Under this treaty's terms, southern Dobruja was ceded to Bulgaria; northern Dobruja was put under the joint administration of the Central Powers; and the latter obtained virtual control of Romania's oil fields and communications. Romania, on the other hand, had some consolation from Bessarabia, whose nationalists, after receiving Romanian assistance against the Bolsheviks, had voted in March 1918 for their country's conditional union with Romania.

Even Transcaucasia began to slide into the German camp. The short-lived federal republic was dissolved by its three members' individual declarations of independence—Georgia's on May 26, Armenia's and Azerbaijan's on May 28. Treaties of friendship were promptly signed between Georgia and Germany and between Armenia and Turkey, and Turkish troops advanced into Azerbaijan, where they occupied Baku on September 15. The western Allies, meanwhile, were hoping that some new semblance of an Eastern Front could be conjured up if they supported the various and growing forces in Russia that were opposed to the peacemaking Bolsheviks. Since the Black Sea and the Baltic were closed to them, the Allies could land troops only on Russia's Arctic and Pacific shores. Thus, the Allied "intervention" in Russia on the side of the anti-Bolshevik ("White") forces, long to be execrated by Soviet historians, began with an Anglo-French landing at Murmansk, in the far north, on March 9, 1918. The subsequent reinforcement of Murmansk made possible the occupation of the Murmansk railway as far south as Soroka (now Belomorsk); and a further landing at Arkhangelsk in the summer raised the total Allied strength in northern Russia to some 48,000 (including 20,000 Russian "Whites"). By this time, moreover, there were some 85,000 interventionist troops in Siberia, where a strong Japanese landing at Vladivostok in April had been followed by British, French, Italian, and U.S. contingents. A "White" provisional government of Russia was set up at Omsk, with Admiral A.V. Kolchak as its dominant personality. The "White" resistance in the south of European Russia, which had been growing since November 1917, was put under the supreme command of General A.I. Denikin in April 1918.

Allied "intervention" in Russia

The Balkan front, 1918. At Salonika the Allies' politically ambitious but militarily ineffective commander in chief, General Sarrail, was replaced at the end of 1917 by General Guillaumat, who was in turn succeeded in July 1918 by General L.-F.-F. Franchet d'Esperey, who launched a major offensive in September with six Serbian and two French divisions against a seven-mile front held by only one Bulgarian division.

The initial assault, preceded by heavy bombardment at night, began in the morning of Sept. 15, 1918, and a five-mile penetration was achieved by nightfall on September 16. The next day the Serbs advanced 20 miles forward, while French and Greek forces on their flanks widened the breach to 25 miles. A British attack, launched on September 18 on the front between the Vardar and Lake Doiran, prevented the Bulgars from transferring troops

westward against the right flank of the penetration; and by September 19 the Serbian cavalry had reached Kavadarci, at the apex of the Crna–Vardar triangle. Two days later the whole Bulgarian front west of the Vardar had collapsed.

While Italian forces in the extreme west advanced on Prilep, the elated Serbs, with the French beside them, pressed on up the Vardar Valley. The British in the east now made such headway as to take Strumica, across the old Bulgarian frontier, on September 26. The Bulgars then sued for an armistice; and on September 29, when a bold French cavalry thrust up the Vardar from Veles (Titov Veles) took Skopje, key to the whole system of communications for the Balkan front, Bulgarian delegates signed the Armistice of Salonika, accepting the Allies' terms unreservedly.

The Turkish fronts, 1918. The British–Turkish front in Palestine in the summer of 1918 ran from the Jordan River westward north of Jericho and Lydda to the Mediterranean just north of Jaffa. North of this front there were three Turkish "armies" (in fact, barely stronger than divisions): one to the east of the Jordan, two to the west. These armies depended for their supplies on the Hejaz Railway, the main line of which ran from Damascus southward, east of the Jordan, and which was joined at Déraa (Dar'ā) by a branch line serving Palestine.

Liman von Sanders, Falkenhayn's successor as commander of the Turkish forces in Syria–Palestine, was convinced that the British would make their main effort east of the Jordan. Allenby, however, was really interested in taking a straight northerly direction, reckoning that the Palestine branch rail line at 'Afula and Beisān, some 60 miles behind the Turkish front, could be reached by a strategic "bound" of his cavalry and that their fall would isolate the two Turkish armies in the west.

Having by ruse and diversion induced the Turks to reduce their strength in the west, Allenby struck there on Sept. 19, 1918, with a numerical superiority of 10 to one.

The Battle of Megiddo

In this Battle of Megiddo, a British infantry attack swept the astonished defenders aside and opened the way for the cavalry, which rode 30 miles north up the coastal corridor before swinging inland to cut the Turks' northward lines of retreat. 'Afula, Beisān, and even Nazareth, farther north, were in British hands the next day.

When the Turks east of the Jordan River began to retreat on September 22, the Arabs had already severed the railway line and were lying in wait for them; and a British cavalry division from Beisān was also about to push eastward to intercept their withdrawal. Simultaneously, two more British divisions and another force of Arabs were racing on toward Damascus, which fell on October 1. The campaign ended with the capture of Aleppo and the junction of the Baghdad Railway. In 38 days Allenby's forces had advanced 350 miles and taken 75,000 prisoners at a cost of less than 5,000 casualties.

In Mesopotamia, meanwhile, the British had taken Kifrī, north of the Diyālā left-bank tributary of the Tigris, in January 1918, and Khān al-Baghdādī, up the Euphrates, in March. Pressing northward from Kifrī, they took Kirkūk in May but soon evacuated it.

The British centre in Mesopotamia, advancing up the Tigris in October, was about to capture Mosul when the hostilities were suspended. The Ottoman government, seeing eastern Turkey defenseless and fearing an Allied advance against Istanbul from the west now that Bulgaria had collapsed, decided to capitulate. On October 30 the Armistice of Mudros was signed, on a British cruiser off Lemnos. The Turks, by its terms, were to open the Straits to the Allies; demobilize their forces; allow the Allies to occupy any strategic point that they might require and to use all Turkey's ports and railways; and order the surrender of their remaining garrisons in Arabia, Syria, and Mesopotamia. The centuries-old Ottoman Empire had come to an end.

The Turkish armistice

Vittorio Veneto. After the stabilization of the Italian front on the Piave River at the end of 1917, the Austrians made no further move until the following June. They then tried not only to force the Tonale Pass and enter northeastern Lombardy but also to make two converging thrusts into central Venetia, the one southeastward from the Trentino, the other southwestward across the lower Piave. The whole offensive came to worse than nothing, the attackers losing 100,000 men.

Diaz, the Italian commander in chief, was meanwhile deliberately abstaining from positive action until Italy should be ready to strike with success assured. In the offensive he planned, three of the five armies lining the front from the Monte Grappa sector to the Adriatic end of the Piave were to drive across the river toward Vittorio Veneto, so as to cut communications between the two Austrian armies opposing them.

When Germany, in October 1918, was at last asking for an armistice (see below), Italy's time had obviously come. On October 24, the anniversary of Caporetto, the offensive opened. An attack in the Monte Grappa sector was repulsed with heavy loss, though it served to attract the Austrian reserves, and the flooding of the Piave prevented two of the three central armies from advancing simultaneously with the third; but the latter, comprising one Italian and one British corps, having under cover of darkness and fog occupied Papadopoli Island farther downstream, won a foothold on the left bank of the river on October 27. The Italian reserves were then brought up to exploit this bridgehead.

Austrian mutiny

Mutiny was already breaking out in the Austrian forces, and on October 28 the Austrian high command ordered a general retreat. Vittorio Veneto was occupied the next day by the Italians, who were also pushing on already toward the Tagliamento. On November 3 the Austrians obtained an armistice (see below).

The collapse of Austria-Hungary. The duality of the Habsburg monarchy had been underlined from the very beginning of the war. Whereas the Austrian parliament, or Reichsrat, had been suspended in March 1914 and was not reconvened for three years, the Hungarian parliament in Budapest continued its sessions, and the Hungarian government proved itself constantly less amenable to dictation from the military than had the Austrian. The Slav minorities, however, showed little sign of anti-Habsburg feeling before Russia's March Revolution of 1917. In May 1917, however, the Reichsrat was reconvened, and just before the opening session the Czech intelligentsia sent a manifesto to its deputies calling for "a democratic Europe . . . of autonomous states." The Bolshevik Revolution of November 1917 and the Wilsonian peace pronouncements from January 1918 onward encouraged socialism, on the one hand, and nationalism, on the other, or alternatively a combination of both tendencies, among all peoples of the Habsburg monarchy.

Early in September 1918 the Austro-Hungarian government proposed in a circular note to the other powers that a conference be held on neutral territory for a general peace. This proposal was quashed by the United States on the ground that the U.S. position had already been enunciated by the Wilsonian pronouncements (the Fourteen Points, etc.). But when Austria-Hungary, after the collapse of Bulgaria, appealed on October 4 for an armistice based on those very pronouncements, the answer on October 18 was that the U.S. government was now committed to the Czechoslovaks and to the Yugoslavs, who might not be satisfied with the "autonomy" postulated heretofore. The emperor Charles had, in fact, granted autonomy to the peoples of the Austrian Empire (as distinct from the Hungarian Kingdom) on October 16, but this concession was ignored internationally and served only to facilitate the process of disruption within the monarchy: Czechoslovaks in Prague and South Slavs in Zagreb had already set up organs ready to take power.

The last scenes of Austria-Hungary's dissolution were performed very rapidly. On October 24 (when the Italians launched their very timely offensive), a Hungarian National Council prescribing peace and severance from Austria was set up in Budapest. On October 27 a note accepting the U.S. note of October 18 was sent from Vienna to Washington—to remain unacknowledged. On October 28 the Czechoslovak committee in Prague passed a "law" for an independent state, while a similar Polish committee was formed in Kraków for the incorporation of Galicia and Austrian Silesia into a unified Poland. On

The national states proclaimed independent

October 29, while the Austrian high command was asking the Italians for an armistice, the Croats in Zagreb declared Slavonia, Croatia, and Dalmatia to be independent, pending the formation of a national state of Slovenes, Croats, and Serbs. On October 30 the German members of the Reichsrat in Vienna proclaimed an independent state of German Austria.

The solicited armistice between the Allies and Austria-Hungary was signed at the Villa Giusti, near Padua, on Nov. 3, 1918, to become effective on November 4. Under its provisions, Austria-Hungary's forces were required to evacuate not only all territory occupied since August 1914 but also South Tirol, Tarvisio, the Isonzo Valley, Gorizia, Trieste, Istria, western Carniola, and Dalmatia. All German forces should be expelled from Austria-Hungary within 15 days or interned, and the Allies were to have free use of Austria-Hungary's internal communications and to take possession of most of its warships.

Count Mihály Károlyi, chairman of the Budapest National Council, had been appointed prime minister of Hungary by his king, the Austrian emperor Charles, on October 31 but had promptly started to dissociate his country from Austria—partly in the vain hope of obtaining a separate Hungarian armistice. Charles, the last Habsburg to rule in Austria-Hungary, renounced the right to participate in Austrian affairs of government on November 11, in Hungarian affairs on November 13.

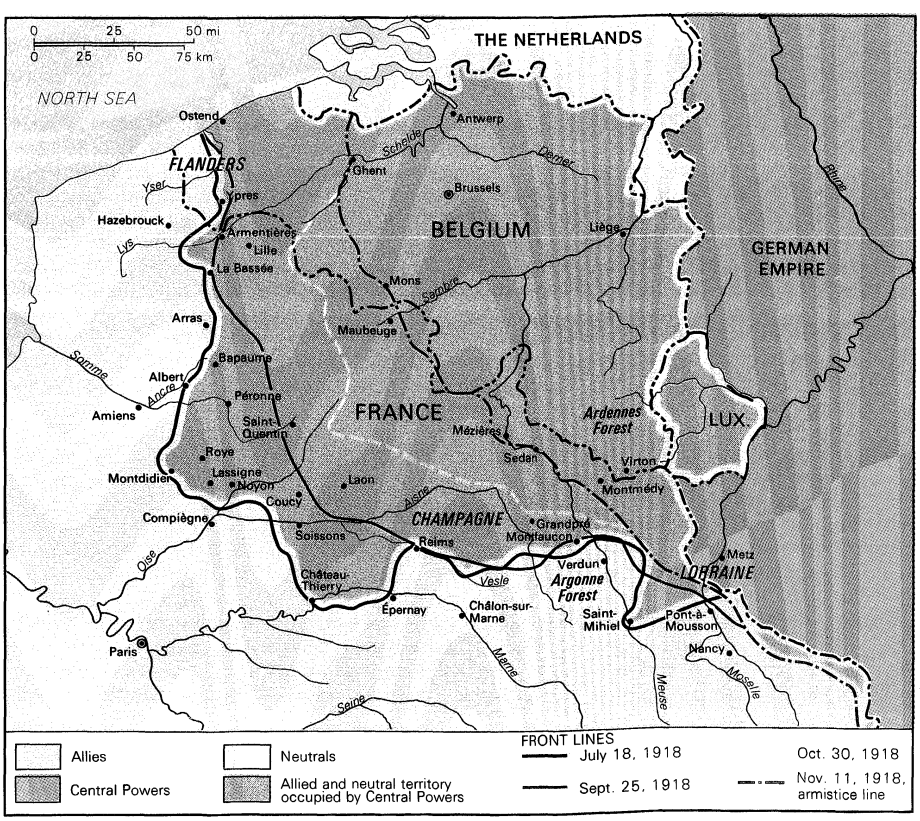
THE FINAL OFFENSIVE ON THE WESTERN FRONT

It was eventually agreed among the Allied commanders that Pershing's American troops should advance across the difficult terrain of the Argonne Forest, so that the combined Allied offensive would consist of converging attacks against the whole German position west of a line drawn from Ypres to Verdun. Thus, the Americans from the front northwest of Verdun and the French from eastern Champagne, the former on the west bank of the Meuse, the latter west of the Argonne Forest, were to launch attacks on September 26, with Mézières as their objective, in order to threaten not only the Germans' supply line along the Mézières–Sedan–Montmédy railway and the natural line of retreat across Lorraine but also the hinge of the Antwerp–Meuse defensive line that the Germans were now preparing. The British were to attack the Hindenburg Line between Cambrai and Saint-Quentin on September 27 and to try to reach the key rail junction of Maubeuge, so as to threaten the Germans' line of retreat through the Liège gap. The Belgians, with Allied support, were to begin a drive from Ypres toward Ghent on September 28.

The Americans took Vauquois and Montfaucon in the first two days of their offensive but were soon slowed down, and on October 14, when their attack was suspended, they had only reached Grandpré, less than halfway to Mézières. The French advance meanwhile was halted on the Aisne. The British, though they had broken through the German defenses by October 5 and thenceforward had open country in front of them, could not pursue the Germans fast enough to endanger their withdrawal. Nevertheless, the piercing of the Hindenburg Line unnerved the German supreme command. The Belgians were in possession of all the heights around Ypres by September 30.

The end of the German War. Georg von Hertling, who had taken the place of Michaelis as Germany's chancellor in November 1917 but had proved no more capable than he of restraining Ludendorff and Hindenburg, tendered his resignation on Sept. 29, 1918, the day of the Bulgarian armistice and of the major development of the British attack on the Western Front. Pending the appointment of a new chancellor, Ludendorff and Hindenburg obtained the Emperor's consent to an immediate peace move. On October 1 they even disclosed their despondency to a meeting of the leaders of all the national political parties, thus undermining the German home front by a sudden revelation of facts long hidden from the public and its civilian leaders. This new and bleak honesty about Germany's deteriorating military situation gave an immense impetus to the native German forces of pacifism and internal discord. On October 3 the new chancellor was appointed: he was Prince Maximilian of Baden, internationally known for his moderation and honorability. Though Max demanded a few days' interval lest Germany's overture for peace should appear too obviously an admission of imminent collapse, the military leaders insisted on an immediate move. A German note to Wilson, requesting an armistice and negotiations on the basis of Wilson's own pronouncements, was sent off in the night of October 3–4.

Prince Max as chancellor



Allied offensives, July–November 1918.

The U.S. answer of October 8 required Germany's preliminary assent (1) to negotiations on the sole question of the means of putting Wilson's principles into practice and (2) to the withdrawal of German forces from Allied soil. The German government's note of October 12 accepted these requirements and suggested a mixed commission to arrange the postulated evacuation. On October 14, however, the U.S. government sent a second note, which coupled allusions to Germany's "illegal and inhuman" methods of warfare with demands that the conditions of the armistice and of the evacuation be determined unilaterally by its own and the Allies' military advisers and that the "arbitrary power" of the German regime be removed in order that the forthcoming negotiations could be conducted with a government representative of the German people.

By this time the German supreme command had become more cheerful, even optimistic, as it saw that the piercing of the Hindenburg Line had not been followed by an actual Allied breakthrough. More encouragement came from reports of a slackening in the force of the Allies' attacks, largely because they had advanced too far ahead of their supply lines. Ludendorff still wanted an armistice, but only to give his troops a rest as a prelude to further resistance and to ensure a secure withdrawal to a shortened defensive line on the frontier. By October 17 he even felt that his troops could do without a rest. It was less that the situation had changed than that his impression of it had been revised; it had never been quite so bad as he had pictured it on September 29. But his dismal first impression had now spread throughout German political circles and the public. Though they had endured increasing privations and were half-starved due to the Allied blockade by mid-1918, the German people had retained their morale surprisingly well as long as they believed Germany had a prospect of achieving victory on the Western Front. When this hope collapsed in October 1918, many, and perhaps even most, Germans wished only that the war would end, though it might mean their nation would have to accept unfavourable peace terms. German public opinion, having been more suddenly disillusioned, was now far more radically defeatist than the supreme command.

A third German note to the United States, sent on October 20, agreed to the unilateral settlement of conditions for the armistice and for the evacuation, in the express belief that Wilson would allow no affront to Germany's honour. The answering U.S. note of October 23 conceded Wilson's readiness to propose an armistice to the Allies but added that the terms must be such as to make Germany incapable of renewing hostilities. Ludendorff saw this, militarily, as a demand for unconditional surrender and would therefore have continued resistance. But the situation had passed beyond his control, and on October 26 he was made to resign by the Emperor, on Prince Max's advice. On October 27 Germany acknowledged the U.S. note.

Wilson now began to persuade the Allies to agree to an armistice and negotiations according to the U.S.–German correspondence. They agreed, with two reservations: they would not subscribe to the second of the Fourteen Points (on the freedom of the seas); and they wanted "compensation . . . for damage done to the civilian population . . . and their property by the aggression of Germany." Wilson's note of November 5 apprised the Germans of these reservations and stated that Foch would communicate armistice terms to Germany's accredited representatives. On November 8 a German delegation, led by Matthias Erzberger, arrived at Rethondes, in the Forest of Compiègne, where the Germans met face to face with Foch and his party and were informed of the Allies' peace terms.

The German revolution

Meanwhile, revolution was shaking Germany. It began with a sailors' mutiny at Kiel on October 29 in reaction to the naval command's order for the High Seas Fleet to go out into the North Sea for a conclusive battle. Though the U-boat crews remained loyal, the mutiny of the surface-ship crews spread to other units of the fleet, developed into armed insurrection on November 3, and progressed to open revolution the next day. There were disturbances in Hamburg and in Bremen; "councils of soldiers and workers," like the Russian soviets, were formed in inland industrial centres; and in the night of November 7–8 a "democratic and socialist Republic of Bavaria" was proclaimed. The Social Democrats of the Reichstag withdrew their support from Prince Max's government in order to be free to contend against the Communists for the leadership of the revolution. While William II, at Spa, was still wondering whether he could abdicate his imperial German title but remain king of Prussia, Prince Max, in Berlin on November 9, on his own initiative, announced William's abdication of both titles. The Hohenzollern monarchy thus came to an end, joining those of the Habsburgs and the Romanovs. Prince Max handed his powers as chancellor over to Friedrich Ebert, a Majority Social Democrat, who formed a provisional government. A member of this government, Philipp Scheidemann, hastily proclaimed a republic. On November 10 William II took refuge in the neutral Netherlands, where on November 28 he signed his own abdication of his sovereign rights.

The Armistice. The Allies' armistice terms presented in the railway carriage at Rethondes were stiff. Germany was required to evacuate not only Belgium, France, and Alsace-Lorraine but also all the rest of the left (west) bank of the Rhine, and it had to neutralize that river's right bank between The Netherlands and Switzerland. The German troops in East Africa were to surrender; the German armies in eastern Europe were to withdraw to the prewar German frontier; the treaties of Brest-Litovsk and Bucharest were to be annulled; and the Germans were to repatriate all prisoners of war and hand over to the Allies a large quantity of war materials, including 5,000 pieces of artillery, 25,000 machine guns, 1,700 aircraft, 5,000 locomotives, and 150,000 railroad cars. And meanwhile, the Allies' blockade of Germany was to continue.

Pleading the danger of Bolshevism in a nation on the verge of collapse, the German delegation obtained some mitigation of these terms: a suggestion that the blockade might be relaxed, a reduction in the quantity of armaments to be handed over, and permission for the German forces in eastern Europe to stay put for the time being. The Germans might have held out longer for further concessions if the fact of revolution on their home front had not been coupled with the imminence of a new blow from the west.

Though the Allied advance was continuing and seemed in some sectors even to be accelerating, the main German forces had managed to retreat ahead of it. The Germans' destruction of roads and railways along the routes of their evacuation made it impossible for supplies to keep pace with the advancing Allied troops; a pause in the advance would occur while Allied communications were being repaired, and that would give the Germans a breathing space in which to rally their resistance. By November 11 the Allied advance on the northern sectors of the front had come more or less to a standstill on a line running from Pont-à-Mousson through Sedan, Mézières, and Mons to Ghent. Foch, however, now had a Franco-U.S. force of 28 divisions and 600 tanks in the south ready to strike through Metz into northeastern Lorraine. Since Foch's general offensive had absorbed the Germans' reserves, this new offensive would fall on their bared left flank and held the promise of outflanking their whole new line of defense (from Antwerp to the line of the Meuse) and of intercepting any German retreat. By this time the number of U.S. divisions in France had risen to 42. In addition, the British were about to bomb Berlin on a scale hitherto unattempted in air warfare.

Whether the Allies' projected final offensive, intended for November 14, would have achieved a breakthrough can never be known. At 5:00 AM on Nov. 11, 1918, the Armistice document was signed in Foch's railway carriage at Rethondes. At 11:00 AM on the same day, World War I came to an end.

The fact that Matthias Erzberger, who was a civilian politician rather than a soldier, headed the German armistice delegation became an integral part of the legend of the "stab in the back" (*Dolchstoss im Rücken*). This legend's theme was that the German Army was "undefeated

Table 4: Armed Forces Mobilized and Casualties in World War I*

country	total mobilized forces	killed and died	wounded	prisoners and missing	total casualties	percentage of mobilized forces in casualties
Allied and Associated Powers						
Russia	12,000,000	1,700,000	4,950,000	2,500,000	9,150,000	76.3
British Empire	8,904,467	908,371	2,090,212	191,652	3,190,235	35.8
France	8,410,000	1,357,800	4,266,000	537,000	6,160,800	73.3
Italy	5,615,000	650,000	947,000	600,000	2,197,000	39.1
United States	4,355,000	116,516	204,002	4,500	323,018	8.1
Japan	800,000	300	907	3	1,210	0.2
Romania	750,000	335,706	120,000	80,000	535,706	71.4
Serbia	707,343	45,000	133,148	152,958	331,106	46.8
Belgium	267,000	13,716	44,686	34,659	93,061	34.9
Greece	230,000	5,000	21,000	1,000	27,000	11.7
Portugal	100,000	7,222	13,751	12,318	33,291	33.3
Montenegro	50,000	3,000	10,000	7,000	20,000	40.0
Total	42,188,810	5,142,631	12,800,706	4,121,090	22,064,427	52.3
Central Powers						
Germany	11,000,000	1,773,700	4,216,058	1,152,800	7,142,558	64.9
Austria-Hungary	7,800,000	1,200,000	3,620,000	2,200,000	7,020,000	90.0
Turkey	2,850,000	325,000	400,000	250,000	975,000	34.2
Bulgaria	1,200,000	87,500	152,390	27,029	266,919	22.2
Total	22,850,000	3,386,200	8,388,448	3,629,829	15,404,477	67.4
Grand total	65,038,810	8,528,831	21,189,154	7,750,919	37,468,904	57.5

*As reported by the U.S. War Department in February 1924. U.S. casualties as amended by the Statistical Services Center, Office of the Secretary of Defense, Nov. 7, 1957.

in the field" (*unbesiegt im Felde*) and had been "stabbed in the back"—*i.e.,* had been denied support at the crucial moment by a weary and defeatist civilian population and their leaders. This theme was adopted soon after the war's end by Ludendorff himself and by other German generals who were unwilling to admit the hopelessness of Germany's military situation in November 1918 and who wanted to vindicate the honour of German arms. The "stab in the back" legend soon found its way into German historiography and was picked up by German right-wing political agitators who claimed that Allied propaganda in Germany in the last stages of the war had undermined civilian morale and that traitors among the politicians had been at hand ready to do the Allies' bidding by signing the Armistice. Adolf Hitler eventually became the foremost of these political agitators, branding Erzberger and the leaders of the Social Democrats as the "November criminals" and advocating militaristic and expansionist policies by which Germany could redeem its defeat in the war, gain vengeance upon its enemies, and become the preeminent power in Europe.

KILLED, WOUNDED, AND MISSING

The casualties suffered by the participants in World War I dwarfed those of previous wars: some 8,500,000 soldiers died as a result of wounds and/or disease. The greatest number of casualties and wounds were inflicted by artillery, followed by small arms, and then by poison gas. The bayonet, which was relied on by the prewar French Army as the decisive weapon, actually produced few casualties. War was increasingly mechanized from 1914 and produced casualties even when nothing important was happening. On even a quiet day on the Western Front, many hundreds of Allied and German soldiers died. The heaviest loss of life for a single day occurred on July 1, 1916, during the Battle of the Somme, when the British Army suffered 57,470 casualties.

Sir Winston Churchill once described the battles of the Somme and Verdun, which were typical of trench warfare in their futile and indiscriminate slaughter, as being waged between double or triple walls of cannons fed by mountains of shells. In an open space surrounded by masses of these guns large numbers of infantry divisions collided. They fought in this dangerous position until battered into a state of uselessness. Then they were replaced by other divisions. So many men were lost in the process and shattered beyond recognition that there is a French monument at Verdun to the 150,000 unlocated dead who are assumed to be buried in the vicinity.

This kind of war made it difficult to prepare accurate casualty lists. There were revolutions in four of the warring countries in 1918, and the attention of the new governments was shifted away from the grim problem of war losses. A completely accurate table of losses may never be compiled. The best available estimates of World War I military casualties are assembled in Table 4.

Similar uncertainties exist about the number of civilian deaths attributable to the war. There were no agencies established to keep records of these fatalities, but it is clear that the displacement of peoples through the movement of the war in Europe and in Asia Minor, accompanied as it was in 1918 by the most destructive outbreak of influenza in history, led to the deaths of large numbers. It has been estimated that the number of civilian deaths attributable to the war was higher than the military casualties, or around 13,000,000. These civilian deaths were largely caused by starvation, exposure, disease, military encounters, and massacres.

WORLD WAR II

Axis initiative and Allied reaction

The outbreak of war. By the early part of 1939 the German dictator Adolf Hitler had become determined to invade and occupy Poland. Poland, for its part, had guarantees of French and British military support should it be attacked by Germany. Hitler intended to invade Poland anyway, but first he had to neutralize the possibility that the Soviet Union would resist the invasion of its western neighbour. Secret negotiations led on August 23–24 to the signing of the German–Soviet Nonaggression Pact in Moscow. In a secret protocol of this pact, the Germans and the Soviets agreed that Poland should be divided between them, with the western third of the country going to Germany and the eastern two-thirds being taken over by the U.S.S.R.

Having achieved this cynical agreement, the other provisions of which stupefied Europe even without divulgence of the secret protocol, Hitler thought that Germany could attack Poland with no danger of Soviet or British intervention and gave orders for the invasion to start on August 26. News of the signing, on August 25, of a formal treaty of mutual assistance between Great Britain and Poland (to supersede a previous though temporary agreement)

caused him to postpone the start of hostilities for a few days. He was still determined, however, to ignore the diplomatic efforts of the western powers to restrain him. Finally, at 12:40 PM on Aug. 31, 1939, Hitler ordered hostilities against Poland to start at 4:45 the next morning. The invasion began as ordered. In response, Great Britain and France declared war on Germany on September 3, at 11:00 AM and at 5:00 PM, respectively. World War II had begun.

Forces and resources of the European combatants, 1939. In September 1939 the Allies, namely Great Britain, France, and Poland, were together superior in industrial resources, population, and military manpower, but the German Army, or Wehrmacht, because of its armament, training, doctrine, discipline, and fighting spirit, was the most efficient and effective fighting force for its size in the world. The index of military strength in September 1939 was the number of divisions that each nation could mobilize. Against Germany's 100 infantry divisions and six armoured divisions, France had 90 infantry divisions in metropolitan France, Great Britain had 10 infantry divisions, and Poland had 30 infantry divisions, 12 cavalry brigades, and one armoured brigade (Poland had also 30 reserve infantry divisions, but these could not be mobilized quickly). A division contained from 12,000 to 25,000 men.

It was the qualitative superiority of the German infantry divisions and the number of their armoured divisions that made the difference in 1939. The firepower of a German infantry division far exceeded that of a French, British, or Polish division; the standard German division included 442 machine guns, 135 mortars, 72 antitank guns, and 24 howitzers. Allied divisions had a firepower only slightly greater than that of World War I. Germany had six armoured divisions in September 1939; the Allies, though they had a large number of tanks, had no armoured divisions at that time.

The six armoured, or panzer, divisions of the Wehrmacht comprised some 2,400 tanks. And though Germany would subsequently expand its tank forces during the first years of the war, it was not the number of tanks that Germany had (the Allies had almost as many in September 1939) but the fact of their being organized into divisions and operated as such that was to prove decisive. In accordance with the doctrines of General Heinz Guderian, the German tanks were used in massed formations in conjunction with motorized artillery to punch holes in the enemy line and to isolate segments of the enemy, which were then surrounded and captured by motorized German infantry divisions while the tanks ranged forward to repeat the process: deep drives into enemy territory by panzer divisions were thus followed by mechanized infantry and foot soldiers. These tactics were supported by dive bombers that attacked and disrupted the enemy's supply and communications lines and spread panic and confusion in its rear, thus further paralyzing its defensive capabilities. Mechanization was the key to the German blitzkrieg, or "lightning war," so named because of the unprecedented speed and mobility that were its salient characteristics. Tested and well-trained in maneuvers, the German panzer divisions constituted a force with no equal in Europe.

The German Air Force, or Luftwaffe, was also the best force of its kind in 1939. It was a ground-cooperation force designed to support the Army, but its planes were superior to nearly all Allied types. In the rearmament period from 1935 to 1939 the production of German combat aircraft steadily mounted. Table 5 shows the production of German aircraft by years.

Table 5: German Aircraft Production by Years

year	combat types	other types
1933	0	368
1934	840	1,128
1935	1,823	1,360
1936	2,530	2,582
1937	2,651	2,955
1938	3,350	1,885
1939	4,733	3,562

The standardization of engines and airframes gave the Luftwaffe an advantage over its opponents. Germany had an operational force of 1,000 fighters and 1,050 bombers in September 1939. The Allies actually had more planes in 1939 than Germany did, but their strength was made up of many different types, some of them obsolescent. Table 6 shows the number of first-line military aircraft available to the Allies at the outbreak of war.

Table 6: Allied Air Strength, September 1939

aircraft	British	French	Polish
Bombers	536	463	200
Fighters	608	634	300
Reconnaissance	96	444	—
Coastal command	216	—	—
Fleet air arm	204	194	—

Great Britain, which was held back by delays in the rearmament program, was producing one modern fighter in 1939, the Hurricane. A higher-performance fighter, the Spitfire, was just coming into production and did not enter the air war in numbers until 1940.

The value of the French Air Force in 1939 was reduced by the number of obsolescent planes in its order of battle: 131 of the 634 fighters and nearly all of the 463 bombers. France was desperately trying to buy high-performance aircraft in the United States in 1939.

At sea the odds against Germany were much greater in September 1939 than in August 1914, since the Allies in 1939 had many more large surface warships than Germany had. At sea, however, there was to be no clash between the Allied and the German massed fleets but only the individual operation of German pocket battleships and commerce raiders.

Technology of war, 1918–39. When World War I ended, the experience of it seemed to vindicate the power of the defensive over the offensive. It was widely believed that a superiority in numbers of at least three to one was required for a successful offensive. Defensive concepts underlay the construction of the Maginot Line between France and Germany and of its lesser counterpart, the Siegfried Line, in the interwar years. Yet by 1918 both of the requirements for the supremacy of the offensive were at hand: tanks and planes. The battles of Cambrai (1917) and Amiens (1918) had proved that when tanks were used in masses, with surprise, and on firm and open terrain, it was possible to break through any trench system.

The Germans learned this crucial, though subtle, lesson from World War I. The Allies on the other hand felt that their victory confirmed their methods, weapons, and leadership, and in the interwar period the French and British armies were slow to introduce new weapons, methods, and doctrines. Consequently, in 1939 the British Army did not have a single armoured division, and the French tanks were distributed in small packets throughout the infantry divisions. The Germans, by contrast, began to develop large tank formations on an effective basis after their rearmament program began in 1935.

In the air the technology of war had also changed radically between 1918 and 1939. Military aircraft had increased in size, speed, and range, and for operations at sea, aircraft carriers were developed that were capable of accompanying the fastest surface ships. Among the new types of planes developed was the dive bomber, a plane designed for accurate low-altitude bombing of enemy strong points as part of the tank-plane-infantry combination. Fast low-wing monoplane fighters were developed in all countries; these aircraft were essentially flying platforms for eight to 12 machine guns installed in the wings. Light and medium bombers were also developed that could be used for the strategic bombardment of cities and military strongpoints. The threat of bomber attacks on both military and civilian targets led directly to the development of radar in England. Radar made it possible to determine the location, the distance, and the height and speed of a distant aircraft no matter what the weather was. By December 1938 there were five radar stations established on the coast of England, and 15 additional stations were begun. So, when

war came in September 1939, Great Britain had a warning chain of radar stations that could tell when hostile planes were approaching.

THE WAR IN EUROPE, 1939–41

The campaign in Poland, 1939. The German conquest of Poland in September 1939 was the first demonstration in war of the new theory of high-speed armoured warfare that had been adopted by the Germans when their rearmament began. Poland was a country all too well suited for such a demonstration. Its frontiers were immensely long—about 3,500 miles in all; and the stretch of 1,250 miles adjoining German territory had recently been extended to 1,750 miles in all by the German occupation of Bohemia-Moravia and of Slovakia, so that Poland's southern flank became exposed to invasion—as the northern flank, facing East Prussia, already was. Western Poland had become a huge salient that lay between Germany's jaws.

It would have been wiser for the Polish Army to assemble farther back, behind the natural defense line formed by the Vistula and San rivers, but that would have entailed the abandonment of some of the most valuable western parts of the country, including the Silesian coalfields and most of the main industrial zone, which lay west of the river barrier. The economic argument for delaying the German approach to the main industrial zone was heavily reinforced by Polish national pride and military overconfidence.

When war broke out the Polish Army was able to mobilize about 1,000,000 men, a fairly large number. The Polish Army was woefully outmoded, however, and was almost completely lacking in tanks, armoured personnel carriers, and antitank and antiaircraft guns. Yet many of the Polish military leaders clung to the double belief that their preponderance of horsed cavalry was an important asset and that they could take the offensive against the German mechanized forces. They also tended to discount the effect of Germany's vastly superior air force, which was nearly 10 times as powerful as their own.

The unrealism of such an attitude was repeated in the Polish Army's dispositions. Approximately one-third of Poland's forces were concentrated in or near the Polish Corridor (in northeastern Poland), where they were perilously exposed to a double envelopment—from East Prussia and the west combined. In the south, facing the main avenues of a German advance, the Polish forces were thinly spread. At the same time, nearly another one-third of Poland's forces were massed in reserve in the north-central part of the country, between Łódź and Warsaw, under the commander in chief, Marshal Edward Rydz-Śmigły. The Poles' forward concentration in general forfeited their chance of fighting a series of delaying actions, since their foot-marching army was unable to retreat to their defensive positions in the rear or to man them before being overrun by the invader's mechanized columns.

The 40-odd infantry divisions employed by the Germans in the invasion counted for much less than their 14 mechanized or partially mechanized divisions: these consisted of six armoured divisions; four light divisions, consisting of motorized infantry (infantry wholly transported by trucks and personnel carriers) with two armoured units; and four motorized divisions. The Germans attacked with about 1,500,000 troops in all. It was the deep and rapid thrusts of these mechanized forces that decided the issue, in conjunction with the overhead pressure of the Luftwaffe, which wrecked the Polish railway system and destroyed most of the Polish Air Force before it could come into action. The Luftwaffe's terror-bombing of Polish cities, bridges, roads, rail lines, and power stations completed the disorganization of the Polish defenses.

On Sept. 1, 1939, the German attack began. Against northern Poland, General Fedor von Bock commanded an army group comprising General Georg von Küchler's 3rd Army, which struck southward from East Prussia, and General Günther von Kluge's 4th Army, which struck eastward across the base of the Corridor. Much stronger in troops and in tanks, however, was the army group in the south under General Gerd von Rundstedt, attacking from Silesia and from the Moravian and Slovakian border: General Johannes Blaskowitz's 8th Army, on the left, was to drive eastward against Łódź; General Wilhelm List's 14th Army, on the right, was to push on toward Kraków and to turn the Poles' Carpathian flank; and General Walter von Reichenau's 10th Army, in the centre, with the bulk of the group's armour, was to deliver the decisive blow with a northwestward thrust into the heart of Poland. By September 3, when Kluge in the north had reached the Vistula and Küchler was approaching the Narew River, Reichenau's armour was already beyond the Warta; two days later his left wing was well to the rear of Łódź and his right wing at Kielce; and by September 8 one of his armoured corps was in the outskirts of Warsaw, having advanced 140 miles in the first week of war. Light divisions on Reichenau's right were on the Vistula between Warsaw and Sandomierz by September 9, while List, in the south, was on the San above and below Przemyśl. At the same time, the 3rd Army tanks, led by Guderian, were across the Narew attacking the line of the Bug River, behind Warsaw. All the German armies had made progress in fulfilling their parts in the great enveloping maneuver planned by General Franz Halder, chief of the general staff, and directed by General Walther von Brauchitsch, the commander in chief. The Polish armies were splitting up into uncoordinated fragments, some of which were retreating while others were delivering disjointed attacks on the nearest German columns.

On September 10 the Polish commander in chief, Marshal Edward Rydz-Śmigły, ordered a general retreat to the southeast. The Germans, however, were by that time not only tightening their net around the Polish forces west of the Vistula (in the Łódź area and, still farther west, around Poznań) but also penetrating deeply into eastern Poland. The Polish defense was already reduced to random efforts by isolated bodies of troops when another blow fell: on Sept. 17, 1939, Soviet forces entered Poland from the east. The next day, the Polish government and high command crossed the Romanian frontier on their way into exile. The Warsaw garrison held out against the Germans until September 28, undergoing terror-bombings and artillery barrages that reduced parts of the city to rubble, with no regard for the civilian population. The last considerable fragment of the Polish Army resisted until October 5; and some guerrilla fighting went on into the winter. The Germans took a total of 700,000 prisoners, and about 80,000 Polish soldiers escaped over neutral frontiers. Polish total casualties (killed, wounded, and missing) remain unknown, while the Germans sustained about 45,000 casualties. Poland was conquered for partition between Germany and the U.S.S.R., the forces of which met and greeted each other on Polish soil. On September 28 another secret German–Soviet protocol modified the arrangements of August: all Lithuania was to be a Soviet sphere of influence, not a German one; but the dividing line in Poland was changed in Germany's favour, being moved eastward to the Bug River.

The Soviet invasion of eastern Poland

The Baltic states and the Russo-Finnish War, 1939–40. Profiting quickly from its understanding with Germany, the U.S.S.R. on Oct. 10, 1939, constrained Estonia, Latvia, and Lithuania to admit Soviet garrisons onto their territories. Approached with similar demands, Finland refused to comply, even though the U.S.S.R. offered territorial compensation elsewhere for the cessions that it was requiring for its own strategic reasons. Finland's armed forces amounted to about 200,000 troops in 10 divisions. The Soviets eventually brought about 70 divisions (about 1,000,000 men) to bear in their attack on Finland, along with about 1,000 tanks. Soviet troops attacked Finland on Nov. 30, 1939.

The invaders succeeded in isolating the little Arctic port of Petsamo in the far north but were ignominiously repulsed on all of the fronts chosen for their advance. On the Karelian Isthmus, the massive reinforced-concrete fortifications of Finland's Mannerheim Line blocked the Soviet forces' direct land route from Leningrad into Finland. The Soviet planners had grossly underestimated the Finns' national will to resist and the natural obstacles constituted by the terrain's numerous lakes and forests.

The western powers exulted overtly over the humiliation

of the Soviet Union. One important effect of Finland's early successes was to reinforce the tendency of both Hitler and the western democracies to underestimate the Soviet military capabilities. But in the meantime, the Soviet strategists digested their hard-learned military lessons.

On Feb. 1, 1940, the Red Army launched 14 divisions into a major assault on the Mannerheim Line. The offensive's weight was concentrated along a 10-mile sector of the line near Summa, which was pounded by a tremendous artillery bombardment. As the fortifications were pulverized, tanks and sledge-carried infantry advanced to occupy the ground while the Soviet Air Force broke up attempted Finnish counterattacks. After little more than a fortnight of this methodical process, a breach was made through the whole depth of the Mannerheim Line. Once the Soviets had forced a passage on the Karelian Isthmus, Finland's eventual collapse was certain. On March 6 Finland sued for peace, and a week later the Soviet terms were accepted: the Finns had to cede the entire Karelian Isthmus, Viipuri, and their part of the Rybachy Peninsula to the Soviets. The Finns had suffered about 70,000 casualties in the campaign, the Soviets more than 200,000.

The war in the west, September 1939–June 1940. During their campaign in Poland, the Germans kept only 23 divisions in the west to guard their frontier against the French, who had nearly five times as many divisions mobilized. The French commander in chief, General Maurice-Gustave Gamelin, proposed an advance against Germany through neutral Belgium and The Netherlands, in order to have room to exercise his ponderous military machine. He was overruled, however, and French assaults on the 100-mile stretch of available front along the Franco-German frontier had barely dented the German defenses when the collapse of Poland prompted the recall of Gamelin's advanced divisions to defensive positions in the Maginot Line (see below). From October 1939 to March 1940, successive plans were developed for counteraction in the event of a German offensive through Belgium—all of them based on the assumption that the Germans would come across the plain north of Namur, not across the hilly and wooded Ardennes. The Germans would indeed have taken the route foreseen by the French if Hitler's desire for an offensive in November 1939 had not been frustrated, on the one hand, by bad weather and, on the other, by the hesitations of his generals; but in March 1940 the bold suggestion of General Erich von Manstein that an offensive through the Ardennes should, in fact, be practicable for tank forces was adopted by Hitler, despite orthodox military opinion.

Franco-British and German strategies

Meanwhile, Hitler's immediate outlook had been changed by considerations about Scandinavia. Originally he had intended to respect Norway's neutrality. Then rumours leaked out, prematurely, of British designs on Norway—as, in fact, Winston Churchill, first lord of the Admiralty, was arguing that mines should be laid in Norwegian waters to stop the export of Swedish iron ore from Gällivare to Germany through Norway's rail terminus and port of Narvik. The British Cabinet, in response to Churchill, authorized at least the preparation of a plan for a landing at Narvik; and in mid-December 1939 a Norwegian politician, Vidkun Quisling, leader of a pro-Nazi party, was introduced to Hitler. On Jan. 27, 1940, Hitler ordered plans for an invasion of Norway, for use if he could no longer respect Norway's neutrality.

After France's failure to interrupt the German conquest of Poland, the western powers and the Germans were so inactive with regard to land operations that journalists began to speak derisively, over the next six months, of the "phony war." At sea, however, the period was somewhat more eventful. German U-boats sank the British aircraft carrier *Courageous* (September 17) and the battleship *Royal Oak* (October 14). The U-boats' main warfare, however, was against merchant shipping: they sank more than 110 vessels in the first four months of the war. Both the Germans and the British, meanwhile, were engaged in extensive mine laying.

In surface warfare at sea, the British were on the whole more fortunate than the Germans. A German pocket battleship in the Atlantic, the *Admiral Graf Spee* sank nine ships before coming to a tragic end: having sustained and inflicted damage in an engagement with three British cruisers off the Río de la Plata on Dec. 13, 1939, she made off to Montevideo and obtained leave to spend four days there for repairs; the British mustered reinforcements for the two cruisers still capable of action after the engagement, namely the *Ajax* and the *Achilles,* and brought the *Cumberland* to the scene in time; but, on December 17, when the *Graf Spee* put to sea again, her crew scuttled her a little way out of the harbour before the fight could be resumed.

The *Admiral Graf Spee* sunk

The invasion of Norway. British plans for landings on the Norwegian coast in the third week of March 1940 were temporarily postponed. Prime Minister Neville Chamberlain, however, was by that time convinced that some aggressive action ought to be taken; and Paul Reynaud, who succeeded Daladier as France's premier on March 21, was of the same opinion. (Reynaud had come into office on the surge of the French public's demand for a more aggressive military policy and quicker offensive action against Germany.) It was agreed that mines should be laid in Norwegian waters and that the mining should be followed by the landing of troops at four Norwegian ports, Narvik, Trondheim, Bergen, and Stavanger.

Because of Anglo-French arguments, the date of the mining was postponed from April 5 to April 8. The postponement was catastrophic. Hitler had on April 1 ordered the German invasion of Norway to begin on April 9; so, when on April 8 the Norwegian government was preoccupied with earnest protest about the British mine laying, the German expeditions were well on their way.

The German invasion of Norway

On April 9, 1940, the major Norwegian ports from Oslo northward to Narvik (1,200 miles away from Germany's naval bases) were occupied by advance detachments of German troops. At the same time, a single parachute battalion (the first ever employed in warfare) took the Oslo and Stavanger airfields, and 800 operational aircraft overawed the Norwegian population. Norwegian resistance at Narvik, at Trondheim (the strategic key to Norway), at Bergen, at Stavanger, and at Kristiansand had been overcome very quickly; and Oslo's effective resistance to the seaborne forces was nullified when German troops from the airfield entered the city.

Simultaneously, along with their Norwegian enterprise, the Germans on April 9 occupied Denmark, sending troopships, covered by aircraft, into Copenhagen harbour and marching over the land frontier into Jutland. This occupation was obviously necessary for the safety of their communications with Norway.

Allied troops began to land at Narvik on April 14. Shortly afterward, British troops were landed also at Namsos and at Åndalsnes, to attack Trondheim from the north and from the south, respectively. The Germans, however, landed fresh troops in the rear of the British at Namsos and advanced up the Gudbrandsdal from Oslo against the force at Åndalsnes. By this time the Germans had about 25,000 troops in Norway. By May 2, both Namsos and Åndalsnes were evacuated by the British. The Germans at Narvik held out against five times as many British and French troops until May 27. By that time the German offensive in France had progressed to such an extent that the British could no longer afford any commitment in Norway, and the 25,000 Allied troops were evacuated from Narvik 10 days after their victory. The Norwegian king Haakon VII and his government left Norway for Britain at the same time. Hitler garrisoned Norway with about 300,000 troops for the rest of the war. By occupying Norway, Hitler had ensured the protection of Germany's supply of iron ore from Sweden and had obtained naval and air bases with which to strike at Britain if necessary.

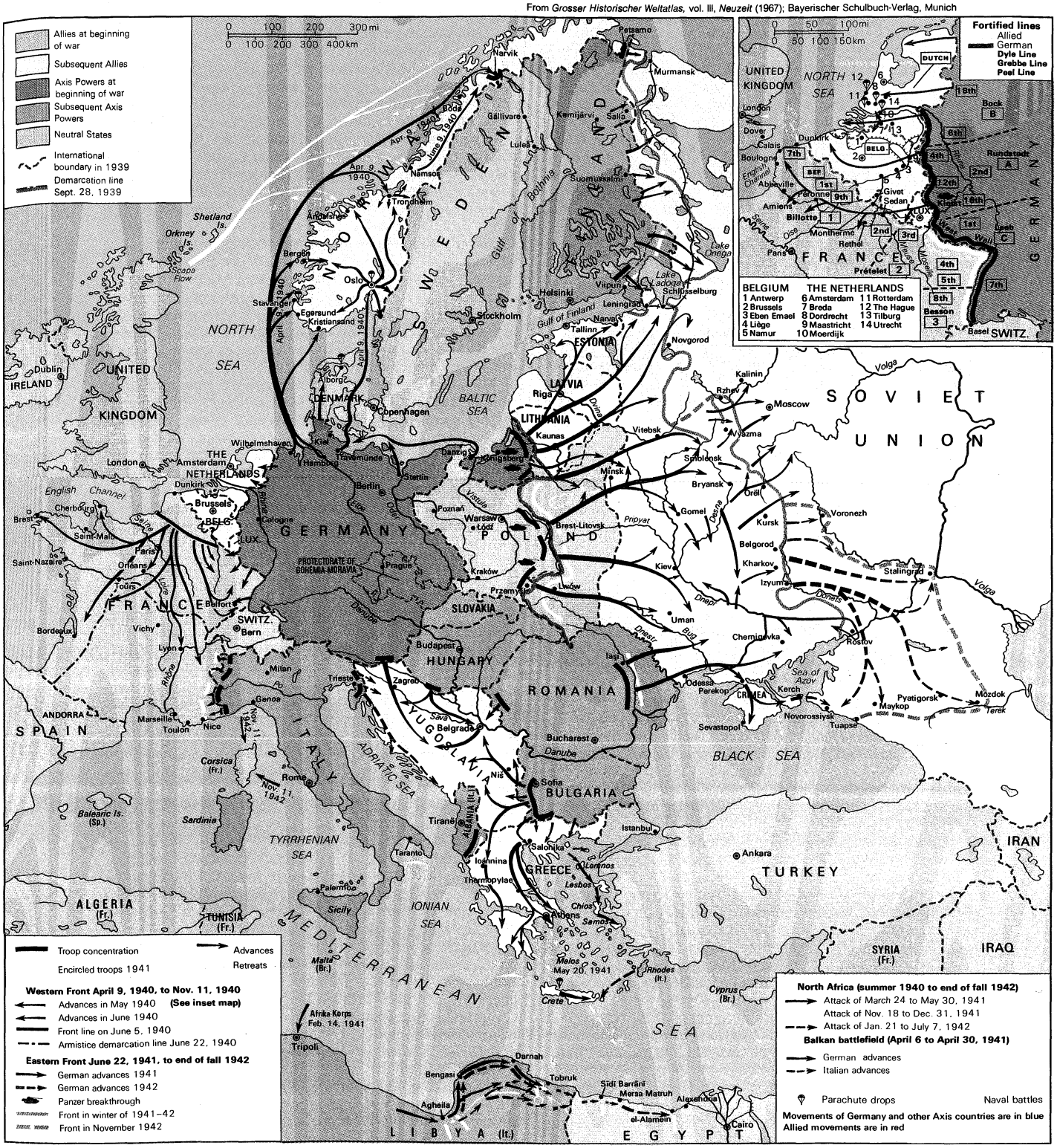
What was to happen in Norway became a less important question for the western powers when, on May 10, 1940, they were surprised by Hitler's long-debated stroke against them through the Low Countries.

The invasion of the Low Countries and France. France's 800,000-man standing army was thought at the time to be the most powerful in Europe. But the French had not progressed beyond the defensive mentality inherited from World War I, and they relied primarily on their

Maginot Line for protection against a German offensive. The Maginot Line was an extremely well-developed chain of fortifications running from the Swiss frontier opposite Basel northward along the left bank of the Rhine and then northwestward no farther than Montmédy, near the Belgian frontier south of the Ardennes Forest. The line consisted of a series of giant pillboxes and other defensive installations constructed in depth, equipped with underground supply and communications facilities, and connected by rail lines, with all its heavy guns pointed east at the German frontier. Depending heavily on the line as a defense against German attack, the French had 41 divisions manning it or backing it, whereas only 39 divisions were watching the long stretch of frontier north of it, from Montmédy through the Ardennes and across Flanders to the English Channel.

In their plan for the invasion of France and the Low Countries, the Germans kept General Wilhelm von Leeb's Army Group C facing the Maginot Line so as to deter the French from diverting forces from it, while launching Bock's Army Group B into the basin of the Lower Maas River north of Liège and Rundstedt's Army Group A

From *Grosser Historischer Weltatlas*, vol. III, *Neuzeit* (1967); Bayerischer Schulbuch-Verlag, Munich



Axis and Allied movements in Europe and North Africa, 1940–42, and (inset) German invasion of the Low Countries and France, 1940.

into the Ardennes. Army Group B comprised Küchler's 18th Army, with one armoured division and airborne support, to attack The Netherlands, and Reichenau's 6th, with two armoured divisions, to advance over the Belgian plain. These two armies would have to deal not only with the Dutch and Belgian armies but also with the forces that the Allies, according to their plan, would send into the Low Countries, namely two French armies and nine British divisions. Rundstedt's Army Group A, however, was much stronger, comprising as it did Kluge's 4th Army, List's 12th, and General Ernst Busch's 16th, with General Maximilian von Weichs's 2nd in reserve, besides a large armoured group under Kleist and a smaller one under General Hermann Hoth, and amounting in all to 44 divisions, seven of them armoured, with 27 divisions in reserve. Army Group A thus amounted to more than 1,500,000 men and more than 1,500 tanks, and it would strike at the weak hinge of the Allies' wheel into Belgium—that is to say, at two French armies, General Charles Huntziger's 2nd and General André Corap's 9th, which together mustered only 12 infantry and four horsed cavalry divisions and stood, respectively, east and west of Sedan on the least-fortified stretch of the French frontier. Against this weak centre of the Allied line were thus massed nearly two-thirds of Germany's forces in the west and nearly three-quarters of its tank forces.

The Dutch Army comprised 10 divisions and the equivalent of 10 more in smaller formations, and thus totaled more than 400,000 men. It apparently had a good chance of withstanding the German invasion, since the attacking German army comprised only seven divisions, apart from the airborne forces it would use. The Dutch, however, had a wide front, a very sensitive and loosely settled rear, very few tanks, and no experience of modern warfare. On May 10, the German attack on The Netherlands began with the capture by parachutists of the bridges at Moerdijk, at Dordrecht, and at Rotterdam and with landings on the airfields around The Hague. On the same day, the weakly held Peel Line, south of the westward-turning arc of the Maas, was penetrated by the German land forces; and on May 11 the Dutch defenders fell back westward past Tilburg to Breda, with the consequence that the French 7th Army, under General Henri Giraud, whose leading forces had sped forward across Belgium over the 140 miles to Tilburg, fell back to Breda likewise. The German tanks thus had a clear road to Moerdijk, and by noon on May 12 they were in the outskirts of Rotterdam. North of the Maas, meanwhile, where the bulk of the Dutch defense was concentrated, the Germans achieved a narrow breach of the Geld Valley line on May 12, whereupon the Dutch, unable to counterattack, retreated to the "Fortress of Holland" Line protecting Utrecht and Amsterdam. Queen Wilhelmina and her government left the country for England on May 13; and the next day the Dutch commander in chief, General Henri Gerard Winkelman, surrendered to the Germans, who had threatened to bomb Rotterdam and Utrecht, as places in the front line of the fighting, if resistance continued. In fact, Rotterdam was bombed, after the capitulation, by 30 planes through a mistake in the Germans' signal communications.

The news of the German onslaught in the Low Countries, dismaying as it was to the Allies, had one effect that was to be of momentous importance to their fortunes: Chamberlain, whose halfhearted conduct of the war had been bitterly criticized in the House of Commons during the debate of May 7–8 on the campaign in Norway, resigned office in the evening of May 10 and was succeeded as prime minister by Churchill, who formed a coalition government.

For the first phase of the invasion of the Belgian plain north of Liège, Reichenau had four army corps, one armoured corps, and only 500 airborne troops; but he also had massive cooperation from the German Luftwaffe, whose dive bombers and fighters played a major role in breaking down the Belgian defenses. West of the Maastricht "appendix" of indefensible Dutch territory separating Belgium from Germany, the fortress of Eben Emael, immediately opposite Maastricht, and the line of the Albert Canal constituted the Belgians' foremost defensive position. On May 10 German airborne troops landed in gliders on the top of the fortress and on bridges over the canal. On May 11 the Belgian front was broken, the German tanks running on westward and some of the infantry turning southward to take Liège from the rear, while the Belgians made a general retreat to the Antwerp–Namur, or Dyle, Line. French and British divisions had just arrived on this Dyle Line, and General René Prioux's two tank divisions went out from it to challenge the German advance. After a big battle on May 14, however, Prioux's tanks had to retire to the consolidated Dyle Line; and on May 15, notwithstanding a successful defense against a German attack, Gamelin ordered the abandonment of the position, because events farther to the south had made it strategically untenable.

The chances for success of the German offensive against France hinged on a German advance through the hilly and dense Ardennes Forest, which the French considered to be impassable to tanks. But the Germans did succeed in moving their tank columns through that difficult belt of country by means of an amazing feat of staff work. While the armoured divisions used such roads through the forest as were available, infantry divisions started alongside them by using field and woodland paths and marched so fast across country that the leading ones reached the Meuse River only a day after the armoured divisions had.

The German breakthrough into France

The decisive operations in France were those of Rundstedt's Army Group A. Kleist's tanks on May 10 took only three hours to cover the 30 miles from the eastern border of independent Luxembourg to the southeastern border of Belgium; and on May 11 the French cavalry divisions that had ridden forward into the Ardennes to oppose them were thrown back over the Semois River. By the evening of May 12 the Germans were across the Franco-Belgian frontier and overlooking the Meuse River. The defenses of this sector were rudimentary, and it was the least-fortified stretch of the whole French front. Worse still, the defending French 2nd and 9th armies had hardly any antitank guns or antiaircraft artillery with which to slow down the German armoured columns and shoot down their dive bombers. Such was the folly of the French belief that a German armoured thrust through the Ardennes was unlikely.

On May 13 Kleist's forces achieved a threefold crossing of the Meuse River. At Sedan wave after wave of German dive bombers swooped on the French defenders of the south bank. The latter could not stand the nerve-racking strain, and the German troops were able to push across the river in rubber boats and on rafts. The tremendous air bombardment was the decisive factor in the crossings. A thousand aircraft supported Kleist's forces, while only a few French aircraft intervened in a gallant but hopeless effort to aid their troops on the ground. Next day, after the tanks had been brought across, Guderian widened the Sedan bridgehead and beat off French counterattacks. On May 15 he broke through the French defenses into open country, turning westward in the direction of the English Channel. On May 16 his forces swept on west for nearly 50 miles. His superiors tried to put on the brake, feeling that such rapid progress was hazardous, but the pace of the German drive upset the French far more, and their collapse spread as Reinhardt's corps joined in the pressure. When more German tanks crossed the Meuse between Givet and Namur, the breach of the French front was 60 miles wide.

Driving westward down the empty corridor between the Sambre and the Aisne rivers, Guderian's tanks crossed the Oise River on May 17 and reached Amiens two days later. Giraud, who on May 15 had superseded Corap in command of the French 9th Army, was thus frustrated in his desperate plan of checking the Germans on the Oise; and Kleist, meanwhile, by lining the Aisne progressively with tanks until the infantry came up to relieve them, was protecting the southwestern flank of the advance against the danger of a counteroffensive from the south. Indeed, when the Germans, on May 15, were reported to be crossing the Aisne River between Rethel and Laon, Gamelin told Reynaud that he had no reserves in that sector and that Paris might fall within two days' time.

Thereupon Reynaud, though he postponed his immediate decision to move the government to Tours, summoned General Maxime Weygand from Syria to take Gamelin's place as commander in chief; but Weygand did not arrive until May 19.

Guderian's tanks were at Abbeville on May 20, and on May 22 he turned northward to threaten Calais and Dunkirk, while Reinhardt, swinging south of the British rear at Arras, headed for the same objectives, the remaining ports by which the British Expeditionary Force (BEF) could be evacuated.

The evacuation from Dunkirk. For the Allies, all communication between their northern and southern forces was severed by the arc of the westward German advance from the Ardennes to the Somme. The Allied armies in the north, having fallen back from the Dyle Line to the Escaut (Schelde), were being encircled, and already on May 19 the British commander, Viscount Gort, was considering the withdrawal of the BEF by sea. On May 21, however, to satisfy orders from London for more positive action, he launched an attack from Arras southward against the right flank of the Germans' corridor; but, though it momentarily alarmed the German high command, this small counterstroke lacked the armoured strength necessary for success. Meanwhile, Guderian's tanks had swept up past Boulogne and Calais and were crossing the canal defense line close to Dunkirk when, on May 24, an inexplicable order from Hitler not only stopped their advance but actually called them back to the canal line just as Guderian was expecting to drive into Dunkirk.

Dunkirk was now the only port left available for the withdrawal of the mass of the BEF from Europe, and the British Cabinet at last decided to save what could be saved. The British Admiralty had been collecting every kind of small craft it could find to help in removing the troops, and the British retreat to the coast now became a race to evacuate the troops before the Germans could occupy Dunkirk. Evacuation began on May 26 and became still more urgent the next day, when the Belgians, their right wing and their centre broken by Reichenau's advance, sued for an armistice. On May 27, likewise, bombing by the Luftwaffe put the harbour of Dunkirk out of use, so that many of the thousands of men thronging the 10-mile stretch of beaches had to be ferried out to sea by petty craft pressed into service by the Royal Navy and manned largely by amateur seamen, though the harbour's damaged breakwater still offered a practicable exit for the majority. By June 4, when the operation came to an end, 198,000 British and 140,000 French and Belgian troops had been saved; but virtually all of their heavy equipment had to be abandoned, and, of the 41 destroyers participating, six were sunk and 19 others damaged. The men who were saved represented a considerable part of the experienced troops possessed by Great Britain and were an inestimable gain to the Allies. The success of the near-miraculous evacuation from Dunkirk was due, on the one hand, to fighter cover by the Royal Air Force from the English coast and on the other to Hitler's fatal order of May 24 halting Guderian. That order had been made for several reasons, chiefly: Hermann Göring, head of the Luftwaffe, had mistakenly assured Hitler that his aircraft alone could destroy the Allied troops trapped on the beaches at Dunkirk; and Hitler himself seems to have believed that Great Britain might accept peace terms more readily if its armies were not constrained into humiliating surrender. Three days passed before Brauchitsch, the German Army commander in chief, was able to persuade Hitler to withdraw his orders and allow the German armoured forces to advance on Dunkirk. But they met stronger opposition from the British, who had had time to solidify their defenses, and almost immediately Hitler stopped the German armoured forces again, ordering them instead to move south and prepare for the attack on the Somme–Aisne line.

The British evacuation from Dunkirk

The campaign in northern France was wound up by Küchler's forces, after both Guderian and Reichenau had been ordered southward. Altogether, the Germans had taken more than 1,000,000 prisoners in three weeks, at a cost of 60,000 casualties. Some 220,000 Allied troops, however, were rescued by British ships from France's northwestern ports (Cherbourg, Saint-Malo, Brest, and Saint-Nazaire), thus bringing the total of Allied troops evacuated to about 558,000.

There remained the French armies south of the Germans' Somme–Aisne front. The French had lost 30 divisions in the campaign so far. Weygand still managed to muster 49 divisions, apart from the 17 left to hold the Maginot Line, but against him the Germans had 130 infantry divisions as well as their 10 divisions of tanks. The Germans, after redisposing their units, began a new offensive on June 5 from their positions on the Somme. The French resisted stiffly for two days, but on June 7 the German tanks in the westernmost sector, led by Major General Erwin Rommel, broke through toward Rouen, and on June 9 they were over the Seine. On June 9 the Germans attacked on the Aisne: the infantry forced the crossings, and then Guderian's armour drove through the breach toward Châlons-sur-Marne before turning eastward for the Swiss frontier, thus isolating all the French forces still holding the Maginot Line.

Italy's entry into the war and the French Armistice. Italy had been unprepared for war when Hitler attacked Poland, but if the Italian leader, Benito Mussolini, was to reap any positive advantages from partnership with Hitler it seemed that Italy would have to abandon its nonbelligerent stance before the western democracies had been defeated by Germany singlehanded. The obvious collapse of France convinced Mussolini that the time to implement his Pact of Steel with Hitler had come, and on June 10, 1940, Italy declared war against France and Great Britain. With about 30 divisions available on their Alpine frontier, the Italians delayed their actual attack on southeastern France until June 20, but it achieved little against the local defense. In any case, the issue in France had already been virtually settled by the victory of Italy's German ally.

Meanwhile, Reynaud had left Paris for Cangé, near Tours; and Weygand, after speaking frankly and despondently to Churchill at the Allied military headquarters at Briare on June 11, told Reynaud and the other ministers at Cangé on June 12 that the battle for France was lost and that a cessation of hostilities was compulsory. There was little doubt that he was correct in this estimate of the military situation: the French armies were now splitting up into fragments. Reynaud's government was divided between the advocates of capitulation and those who, with Reynaud, wanted to continue the war from French North Africa. The only decision that it could make was to move itself from Tours to Bordeaux.

The Germans entered Paris on June 14, 1940, and were driving still deeper southward along both the western and eastern edges of France. Two days later they were in the Rhône Valley. Meanwhile, Weygand was still pressing for an armistice, backed by all the principal commanders. Reynaud resigned office on June 16, whereupon a new government was formed by Marshal Philippe Pétain, the revered and aged hero of the Battle of Verdun in World War I. In the night of June 16 the French request for an armistice was transmitted to Hitler. While discussion of the terms went on, the German advance went on too. Finally, on June 22, 1940, at Rethondes, the scene of the signing of the Armistice of 1918, the new Franco-German Armistice was signed. The Franco-Italian Armistice was signed on June 24. Both armistices came into effect early on June 25.

The Armistice of June 22 divided France into two zones, one to be under German military occupation, one to be left to the French in full sovereignty. The occupied zone comprised all northern France from the northwestern frontier of Switzerland to the Channel and from the Belgian and German frontiers to the Atlantic, together with a strip extending from the lower Loire southward along the Atlantic coast to the western end of the Pyrenees; the unoccupied zone comprised only two-fifths of France's territory, the southeast. The French Navy and Air Force were to be neutralized, but it was not required that they be handed over to the Germans. The Italians granted very generous terms to the French: the only French territory that they claimed to occupy was the small frontier tract that their forces had succeeded in overrunning since June

20. Meanwhile, from June 18, General Charles de Gaulle, whom Reynaud had sent on a military mission to London on June 5, was broadcasting appeals for the continuance of France's war.

The collapse of France in June 1940 posed a severe naval problem to the British, because the powerful French Navy still existed: strategically, it was of immense importance to the British that these French ships not fall into German hands, since they would have tilted the balance of sea power decidedly in favour of the Axis—the Italian Navy being now also at war with Britain. Mistrustful of promises that the French ships would be used only for "supervision and minesweeping," the British decided to immobilize them. Thus, on July 3, 1940, the British seized all French ships in British-controlled ports, encountering only nominal resistance. But when British ships appeared off Mers el-Kébir, near Oran on the Algerian coast, and demanded that the ships of the important French naval force there either join the Allies or sail out to sea, the French refused to submit, and the British eventually opened fire, damaging the battleship *Dunkerque,* destroying the *Bretagne,* and disabling several other vessels. Thereupon, Pétain's government, which on July 1 had installed itself at Vichy, on July 4 severed diplomatic relations with the British. In the eight following days, the constitution of France's Third Republic was abolished and a new French state created, under the supreme authority of Pétain himself. The few French colonies that rallied to General de Gaulle's Free French movement were strategically unimportant.

The Battle of Britain. With France conquered, Hitler could now turn his forces on Germany's sole remaining enemy: Great Britain, which was protected from the formidable German Army by the waters of the English Channel. On July 16, 1940, Hitler issued a directive ordering the preparation and, if necessary, the execution of a plan for the invasion of Great Britain. But an amphibious invasion of Britain would only be possible, given Britain's large navy, if Germany could establish control of the air in the battle zone. To this end, the Luftwaffe chief, Göring, on August 2 issued the "Eagle Day" directive, laying down a plan of attack in which a few massive blows from the air were to destroy British air power and so open the way for the amphibious invasion, termed Operation "Sea Lion." Victory in the air battle for the Luftwaffe would indeed have exposed Great Britain to invasion and occupation. The victory by the Royal Air Force (RAF) Fighter Command blocked this possibility and, in fact, created the conditions for Great Britain's survival, for the extension of the war, and for the eventual defeat of Nazi Germany.

The forces engaged in the battle were relatively small. The British disposed some 600 frontline fighters to defend the country. The Germans made available about 1,300 bombers and dive bombers, and about 900 single-engined and 300 twin-engined fighters. These were based in an arc around England from Norway to the Cherbourg Peninsula in northern coastal France. The preliminaries of the Battle of Britain occupied June and July 1940, the climax August and September, and the aftermath—the so-called Blitz—the winter of 1940–41. In the campaign, the Luftwaffe had no systematic or consistent plan of action: sometimes it tried to establish a blockade by the destruction of British shipping and ports; sometimes, to destroy Britain's Fighter Command by combat and by the bombing of ground installations; and sometimes, to seek direct strategic results by attacks on London and other populous centres of industrial or political significance. The British, on the other hand, had prepared themselves for the kind of battle that in fact took place. Their radar early warning, the most advanced and the most operationally adapted system in the world, gave Fighter Command adequate notice of where and when to direct their fighter forces to repel German bombing raids. The Spitfire, moreover, though still in short supply, was unsurpassed as an interceptor by any fighter in any other air force.

The British fought not only with the advantage—unusual for them—of superior equipment and undivided aim but also against an enemy divided in object and condemned by circumstance and by lack of forethought to fight at a tactical disadvantage. The German bombers lacked the bomb-load capacity to strike permanently devastating blows and also proved, in daylight, to be easily vulnerable to the Spitfires and Hurricanes. Britain's radar, moreover, largely prevented them from exploiting the element of surprise. The German dive bombers were even more vulnerable to being shot down by British fighters, and long-range fighter cover was only partially available from German fighter aircraft, since the latter were operating at the limit of their flying range.

The German air attacks began on ports and airfields along the English Channel, where convoys were bombed and the air battle was joined. In June and July 1940, as the Germans gradually redeployed their forces, the air battle moved inland over the interior of Britain. On August 8 the intensive phase began, when the Germans launched bombing raids involving up to nearly 1,500 aircraft a day and directed them against the British fighter airfields and radar stations. In four actions, on August 8, 11, 12, and 13, the Germans lost 145 aircraft as against the British loss of 88. By late August the Germans had lost more than 600 aircraft, the RAF only 260, but the RAF was losing badly needed fighters and experienced pilots at too great a rate, and its effectiveness was further hampered by bombing damage done to the radar stations. At the beginning of September the British retaliated by unexpectedly launching a bombing raid on Berlin, which so infuriated Hitler that he ordered the Luftwaffe to shift its attacks from Fighter Command installations to London and other cities. These assaults on London, Coventry, Liverpool, and other cities went on unabated for several months. But already, by September 15, on which day the British believed, albeit incorrectly, that they had scored their greatest success by destroying 185 German aircraft, Fighter Command had demonstrated to the Luftwaffe that it could not gain air ascendancy over Britain. This was because British fighters were simply shooting down German bombers faster than German industry could produce them. The Battle of Britain was thus won, and the invasion of England was postponed indefinitely by Hitler. The British had lost more than 900 fighters but had shot down about 1,700 German aircraft.

During the following winter, the Luftwaffe maintained a bombing offensive, carrying out night-bombing attacks on Britain's larger cities. By February 1941 the offensive had declined, but in March and April there was a revival, and nearly 10,000 sorties were flown, with heavy attacks made on London. Thereafter German strategic air operations over England withered.

Central Europe and the Balkans, 1940–41. The continued resistance of the British caused Hitler once more to change his timetable. His great design for a campaign against the U.S.S.R. had originally been scheduled to begin about 1943—by which time he should have secured the German position on the rest of the European continent by a series of "localized" campaigns and have reached some sort of compromise with Great Britain. But in July 1940, seeing Great Britain still undefeated and the United States increasingly inimical to Germany, he decided that the conquest of the European part of the Soviet Union must be undertaken in May 1941, in order both to demonstrate Germany's invincibility to Great Britain and to deter the United States from intervention in Europe (because the elimination of the U.S.S.R. would strengthen the Japanese position in the Far East and in the Pacific). Events in the interval, however, were to make him change his plan once again.

German–Romanian relations

While the invasion of the U.S.S.R. was being prepared, Hitler was much concerned to extend German influence across Slovakia and Hungary into Romania, the oil fields of which he was anxious to secure against Soviet attack and the military manpower of which might be joined to the forces of the German coalition. In May 1940 he obtained an oil and arms pact from Romania; but, when Romania, after being constrained by a Soviet ultimatum in June to cede Bessarabia and northern Bukovina to the U.S.S.R., requested a German military mission and a German guarantee of its remaining frontiers, Hitler refused to comply until the claims of other states against Romania had been met. Romania was compelled to cede

southern Dobruja to Bulgaria on August 21 (an act that was formalized in the Treaty of Craiova on September 7); but its negotiations with Hungary about Transylvania were broken off on August 23. Since, if war had broken out between Romania and Hungary, the U.S.S.R. might have intervened and won control over the oil wells, Hitler decided to arbitrate immediately: by the Vienna Award of August 30, Germany and Italy assigned northern Transylvania, including the Szekler district, to Hungary, and Germany then guaranteed what was left of Romania. In the face of the Romanian nationalists' outcry against these proceedings, the king, Carol II, transferred his dictatorial powers to General Ion Antonescu on Sept. 4, 1940, and abdicated his crown in favour of his young son Michael two days later. Antonescu had already repeated the request for a German military mission, which arrived in Bucharest on October 12.

Though Hitler had apprised the Italian foreign minister, Galeazzo Ciano, of his intention to send a military mission to Romania, Ciano had not apprised Mussolini. So, since the latter's Balkan ambitions had been continually restrained by Hitler, particularly with regard to Yugoslavia, the sudden news of the mission annoyed him. On Oct. 28, 1940, therefore, having given Hitler only the barest hints of his project, Mussolini launched seven Italian divisions (155,000 men) from Albania into a separate war of his own against Greece.

Italy's attack on Greece

The result was exasperating for Hitler. His ally's forces were not only halted by the Greeks, a few miles over the border, on Nov. 8, 1940, but were also driven back by General Alexandros Papagos' counteroffensive of November 14, which was to put the Greeks in possession of one-third of Albania by mid-December. Moreover, British troops landed in Crete, and some British aircraft were sent to bases near Athens, whence they might have attacked the Romanian oil fields. Lastly, the success of the Greeks caused Yugoslavia and Bulgaria, who had hitherto been attentive to overtures from the Axis powers, to revert to a strictly neutral policy.

Anticipating Mussolini's appeal for German help in his "separate" or "parallel" war, Hitler in November 1940 drew Hungary, Romania, and Slovakia successively into the Axis, or Tripartite, Pact that Germany, Italy, and Japan had concluded on September 27 (see below); and he also obtained Romania's assent to the assembling of German troops in the south of Romania for an attack on Greece through Bulgaria. Hungary consented to the transit of these troops through its territory lest Romania take Hungary's place in Germany's favour and so be secured in possession of the Transylvanian lands left to it by the Vienna Award. Bulgaria, however, for fear of Soviet reaction, on the one hand, and of Turkish, on the other (Turkey had massed 28 divisions in Thrace when Italy attacked Greece), delayed its adhesion to the Axis until March 1, 1941. Only thereafter, on March 18, did the Yugoslav regent, Prince Paul, and his ministers Dragiša Cvetković and Aleksandar Cincar-Marković agree to Yugoslavia's adhesion to the Axis.

Meanwhile, the German 12th Army had crossed the Danube from Romania into Bulgaria on March 2, 1941. Consequently, in accordance with a Greco-British agreement of February 21, a British expeditionary force of 58,000 men from Egypt landed in Greece on March 7, to occupy the Olympus–Vermion line. Then, on March 27, 1941, two days after the Yugoslav government's signature, in Vienna, of its adhesion to the Axis Pact, a group of Yugoslav Army officers, led by General Dušan Simović, executed a coup d'état in Belgrade, overthrowing the regency in favour of the 17-year-old king Peter II and reversing the former government's policy.

The Yugoslav coup d'état and the Battle of Cape Matapan

Almost simultaneously with the Belgrade coup d'état, the decisive Battle of Cape Matapan took place between the British and Italian fleets in the Mediterranean, off the Peloponnesian mainland northwest of Crete. Hitherto, Italo-British naval hostilities in the Mediterranean area since June 1940 had comprised only one noteworthy action: the sinking in November at the Italian naval base of Taranto of three battleships by aircraft from the British carrier *Illustrious*. In March 1941, however, some Italian naval forces, including the battleship *Vittorio Veneto*, with several cruisers and destroyers, set out to threaten British convoys to Greece; and British forces, including the battleships *Warspite, Valiant,* and *Barham* and the aircraft carrier *Formidable,* likewise with cruisers and destroyers, were sent to intercept them. When the forces met in the morning of March 28, off Cape Matapan, the *Vittorio Veneto* opened fire on the lighter British ships but was soon trying to escape from the engagement, for fear of the torpedo aircraft from the *Formidable.* The battle then became a pursuit, which lasted long into the night. Finally, though the severely damaged *Vittorio Veneto* made good her escape, the British sank three Italian cruisers and two destroyers. The Italian Navy made no more surface ventures into the eastern Mediterranean.

The German attack on Greece, scheduled for April 1, 1941, was postponed for a few days when Hitler, because of the Belgrade coup d'état, decided that Yugoslavia was to be destroyed at the same time. While Great Britain's efforts to draw Yugoslavia into the Greco-British defensive system were fruitless, Germany began canvasing allies for its planned invasion of Yugoslavia and Greece. Italy agreed to collaborate in the attack, and Hungary and Bulgaria agreed to send troops to occupy the territories that they coveted as soon as the Germans should have destroyed the Yugoslav state.

On April 6, 1941, the Germans, with 24 divisions and 1,200 tanks, invaded both Yugoslavia (which had 32 divisions) and Greece (which had 15 divisions). The operations were conducted in the same way as Germany's previous blitzkrieg campaigns. While massive air raids struck Belgrade, List's 12th Army drove westward and southward from the Bulgarian frontiers, Kleist's armoured group northwestward from Sofia, and Weichs's 2nd Army southward from Austria and from western Hungary. The 12th Army's advance through Skopje to the Albanian border cut communications between Yugoslavia and Greece in two days; Niš fell to Kleist on April 9, Zagreb to Weichs on April 10; and on April 11 the Italian 2nd Army (comprising 15 divisions) advanced from Istria into Dalmatia. After the fall of Belgrade to the German forces from bases in Romania (April 12), the remnant of the Yugoslav Army—whose only offensive, in northern Albania, had collapsed—was encircled in Bosnia. Its capitulation was signed, in Belgrade, on April 17.

Yugoslavia and Greece invaded by the Germans

In Greece, meanwhile, the Germans took Salonika (Thessaloníki) on April 9, 1941, and then initiated a drive toward Ioánnina (Yannina), thus severing communication between the bulk of the Greek Army (which was on the Albanian frontier) and its rear. The isolated main body capitulated on April 20, the Greek Army as a whole on April 22. Two days later the pass of Thermopylae, defended by a British rear guard, was taken by the Germans, who entered Athens on April 27. All mainland Greece and all the Greek Aegean islands except Crete were under German occupation by May 11, the Ionian islands under Italian. The remainder of Britain's 50,000-man force in Greece was hastily evacuated with great difficulty after leaving all of their tanks and other heavy equipment behind.

The campaign against Yugoslavia brought 340,000 soldiers of the Yugoslav Army into captivity as German prisoners of war. In the campaign against Greece the Germans took 220,000 Greek and 20,000 British or Commonwealth prisoners of war. The combined German losses in the Balkan campaigns were about 2,500 dead, 6,000 wounded, and 3,000 missing.

German airborne troops began to land in Crete on May 20, 1941, at Máleme, in the Canea-Suda area, at Réthimnon, and at Iráklion. Fighting, on land and on the sea, with heavy losses on both sides, went on for a week before the Allied commander in chief, General Bernard Cyril Freyberg of the New Zealand Expeditionary Force, was authorized to evacuate the island. The last defenders were overwhelmed at Réthimnon on May 31. The prisoners of war taken by the Germans in Crete numbered more than 15,000 British or Commonwealth troops, besides the Greeks taken. In battles around the island, German air attacks sank three light cruisers and six destroyers of the British Mediterranean fleet and damaged three battleships,

one aircraft carrier, six light cruisers, and five destroyers.

Dissolution of Yugoslavia

Both the Yugoslav and the Greek royal governments went into exile on their armies' collapse. The Axis powers were left to dispose as they would of their conquests. Yugoslavia was completely dissolved: Croatia, the independence of which had been proclaimed on April 10, 1941, was expanded to form Great Croatia, which included Srem (Syrmia, the zone between the Sava and the Danube south of the Drava confluence) and Bosnia and Hercegovina; most of Dalmatia was annexed to Italy; Montenegro was restored to independence; Yugoslav Macedonia was partitioned between Bulgaria and Albania; Slovenia was partitioned between Italy and Germany; the Baranya triangle and the Bačka went to Hungary; the Banat and Serbia were put under German military administration. Of the independent states, Great Croatia, ruled by Ante Pavelić's nationalist Ustaše ("Insurgents"), and Montenegro were Italian spheres of influence, although German troops still occupied the eastern part of Great Croatia. A puppet government of Serbia was set up by the Germans in August 1941.

While Bulgarian troops occupied eastern Macedonia and most of western Thrace, the rest of mainland Greece, theoretically subject to a puppet government in Athens, was militarily occupied by the Italians except for three zones, namely the Athens district, the Salonika district, and the Dimotika strip of Thrace, which the German conquerors reserved for themselves. The Germans also remained in occupation of Lesbos, Chios, Samos, Melos, and Crete.

OTHER FRONTS, 1940–41

Egypt and Cyrenaica, 1940–summer 1941. The contemporary course of events in the Balkans, described above, nullified the first great victory won by British land forces in World War II, which took place in North Africa. When Italy declared war against Great Britain in June 1940, it had nearly 300,000 men under Marshal Rodolfo Graziani in Cyrenaica (present-day Libya), to confront the 36,000 troops whom the British commander in chief in the Middle East, General Sir Archibald Wavell, had in Egypt to protect the North African approaches to the Suez Canal. Between these forces lay the Western Desert, in which the westernmost position actually held by the British was Mersa Matruh (Marsā Maṭrūḥ), 120 miles east of the Cyrenaican frontier. The Italians in September 1940 occupied Sīdī Barrānī, 170 miles west of Mersa Matruh; but, after settling six divisions into a chain of widely separated camps, they did nothing more for weeks, and during that time Wavell received some reinforcements.

Wavell's offensive

Wavell, whose command included not only Egypt but also the East African fronts against the Italians, decided to strike first in North Africa. On Dec. 7, 1940, some 30,000 men, under Major General Richard Nugent O'Connor, advanced westward, from Mersa Matruh, against 80,000 Italians; but, whereas the Italians at Sīdī Barrānī had only 120 tanks, O'Connor had 275. Having passed by night through a gap in the chain of forts, O'Connor's forces stormed three of the Italian camps, while the 7th Armoured Division was already cutting the Italians' road of retreat along the coast to the west. On December 10 most of the positions closer to Sīdī Barrānī were overrun; and on December 11 the reserve tanks made a further enveloping bound to the coast beyond Buqbuq, intercepting a large column of retreating Italians. In three days the British had taken nearly 40,000 prisoners.

Falling back across the frontier into Cyrenaica, the remnant of the Italian forces from Sīdī Barrānī shut itself up in the fortress of Bardia (Bardīyah), which O'Connor's tanks speedily isolated. On Jan. 3, 1941, the British assault on Bardia began, and three days later the whole garrison of Bardia surrendered—45,000 men. The next fortress to the west, Tobruk (Ṭubruq), was assaulted on January 23 and captured the next day (30,000 more prisoners).

To complete their conquest of Cyrenaica, it remained for the British to take the port of Benghazi. On Feb. 3, 1941, however, O'Connor learned that the Italians were about to abandon Benghazi and to retreat westward down the coast road to Agheila (al-'Uqaylah). Thereupon he boldly ordered the 7th Armoured Division to cross the desert hinterland and intercept the Italian retreat by cutting the coast road well to the east of Agheila. On February 5, after an advance of 170 miles in 33 hours, the British were blocking the Italians' line of retreat south of Beda Fomm (Bayḍā' Fumm); and in the morning of February 6, as the main Italian columns appeared, a day of battle began. Though the Italians had, altogether, nearly four times as many cruiser tanks as the British, by the following morning 60 Italian tanks had been crippled, 40 more abandoned, and the rest of Graziani's army was surrendering in crowds. The British, only 3,000 strong and having lost only three of their 29 tanks, took 20,000 prisoners, 120 tanks, and 216 guns.

The British, having occupied Benghazi on February 6 and Agheila on February 8, could now have pushed on without hindrance to Tripoli, but the chance was foregone: the Greek government had accepted Churchill's reiterated offer of British troops to be sent to Greece from Egypt, which meant a serious reduction of British strength in North Africa.

The reduction was to have serious consequences, because on February 6, the very day of Beda Fomm, a young general, Erwin Rommel, had been appointed by Hitler to command two German mechanized divisions that were to be sent as soon as possible to help the Italians. Arriving in Tripolitania, Rommel decided to try an offensive with what forces he had. Against the depleted British strength, he was rapidly and brilliantly successful. After occupying Agheila with ease on March 24 and Mersa Bréga (Qasr al-Burayqah) on March 31, he resumed his advance on April 2—despite orders to stand still for two months—with 50 tanks backed by two new Italian divisions. The British evacuated Benghazi the next day and began a precipitate retreat into Egypt, losing great numbers of their tanks on the way (a large force of armour, surrounded at Mechili, had to surrender on April 7). By April 11 all Cyrenaica except Tobruk had been reconquered by Rommel's audacious initiative.

Rommel's counter-attack

Tobruk, garrisoned mainly by the 9th Australian Division, held out against siege; and Rommel, though he defeated two British attempts to relieve the place (May and June 1941), was obliged to suspend his offensive on the Egyptian frontier, since he had overstretched his supply lines.

East Africa. Wavell, the success of whose North African strategy had been sacrificed to Churchill's recurrent fantasy of creating a Balkan front against Germany (Greece in 1941 was scarcely less disastrous for the British than the Dardanelles in 1915), nevertheless enjoyed one definitive triumph before Churchill, doubly chagrined at having lost Cyrenaica for Greece's sake and Greece for no advantage at all, removed him, in the summer of 1941, from his command in the Middle East. That triumph was the destruction of Italian East Africa and the elimination, thereby, of any threat to the Suez Canal from the south or to Kenya from the north.

In August 1940 Italian forces mounted a full-scale offensive and overran British Somaliland. Wavell, however, was already assured of the collaboration of the former Ethiopian emperor Haile Selassie in raising the Ethiopians in patriotic revolt against the Italians; and, whereas in June he had disposed only of meagre resources against the 200,000 men and 325 aircraft under the Duca d'Aosta, Amedeo di Savoia, his troops in the Sudan were reinforced by two Indian divisions before the end of the year. After Haile Selassie and a British major, Orde Wingate, with two battalions of Ethiopian exiles, had crossed the Sudanese frontier directly into Ethiopia, General William Platt and the Indian divisions invaded Eritrea on Jan. 19, 1941 (the Italians had already abandoned Kassala); and, almost simultaneously, British troops from Kenya, under General Alan Cunningham, advanced into Italian Somaliland.

Platt's drive eastward into Eritrea was checked on February 5, at Keren, where the best Italian troops, under General Nicolangelo Carnimeo, put up a stiff defense facilitated by a barrier of cliffs. But when Keren fell on March 26, Platt's way to Asmara (Asmera), to Massawa (Mitsiwa), and then from Eritrea southward into Ethiopia was comparatively easy. Meanwhile, Cunningham's troops

Liberation of Ethiopia

were advancing northward into Ethiopia; and on April 6 they entered the Ethiopian capital, Addis Ababa. Finally, the Duca d'Aosta was caught between Platt's column and Cunningham's; and at Amba Alaji, on May 20, he and the main body of his forces surrendered.

Iraq and Syria, 1940–41. In 1940 Prince ʿAbd al-Ilāh, regent of Iraq for King Fayṣal, had a government divided within itself about the war; he himself and his foreign minister, Nuri as-Said, were pro-British, but his prime minister, Rashid Ali al-Gailani, had pro-German leanings. Having resigned office in January 1941, Rashid Ali on April 3 seized power in Baghdad with help from some army officers and announced that the temporarily absent regent was deposed. The British, ostensibly exercising their right under the Anglo-Iraqi Treaty of 1930 to move troops across Iraqi territory, landed troops at Basra on April 19 and rejected Iraqi demands that these troops be sent on into Palestine before any further landings. Iraqi troops were then concentrated around the British air base at Ḥabbānīyah, west of Baghdad; and on May 2 the British commander there opened hostilities, lest the Iraqis should attack first. Having won the upper hand at Ḥabbānīyah and been reinforced from Palestine, the British troops from the air base marched on Baghdad; and on May 30 Rashid Ali and his friends took refuge in Iran. ʿAbd al-Ilāh was reinstated as regent; Nuri became prime minister; and the British military presence remained to uphold them.

German military supplies for Rashid Ali were dispatched too late to be useful to him; but they reached Iraq via Syria, whose high commissioner, General H.-F. Dentz, was a nominee of the Vichy government of France. Lest Syria and Lebanon should fall altogether under Axis control, the British decided to intervene there. Consequently, Free French forces, under General Georges Catroux, with British, Australian, and Indian support, were sent into both countries from Palestine on June 8, 1941; and a week later British forces invaded Syria from Iraq. Dentz's forces put up an unexpectedly stiff resistance, particularly against the Free French, but were finally obliged to capitulate: an armistice was signed at Acre on July 14. By an arrangement of July 25 the Free French retained territorial command in Syria and Lebanon subject to strategic control by the British.

The beginning of lend-lease. On June 10, 1940, when Italy entered the war on the German side and when the fall of France was imminent, U.S. president Franklin D. Roosevelt declared that the United States would "extend to the opponents of force the material resources of this nation." After France fell, he pursued this policy by aiding the British in their struggle against Germany. Roosevelt arranged for the transfer of surplus American war matériel to the British under various arrangements, including the exchange of 50 old American destroyers for certain British-held Atlantic bases, and he facilitated the placing of British orders for munitions in the United States. The British decided to rely on the United States unreservedly and without regard to their ability to pay. By December 1940 they had already placed orders for war materials that were far more than they could possibly muster the dollar exchange to finance.

Churchill suggested the concept of lend-lease to Roosevelt in December 1940, proposing that the United States provide war materials, foodstuffs, and clothing to the democracies (and particularly to Great Britain). Roosevelt assented, and a bill to achieve this purpose was passed by the Congress in early 1941. The Lend-Lease Act not only empowered the president to transfer defense materials, services, and information to any foreign government whose defense he deemed vital to that of the United States, but also left to his discretion what he should ask in return. An enormous grant of power, it gave Roosevelt virtually a free hand to pursue his policy of material aid to the "opponents of force." Congress appropriated funds generously, amounting to almost $13,000,000,000 by November 1941. Other countries besides Britain began receiving lend-lease aid by this time, including China and the Soviet Union. From the time of the German invasion of the U.S.S.R., Roosevelt had been clearly determined to aid the Soviet Union, but the American public's suspicions of Communism delayed his declaring that country eligible for lend-lease until November 1941. American deliveries of aircraft, tanks, and other supplies to the U.S.S.R. began shortly thereafter.

The Atlantic and the Mediterranean, 1940–41. At the outbreak of World War II, the primary concerns of the British Navy were to defend Great Britain from invasion and to retain command of the ocean trading routes, both in order to protect the passage of essential supplies of food and raw materials for Britain and to deny the trading routes to the Axis powers, thus drawing tight once again the blockade that had proved so successful during World War I. Britain had adequate forces of battleships, aircraft carriers, cruisers, and other ships to fulfill these tasks.

The German Navy's role was to protect Germany's coasts, to defend its sea communications and to attack those of the Allies', and to support land and air operations. These modest goals were in keeping with Germany's position as the dominant land-based power in continental Europe. Germany's main naval weapon during the war was to be the submarine, or U-boat, with which it attacked Allied shipping much as it had in World War I.

German control of the Biscay ports after the fall of France in June 1940 provided the U-boats with bases from which they could infest the Atlantic without having to pass either through the Channel or around the north of the British Isles at the end of every sortie. Thenceforward, so long as naval escorts for outgoing convoys from the British Isles could go only 200 or 300 miles out to sea before having to turn back to escort incoming convoys, the U-boats had a very wide field for free-ranging activity: sinkings rose sharply from 55,580 tons in May 1940 to 352,407 tons in October, achieved mainly by solitary attacks by single U-boats at night. But the beginning of lend-lease and the freeing of British warships after the German invasion threat waned enabled the British to escort their convoys for 400 miles by October 1940 and halfway across the Atlantic by April 1941. Since air cover for shipping could also be provided from the British Isles, from Canada, and from Iceland, the Atlantic space left open to the U-boats was reduced by May 1941 to a width of only 300 miles. Moreover, British surface vessels had the ASDIC (Anti-Submarine Detection Investigation Committee) device to detect submerged U-boats. By the spring of 1941, under the guidance of Admiral Karl Dönitz, the U-boat commanders were changing their tactic of individual operation to one of wolf-pack attacks: groups of U-boats, disposed in long lines, would rally when one of them by radio signaled a sighting and overwhelm the convoy by weight of numbers. Between July and December 1941 the German U-boat strength was raised from 65 to more than 230.

Furthermore, the German surface fleet became more active against Allied seaborne trade. Six armed German raiders disguised as merchantmen, with orders to leave convoys alone and to confine their attacks to unescorted ships, roamed the oceans with practical impunity from the spring of 1940 and had sunk 366,644 tons of shipping by the end of the year. German battleships—the *Admiral Scheer,* the *Admiral Hipper,* the *Scharnhorst,* and the *Gneisenau*—one after another began similar raiding operations, with considerable success, from October 1940; and in May 1941 a really modern battleship, the *Bismarck,* and a new cruiser, the *Prinz Eugen,* put out to sea from Germany. The *Bismarck* and the *Prinz Eugen,* however, were located by British reconnaissance in the North Sea near Bergen, and an intensive hunt for them was immediately set in motion. Tracked from a point northwest of Iceland by two British cruisers, the two German ships were engaged on May 24 by the battle cruiser *Hood* and by the new battleship *Prince of Wales;* and, though the *Hood* was sunk, the *Bismarck* 's fuel supply was put out of action, so that her commander, Admiral Günther Lütjens, decided to make for the French coast. Separating from the *Prinz Eugen* (which escaped), the *Bismarck* threw off her pursuers early on May 25 but was sighted again the next day some 660 miles west of Brest. Paralyzed by torpedo aircraft from the *Ark Royal,* she was bombarded and sunk by the *King George V,* the *Rodney,* and the *Dorsetshire* on May 27.

The sinking of the *Bismarck*

In the Mediterranean the year 1941 ended with some naval triumphs for the Axis: U-boats torpedoed the *Ark Royal* on November 13 and the *Barham* 12 days later; Italian frogmen, entering the harbour of Alexandria, on December 19 crippled the battleships *Queen Elizabeth* and *Valiant;* and two British cruisers and a destroyer were also sunk in Mediterranean waters in December.

German strategy, 1939–42. German strategy in World War II is wholly intelligible only if Hitler's far-reaching system of power politics and his racist ideology are borne in mind. Since the 1920s his program had been first to win power in Germany proper, next to consolidate Germany's domination over Central Europe, and then to raise Germany to the status of a world power by two stages: (1) the building up of a continental empire embracing all Europe, including the European portion of the Soviet Union, and (2) the attainment for Germany of equal rank with the British Empire, Japan, and the United States—the only world powers to be left after the elimination of France and the U.S.S.R.—through the acquisition of colonies in Africa and the construction of a strong fleet with bases on the Atlantic. In the succeeding generation Hitler foresaw a decisive conflict between Germany and the United States, during which he hoped that Great Britain would be Germany's ally.

The conquest of the European part of the Soviet Union, which in Hitler's calendar was dated approximately for 1943–45, was to be preceded, he thought, by short localized campaigns elsewhere in Europe to provide a strategic shield and to secure Germany's rear for the great expedition of conquest in the East, which was also bound up with the extermination of the Jews. The most important of the localized campaigns would be that against France. While this European program remained unfulfilled, it was imperative to avoid any world war, since only after the German Reich had come to dominate the whole European continent would it have the economic base and the territorial extent that were prerequisite for success in a great war, especially against maritime world powers.

Hitler had always contemplated the overthrow of the Soviet regime, and though he had congratulated himself on the German-Soviet Nonaggression Pact of 1939 as a matter of expediency, anti-Bolshevism had remained his most profound emotional conviction. His feelings had been stirred up afresh by the Soviet occupation of the Baltic states and of Bessarabia and northern Bukovina in June 1940 and by the consequent proximity of Soviet forces to the Romanian oil fields on which Germany depended. Hitler became acutely suspicious of the intentions of the Soviet leader, Joseph Stalin, and he began to feel that he could not afford to wait to complete the subjugation of western Europe before dealing with the Soviet Union. Hitler and his generals had originally scheduled the invasion of the U.S.S.R. for mid-May 1941, but the unforeseen necessity of invading Yugoslavia and Greece in April of that year had forced them to postpone the Soviet campaign to late June. The swiftness of Hitler's Balkan victories enabled him to keep to this revised timetable, but the five weeks' delay shortened the time for carrying out the invasion of the U.S.S.R. and was to prove the more serious because in 1941 the Russian winter would arrive earlier than usual. Nevertheless, Hitler and the heads of the Oberkommando des Heeres (OKH, or German Army High Command), namely the army commander in chief Werner von Brauchitsch and the army general staff chief Franz Halder, were convinced that the Red Army could be defeated in two or three months, and that, by the end of October, the Germans would have conquered the whole European part of Russia and the Ukraine west of a line stretching from Archangel to Astrakhan. The invasion of the Soviet Union was given the code name "Operation Barbarossa."

INVASION OF THE SOVIET UNION, 1941

The German attack on the Soviet Union, 1941. For the campaign against the Soviet Union, the Germans allotted almost 150 divisions containing a total of about 3,000,000 men. Among these were 19 panzer divisions, and in total the "Barbarossa" force had about 3,000 tanks, 7,000 artillery pieces, and 2,500 aircraft. It was in effect the largest and most powerful invasion force in human history. The Germans' strength was further increased by more than 30 divisions of Finnish and Romanian troops.

The Soviet Union had twice or perhaps three times the number of both tanks and aircraft as the Germans had, but their aircraft were mostly obsolete. The Soviet tanks were about equal to those of the Germans, however. A greater hindrance to Hitler's chances of victory was that the German intelligence service underestimated the troop reserves that Stalin could bring up from the depths of the U.S.S.R. The Germans correctly estimated that there were about 150 divisions in the western parts of the U.S.S.R. and reckoned that 50 more might be produced. But the Soviets actually brought up more than 200 fresh divisions by the middle of August, making a total of 360. The consequence was that, though the Germans succeeded in shattering the original Soviet armies by superior technique, they then found their path blocked by fresh ones. The effects of the miscalculations were increased because much of August was wasted while Hitler and his advisers were having long arguments as to what course they should follow after their initial victories. Another factor in the Germans' calculations was purely political, though no less mistaken; they believed that within three to six months of their invasion, the Soviet regime would collapse from lack of domestic support.

The German attack on the Soviet Union was to have an immediate and highly salutary effect on Great Britain's situation. Until then Britain's prospects had appeared hopeless in the eyes of most people except the British themselves; and the government's decision to continue the struggle after the fall of France and to reject Hitler's peace offers could spell only slow suicide unless relief came from either the United States or the U.S.S.R. Hitler brought Great Britain relief by turning eastward and invading the Soviet Union just as the strain on Britain was becoming severe.

On June 22, 1941, the German offensive was launched by three army groups under the same commanders as in the invasion of France in 1940: on the left (north), an army group under Leeb struck from East Prussia into the Baltic states toward Leningrad; on the right (south), another army group, under Rundstedt, with an armoured group under Kleist, advanced from southern Poland into the Ukraine against Kiev, whence it was to wheel southeastward to the coasts of the Black Sea and the Sea of Azov; and in the centre, north of the Pripet Marshes, the main blow was delivered by Bock's army group, with one armoured group under Guderian and another under Hoth, thrusting northeastward at Smolensk and Moscow.

The invasion along a 1,800-mile front took the Soviet leadership completely by surprise and caught the Red Army in an unprepared and partially demobilized state. Piercing the northern border, Guderian's tanks raced 50 miles beyond the frontier on the first day of the invasion and were at Minsk, 200 miles beyond it, on June 27. At Minsk they converged with Hoth's tanks, which had pierced the opposite flank, but Bock's infantry could not follow up quickly enough to complete the encirclement of the Soviet troops in the area; though 300,000 prisoners were taken in the salient, a large part of the Soviet forces was able to escape to the east. The Soviet armies were clumsily handled and frittered their tank strength away in piecemeal action like that of the French in 1940. But the isolated Soviet troops fought with a stubbornness that the French had not shown, and their resistance imposed a brake by continuing to block road centres long after the German tide had swept past them. The result was similar when Guderian's tanks, having crossed the Dnepr River on July 10, entered Smolensk six days later and converged with Hoth's thrust through Vitebsk: 200,000 Soviet prisoners were taken; but some Soviet forces were withdrawn from the trap to the line of the Desna, and a large pocket of resistance lay behind the German armour. By mid-July, moreover, a series of rainstorms were turning the sandy Russian roads into clogging mud, over which the wheeled vehicles of the German transport behind the tanks could make only very slow progress. The Germans also began

Smolensk taken by the Germans

to be hampered by the scorched earth policy adopted by the retreating Soviets. The Soviet troops burned crops, destroyed bridges, and evacuated factories in the face of the German advance. Entire steel and munitions plants in the westernmost portions of the U.S.S.R. were dismantled and shipped by rail to the east, where they were put back into production. The Soviets also destroyed or evacuated most of their rolling stock (railroad cars), thus depriving the Germans of the use of the Soviet rail system, since Soviet railroad track was of a different gauge than German track and German rolling stock was consequently useless on it.

Nevertheless, by mid-July the Germans had advanced more than 400 miles and were only 200 miles from Moscow. They still had ample time to make decisive gains before the onset of winter, but they lost the opportunity, primarily because of arguments throughout August between Hitler and the OKH about the destination of the next thrusts thence: whereas the OKH proposed Moscow as the main objective, Hitler wanted the major effort to be directed southeastward, through the Ukraine and the Donets Basin into the Caucasus, with a minor swing northwestward against Leningrad (to converge with Leeb's army group).

In the Ukraine, meanwhile, Rundstedt and Kleist had made short work of the foremost Soviet defenses, stronger though the latter had been. A new Soviet front south of Kiev was broken by the end of July; and in the next fortnight the Germans swept down to the Black Sea mouths of the Bug and Dnepr rivers—to converge with Romania's simultaneous offensive. Kleist was then ordered to wheel northward from the Ukraine, Guderian southward from Smolensk, for a pincer movement around the Soviet forces behind Kiev; and by the end of September the claws of the encircling movement had caught 520,000 men. These gigantic encirclements were partly the fault of inept Soviet high commanders and partly the fault of Stalin, who as commander in chief stubbornly overrode the advice of his generals and ordered his armies to stand and fight instead of allowing them to retreat eastward and regroup in preparation for a counteroffensive.

Winter was approaching, and Hitler stopped Leeb's northward drive on the outskirts of Leningrad. He ordered Rundstedt and Kleist, however, to press on from the Dnepr toward the Don and the Caucasus; and Bock was to resume the advance on Moscow.

Bock's renewed advance on Moscow began on Oct. 2, 1941. Its prospects looked bright when Bock's armies brought off a great encirclement around Vyazma, where 600,000 more Soviet troops were captured. That left the Germans momentarily with an almost clear path to Moscow. But the Vyazma battle had not been completed until late October; the German troops were tired, the country became a morass as the weather got worse, and fresh Soviet forces appeared in the path as they plodded slowly forward. Some of the German generals wanted to break off the offensive and to take up a suitable winter line. But Bock wanted to press on, believing that the Soviets were on the verge of collapse, while Brauchitsch and Halder tended to agree with his view. As that also accorded with Hitler's desire, he made no objection. The temptation of Moscow, now so close in front of their eyes, was too great for any of the topmost leaders to resist. On December 2 a further effort was launched, and some German detachments penetrated into the suburbs of Moscow; but the advance as a whole was held up in the forests covering the capital. The stemming of this last phase of the great German offensive was partly due to the effects of the Russian winter, whose subzero temperatures were the most severe in several decades. In October and November a wave of frostbite cases had decimated the ill-clad German troops, for whom provisions of winter clothing had not been made, while the icy cold paralyzed the Germans' mechanized transport, tanks, artillery, and aircraft. The Soviets, by contrast, were well clad and tended to fight more effectively in winter than did the Germans. By this time German casualties had mounted to levels that were unheard of in the campaigns against France and the Balkans; by November the Germans had suffered about 730,000 casualties.

The Soviet counter-offensives

In the south, Kleist had already reached Rostov-on-Don, gateway to the Caucasus, on November 22, but had exhausted his tanks' fuel in doing so. Rundstedt, seeing the place to be untenable, wanted to evacuate it but was overruled by Hitler. A Soviet counteroffensive recaptured Rostov on November 28, and Rundstedt was relieved of his command four days later. The Germans, however, managed to establish a front on the Mius River—as Rundstedt had recommended.

As the German drive against Moscow slackened, the Soviet commander on the Moscow front, General Georgy Konstantinovich Zhukov, on December 6 inaugurated the first great counteroffensive with strokes against Bock's right in the Elets (Yelets) and Tula sectors south of Moscow and against his centre in the Klin and Kalinin sectors to the northwest. Levies of Siberian troops, who were extremely effective fighters in cold weather, were used for these offensives. There followed a blow at the German left, in the Velikie Luki sector; and the counteroffensive, which was sustained throughout the winter of 1941–42, soon took the form of a triple convergence toward Smolensk.

These Soviet counteroffensives tumbled back the exhausted Germans, lapped around their flanks, and produced a critical situation. From generals downward, the invaders were filled with ghastly thoughts of Napoleon's retreat from Moscow. In that emergency Hitler forbade any retreat beyond the shortest possible local withdrawals. His decision exposed his troops to awful sufferings in their advanced positions facing Moscow, for they had neither the clothing nor the equipment for a Russian winter campaign; but if they had once started a general retreat it might easily have degenerated into a panic-stricken rout.

The Red Army's winter counteroffensive continued for more than three months after its December launching, though with diminishing progress. By March 1942 it had advanced more than 150 miles in some sectors. But the Germans maintained their hold on the main bastions of their winter front—such towns as Schlüsselburg, Novgorod, Rzhev, Vyazma, Bryansk, Orël (Oryol), Kursk, Kharkov, and Taganrog—despite the fact that the Soviets had often advanced many miles beyond these bastions, which were in effect cut off. In retrospect, it became clear that Hitler's veto on any extensive withdrawal worked out in such a way as to restore the confidence of the German troops and probably saved them from a widespread collapse. Nevertheless, they paid a heavy price indirectly for that rigid defense. One immediate handicap was that the strength of the Luftwaffe was drained in the prolonged effort to maintain supplies by air, under winter conditions, to the garrisons of these more or less isolated bastion towns. The tremendous strain of that winter campaign, on armies which had not been prepared for it, had other serious effects. Before the winter ended, many German divisions were reduced to barely a third of their original strength, and they were never fully built up again.

The German plan of campaign had begun to miscarry in August 1941, and its failure was patent when the Soviet counteroffensive started. Nevertheless, having dismissed Brauchitsch and appointed himself army commander in chief in December, Hitler persisted in overruling the tentative opposition of the general staff to his strategy.

The first three months of the German–Soviet conflict produced cautious rapprochements between the U.S.S.R. and Great Britain and between the U.S.S.R. and the United States. The Anglo-Soviet agreement of July 12, 1941, pledged the signatory powers to assist one another and to abstain from making any separate peace with Germany. On Aug. 25, 1941, British and Soviet forces jointly invaded Iran, to forestall the establishment of a German base there and to divide the country into spheres of occupation for the duration of the war; and late in September—at a conference in Moscow—Soviet, British, and U.S. representatives formulated the monthly quantities of supplies, including aircraft, tanks, and raw materials, that Great Britain and the United States should try to furnish to the Soviet Union.

The critical situation on the Eastern Front did not deter Hitler from declaring Germany to be at war with the United States on Dec. 11, 1941, after the Japanese attack

on the U.S., British, and Dutch positions in the Pacific and in the Far East (see below), since this extension of hostilities did not immediately commit the German land forces to any new theatre but at the same time had the merit of entitling the German Navy to intensify the war at sea.

THE WAR IN THE PACIFIC, 1938–41

The war in China, 1937–41. In 1931–32 the Japanese had invaded Manchuria (Northeast China) and, after overcoming ineffective Chinese resistance there, had created the Japanese-controlled puppet state of Manchukuo. In the following years the Nationalist government of China, headed by Chiang Kai-shek, temporized in the face of Japanese military and diplomatic pressures and instead waged an internal war against the Chinese Communists, led by Mao Zedong, who were based in Shensi Province in north-central China. Meanwhile, the Japanese began a military buildup in North China proper, which in turn stimulated the formation of a unified resistance by the Nationalists and the Communists. Overt hostilities between Japan and China began after the Marco Polo Bridge incident of July 7, 1937, when shots were exchanged between Chinese and Japanese troops on the outskirts of Peking. Open fighting broke out in that area, and in late July the Japanese captured the Peking-Tientsin area. Thereupon full-scale hostilities began between the two nations. The Japanese landed near Shanghai, at the mouth of the Yangtze River, and took Shanghai in November and the Chinese capital, Nanking, in December 1937. Chiang Kai-shek moved his government to Han-k'ou (one of the Wu-han cities), which lay 435 miles west of Shanghai along the Yangtze. The Japanese also pushed southward and westward from the Peking area into Hopeh and Shansi provinces. In 1938 the Japanese launched several ambitious military campaigns that brought them deep into the heart of central China. They advanced to the northeast and west from Nanking, taking Suchow and occupying the Wu-han cities. The Nationalists were forced to move their government to Chungking in Szechwan Province, about 500 miles west of the Wu-han cities. The Japanese also occupied Canton and several other coastal cities in South China in 1938.

Nationalist Chinese resistance to these Japanese advances was ineffective, primarily because the Nationalist leadership was still more interested in holding their forces in reserve for a future struggle with the Communists than in repelling the Japanese. By contrast, the Communists, from their base in north-central China, began an increasingly effective guerrilla war against the Japanese troops in Manchuria and North China. The Japanese needed large numbers of troops to maintain their hold on the immense Chinese territories and populations they controlled. Of the 51 infantry divisions making up the Japanese Army in 1941, 38 of them, comprising about 750,000 men, were stationed in China (including Manchuria).

Japanese policy, 1939–41. When war broke out in Europe in September 1939, the Japanese, despite a series of victorious battles, had still not brought their war in China to an end: on the one hand, the Japanese strategists had made no plans to cope with the guerrilla warfare pursued by the Chinese; on the other, the Japanese commanders in the field often disregarded the orders of the supreme command at the Imperial headquarters and occupied more Chinese territory than they had been ordered to take. Half of the Japanese Army was thus still tied down in China when the commitment of Great Britain and France to war against Germany opened up the prospect of wider conquests for Japan in Southeast Asia and in the Pacific. Japan's military ventures in China proper were consequently restricted rather more severely henceforth.

The German victories over The Netherlands and France in the summer of 1940 further encouraged the Japanese premier, Prince Konoe, to look southward at those defeated powers' colonies and also, of course, at the British and U.S. positions in the Far East. The island archipelago of the Dutch East Indies (now Indonesia) along with French Indochina and British-held Malaya contained raw materials (tin, rubber, petroleum) that were essential to Japan's industrial economy, and if Japan could seize these regions and incorporate them into the empire, it could make itself virtually self-sufficient economically and thus become the dominant power in the Pacific Ocean. Since Great Britain, single-handedly, was confronting the might of the Axis in Europe, the Japanese strategists had to reckon, primarily, with the opposition of the United States to their plans for territorial aggrandizement. When Japanese troops entered northern Indochina in September 1940 (in pursuance of an agreement extorted in August from the Vichy government of France), the United States uttered a protest. Germany and Italy, by contrast, recognized Japan as the leading power in the Far East by concluding with it the Tripartite, or Axis, Pact of Sept. 27, 1940: negotiated by Japanese foreign minister Matsuoka Yosuke, the pact pledged its signatories to come to one another's help in the event of an attack "by a power not already engaged in war." Japan also concluded a neutrality pact with the U.S.S.R. on April 13, 1941.

Berlin–Rome–Tokyo Axis

On July 2, 1941, the Imperial Conference decided to press the Japanese advance southward even at the risk of war with Great Britain and the United States; and this policy was pursued even when Matsuoka was relieved of office a fortnight later. On July 26, in pursuance of a new agreement with Vichy France, Japanese forces began to occupy bases in southern Indochina.

This time the United States reacted vigorously, not only freezing Japanese assets under U.S. control but also imposing an embargo on supplies of oil to Japan. Dismay at the embargo drove the Japanese naval command, which had hitherto been more moderate than the army, into collusion with the army's extremism. When negotiations with the Dutch of Indonesia for an alternative supply of oil produced no satisfaction, the Imperial Conference on September 6, at the high command's insistence, decided that war must be undertaken against the United States and Great Britain unless an understanding with the United States could be reached in a few weeks' time.

General Tōjō Hideki, who succeeded Konoe as premier in mid-October 1941, continued the already desperate talks. The United States, however, persisted in making demands that Japan could not concede: renunciation of the Tripartite Pact (which would have left Japan diplomatically isolated); the withdrawal of Japanese troops from China and from Southeast Asia (a humiliating retreat from an overt commitment of four years' standing); and an open-door regime for trade in China. When Cordell Hull, the U.S. secretary of state, on Nov. 26, 1941, sent an abrupt note to the Japanese bluntly requiring them to evacuate China and Indochina and to recognize no Chinese regime other than that of Chiang Kai-shek, the Japanese could see no point in continuing the talks.

Since peace with the United States seemed impossible, Japan set in motion its plans for war, which would now necessarily be waged not only against the United States but also against Great Britain (the existing war effort of which depended on U.S. support and the Far Eastern colonies of which lay within the orbit of the projected Japanese expansion) and against the Dutch East Indies (the oil of which was essential to Japanese enterprises, even apart from geopolitical considerations).

The evolving Japanese military strategy was based on the peculiar geography of the Pacific Ocean and on the relative weakness and unpreparedness of the Allied military presence in that ocean. The western half of the Pacific is dotted with many islands, large and small, while the eastern half of the ocean is, with the exception of the Hawaiian Islands, almost devoid of landmasses (and hence of usable bases). The British, French, American, and Dutch military forces in the entire Pacific region west of Hawaii amounted to only about 350,000 troops, most of them lacking combat experience and being of disparate nationalities. Allied air power in the Pacific was weak and consisted mostly of obsolete planes. If the Japanese, with their large, well-equipped armies that had been battle-hardened in China, could quickly launch coordinated attacks from their existing bases on certain Japanese-mandated Pacific islands, on Formosa (Taiwan), and from Japan itself, they could overwhelm the Allied forces, overrun the entire western Pacific Ocean as well as Southeast Asia, and then develop

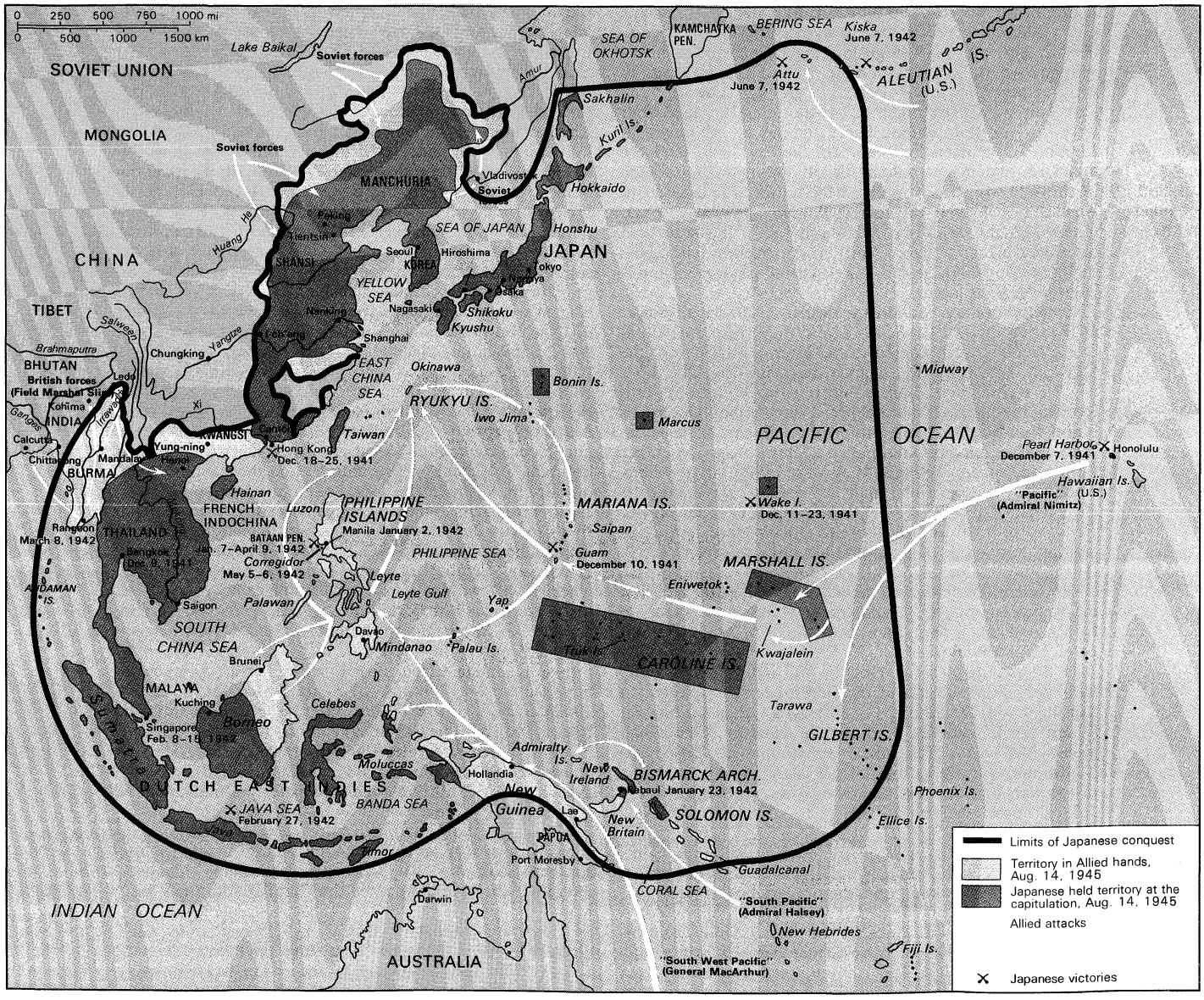
those areas' resources to their own military-industrial advantage. If successful in their campaigns, the Japanese planned to establish a strongly fortified defensive perimeter extending from Burma in the west to the southern rim of the Dutch East Indies and northern New Guinea in the south and sweeping around to the Gilbert and Marshall islands in the southeast and east. The Japanese believed that any American and British counteroffensives against this perimeter could be repelled, after which those nations would eventually seek a negotiated peace that would allow Japan to keep her newly won empire.

Yamamoto's plan

Until the end of 1940 the Japanese strategists had assumed that any new war to be waged would be against a single enemy. When it became clear, in 1941, that the British and the Dutch as well as the Americans must be attacked, a new and daring war plan was successfully sponsored by the commander in chief of the Combined Fleet, Admiral Yamamoto Isoroku.

Yamamoto's plan prescribed two operations, together involving the whole strength of his navy, which was composed of the following ships: 10 battleships, six regular aircraft carriers, four auxiliary carriers, 18 heavy cruisers, 20 light cruisers, 112 destroyers, 65 submarines, and 2,274 combat planes. The first operation, to which all six regular aircraft carriers, two battleships, three cruisers, and 11 destroyers were allocated, was to be a surprise attack, scheduled for December 7 (December 8 by Japanese time), on the main U.S. Pacific Fleet in its base at Pearl Harbor in the Hawaiian Islands. The rest of the Japanese Navy was to support the army in the "Southern Operation": 11 infantry divisions and seven tank regiments, assisted by 795 combat planes, were to undertake two drives, one from Formosa through the Philippines, the other from French Indochina and Hainan Island through Malaya, so as to converge on the Dutch East Indies, with a view to the capture of Java as the culmination of a campaign of 150 days—during which, moreover, Wake Island, Guam, the Gilbert Islands, and Burma should also have been secured as outer bastions, besides Hong Kong.

Pearl Harbor and the Japanese expansion, to July 1942. In accordance with Yamamoto's plan, the aircraft carrier strike force commanded by Admiral Nagumo Chuichi sailed westward undetected by any U.S. reconnaissance until it had reached a point 275 miles north of Hawaii. From there, on Sunday, Dec. 7, 1941, a total of about 360 aircraft, composed of dive-bombers, torpedo bombers, and a few fighters, was launched in two waves in the early morning at the giant U.S. naval base at Pearl Harbor. The base at that time was accommodating 70 U.S. fighting ships, 24 auxiliaries, and some 300 planes. The Americans were taken completely by surprise, and all eight battleships in the harbour were hit (though six were eventually repaired and returned to service); three cruisers, three destroyers, a minelayer, and other vessels



The Pacific Theatre of Operations, 1941–45.

were damaged; more than 180 aircraft were destroyed and others damaged (most while parked at airfields); and more than 2,330 troops were killed and over 1,140 wounded. Japanese losses were comparatively small. The Japanese attack failed in one crucial respect, however; the Pacific Fleet's three aircraft carriers were at sea at the time of the attack and escaped harm, and these were to become the nucleus of the United States' incipient naval defense in the Pacific. Pearl Harbor's shore installations and oil-storage facilities also escaped damage. The Pearl Harbor attack, unannounced beforehand by the Japanese as it was, unified the American public and swept away any remaining support for American neutrality in the war. On December 8 the U.S. Congress declared war on Japan with only one dissenting vote.

On the day of the attack, December 8 by local time, Formosa-based Japanese bombers struck Clark and Iba airfields in the Philippines, destroying more than 50 percent of the U.S. Army's Far East aircraft; and, two days later, further raids destroyed not only more U.S. fighters but also Cavite Naval Yard, likewise in the Philippines. Part of the U.S. Asiatic Fleet, however, had already gone south in November; and the surviving major ships and bomber aircraft, which were vulnerable for lack of fighter protection, were withdrawn in the next fortnight to safety in bases in Java and Australia.

Japanese forces began to land on the island of Luzon in the Philippines on December 10. The main assault, consisting of the bulk of one division, was made at Lingayen Gulf, 100 miles north-northwest of Manila, on December 22, and a second large landing took place south of Manila two days later. Manila itself fell unopposed to the Japanese on Jan. 2, 1942, but by that time the U.S. and Filipino forces under General Douglas MacArthur were ready to hold Bataan Peninsula (across the bay from Manila) and Corregidor Island (in the bay). The Japanese attack on Bataan was halted initially, but it was reinforced in the following eight weeks. MacArthur was ordered to Australia on March 11, leaving Bataan's defense to Lieutenant General Jonathan M. Wainwright. The latter and his men surrendered on April 9; Corregidor fell in the night of May 5–6; and the southern Philippines capitulated three days later.

Bataan and Corregidor defended and lost

Japanese bombers had already destroyed British air power at Hong Kong on Dec. 8, 1941, and the British and Canadian defenders surrendered to the ground attack from the Kowloon Peninsula (the nearest mainland) on December 25. To secure their flank while pushing southward into Malaya, the Japanese also occupied Bangkok on December 9 and Victoria Point in southernmost Burma on December 16. The Japanese landings in Malaya, from December 8 onward, accompanied as they were by air strikes, overwhelmed the small Australian and Indian forces; and the British battleship *Prince of Wales* and the battle cruiser *Repulse,* sailing from Singapore to cut Japanese communications, were sunk by Japanese aircraft on December 10. By the end of January 1942, two Japanese divisions, with air and armoured support, had occupied all Malaya except Singapore Island. In Burma, meanwhile, other Japanese troops had taken Moulmein and were approaching Rangoon and Mandalay.

On the eastern perimeter of the war zone, the Japanese had bombed Wake Island on December 8, attempted to capture it on December 11, and achieved a landing on December 23, quickly subduing the garrison. Guam had already fallen on December 10. Having also occupied Makin and Tarawa in the Gilbert Islands in the first days of the war, the Japanese successfully attacked Rabaul, the strategic base on New Britain (now part of Papua New Guinea), on Jan. 23, 1942.

A unified American–British–Dutch–Australian Command, ABDACOM, under Wavell, responsible for holding Malaya, Sumatra, Java, and the approaches to Australia, became operative on Jan. 15, 1942; but the Japanese had already begun their advance on the oil-rich Dutch East Indies. They occupied Kuching (December 17), Brunei Bay (January 6), and Jesselton (January 11), on the northern coast of Borneo, as well as Tarakan Island (off northeastern Borneo) and points on Celebes. Balikpapan (on Borneo's east coast) and Kendari (in southeastern Celebes) fell to the Japanese on Jan. 24, 1942, Amboina on February 4, Makasar City (in southwestern Celebes) on February 8, and Bandjarmasin (in southern Borneo) on February 16. Bali was invaded on February 18, and by February 24 the Japanese were also in possession of Timor.

Meanwhile, on February 8 and 9, three Japanese divisions had landed on Singapore Island; and on February 15 they forced the 90,000-strong British, Australian, and Indian garrison there, under Lieutenant General A.E. Percival, to surrender. Singapore was the major British base in the Pacific and had been regarded as unassailable due to its strong seaward defenses. The Japanese took it with comparative ease by advancing down the Malay Peninsula and then assaulting the base's landward side, which the British had left inadequately defended. On February 13, moreover, Japanese paratroopers had landed at Palembang in Sumatra, which fell to an amphibious assault three days later.

The fall of Singapore

When ABDACOM was dissolved on Feb. 25, 1942, only Java remained to complete the Japanese program of conquest. The Allies' desperate attempt to intercept the Japanese invasion fleet was defeated in the seven-hour Battle of the Java Sea on February 27, in which five Allied warships were lost and only one Japanese destroyer damaged. The Japanese landed at three points on Java on February 28 and rapidly expanded their beachheads. On March 9 the 20,000 Allied troops in Java surrendered. In the Indian Ocean, the Japanese captured the Andaman Islands on March 23, and began a series of attacks on British shipping. After the failure of ABDACOM, the U.S.–British Combined Chiefs of Staff placed the Pacific under the U.S. Joint Chiefs' strategic direction. MacArthur became supreme commander of the Southwest Pacific Area, which comprised the Dutch East Indies (less Sumatra), the Philippines, Australia, the Bismarck Archipelago, and the Solomons; and Admiral Chester W. Nimitz became commander in chief of the Pacific Ocean Areas, which comprised virtually every area not under MacArthur. Their missions were to hold the U.S.–Australia line of communications, to contain the Japanese within the Pacific, to support the defense of North America, and to prepare for major amphibious counteroffensives.

Japan's initial war plans were realized with the capture of Java. But despite their military triumphs, the Japanese saw no indication that the Allies were ready for a negotiated peace. On the contrary, it seemed evident that an Allied counterstroke was in the making. The U.S. Pacific Fleet bombed the Marshall Islands on Feb. 1, 1942, Wake Island on February 23, and Marcus Island (between Wake and Japan) on March 1. These moves, together with the bombing of Rabaul on February 23 and the establishment of bases in Australia and a line of communications across the South Pacific, made the Japanese decide to expand so as to cut the Allied line of communications to Australia. They planned to occupy New Caledonia, the Fiji Islands, and Samoa and also to seize eastern New Guinea, whence they would threaten Australia from an air base to be established at Port Moresby. They planned also to capture Midway Island in the North Pacific and to establish air bases in the Aleutians. In pursuance of this new program, Japanese troops occupied Lae and Salamaua in New Guinea and Buka in the Solomon Islands in March 1942 and Bougainville in the Solomons and the Admiralty Islands (north of New Guinea) early in April.

Something to raise the Allies' morale was achieved on April 18, 1942, when 16 U.S. bombers raided Tokyo—though they did little real damage except to the Japanese government's prestige. Far more important were the consequences of the U.S. intelligence services' detection of Japanese plans to seize Port Moresby and Tulagi (in the southern Solomons). Had these two places fallen, Japanese aircraft could have dominated the Coral Sea. In the event, after U.S. aircraft on May 3, 1942, had interfered with the Japanese landing on Tulagi, U.S. naval units, with aircraft, challenged the Japanese ships on their circuitous detour from Rabaul to Port Moresby. On May 5 and 6 the opposing carrier groups sought each other out, and the four-day Battle of the Coral Sea ensued. On May 7 planes

The Battle of the Coral Sea

from the Japanese carriers sank a U.S. destroyer and an oil tanker, but U.S. planes sank the Japanese light carrier *Shoho* and a cruiser; and the next day, though Japanese aircraft sank the U.S. carrier *Lexington* and damaged the carrier *Yorktown,* the large Japanese carrier *Shokaku* had to retire crippled. Finally, the Japanese lost so many planes in the battle that their enterprise against Port Moresby had to be abandoned.

Despite the mixed results of the Battle of the Coral Sea, the Japanese continued with their plan to seize Midway Island. Seeking a naval showdown with the remaining ships of the U.S. Pacific Fleet and counting on their own numerical superiority to secure a victory, the Japanese mustered four heavy and three light aircraft carriers, two seaplane carriers, 11 battleships, 15 cruisers, 44 destroyers, 15 submarines, and miscellaneous small vessels. The U.S. Pacific Fleet had only three heavy carriers, eight cruisers, 18 destroyers, and 19 submarines, though there were some 115 aircraft in support of it. The Americans, however, had the incomparable advantage of knowing the intentions of the Japanese in advance, thanks to the U.S. intelligence services' having broken the Japanese Navy's code and deciphered key radio transmissions. In the ensuing Battle of Midway, the Japanese ships destined to take Midway Island were attacked while still 500 miles from their target by U.S. bombers on June 3. The Japanese carriers were still able to launch their aircraft against Midway early on June 4, but in the ensuing battle, waves of carrier- and Midway-based U.S. bombers sank all four of the Japanese heavy carriers and one heavy cruiser. Appalled by this disaster, the Japanese began to retreat in the night of June 4–5. Though the U.S. carrier *Yorktown* was sunk by torpedo on June 6, Midway was saved from invasion. In the Aleutians, the Japanese bombed Dutch Harbor effectively and on June 7 occupied Attu and Kiska.

The Battle of Midway

The Battle of Midway was probably the turning point of the war in the Pacific, for Japan lost its first-line carrier strength and most of its navy's best trained pilots. Henceforth, the naval strengths of the Japanese and of the Allies were virtually equal. Having lost the strategic initiative, Japan canceled its plans to invade New Caledonia, Fiji, and Samoa.

The Chinese front and Burma, 1941–42. Japan's entry into war against the western Allies had its repercussions in China. Chiang Kai-shek's government on Dec. 9, 1941, formally declared war not only against Japan (a formality long overdue) but also, with political rather than military intent, against Germany and Italy. Three Chinese armies were rushed to the Burmese frontier, since the Burma Road was the only land route whereby the western Allies could send supplies to the Nationalist Chinese government. On Jan. 3, 1942, Chiang was recognized as supreme Allied commander for the China theatre of war; and a U.S. general, Joseph W. Stilwell, was sent to him to be his chief of staff. In the first eight weeks after Pearl Harbor, however, the major achievement of the Chinese was the definitive repulse, on Jan. 15, 1942, of a long-sustained Japanese drive against Ch'ang-sha, on the Canton–Hank'ou railway.

Stilwell as Chiang's chief of staff

Thereafter, Chiang and Stilwell were largely preoccupied by efforts to check the Japanese advance into Burma. By mid-March 1942 two Chinese armies, under Stilwell's command, had crossed the Burmese frontier; but before the end of the month the Chinese force defending Toungoo, in central Burma between Rangoon and Mandalay, was nearly annihilated by the more soldierly Japanese. British and Indian units in Burma fared scarcely better, being driven into retreat by the enemy's numerical superiority both in the air and on the ground. On April 29 the Japanese took Lashio, the Burma Road's southern terminus, thus cutting the supply line to China and turning the Allies' northern flank. Under continued pressure, the British and Indian forces in the following month fell back through Kalewa to Imphāl (across the Indian border), while most of the Chinese retreated across the Salween River into China. By the end of 1942 all of Burma was in Japanese hands, China was effectively isolated (except by air), and India was exposed to the danger of a Japanese invasion through Burma.

Since the U.S. bombers that raided Tokyo on April 18 flew on to Chinese airfields, particularly to those in Chekiang (the coastal province south of Shanghai), the Japanese reacted by launching a powerful offensive to seize those airfields. By the end of July they had generally achieved their objectives.

DEVELOPMENTS FROM AUTUMN 1941 TO SPRING 1942

Allied strategy and controversies, 1940–42. In the year following the collapse of France in June 1940, British strategists, relying as they could on supplies from the nonbelligerent United States, were concerned first with home defense, second with the security of the British positions in the Middle East, and third with the development of a war of attrition against the Axis powers, pending the buildup of adequate forces for an invasion of the European continent. For the United States, President Roosevelt's advisers, from November 1940, based their strategic plans on the "Europe first" principle: that is to say, if the United States became engaged in war simultaneously against Germany, Italy, and Japan, merely defensive operations should be conducted in the Pacific (to protect at least the Alaska–Hawaii–Panama triangle) while an offensive was being mounted in Europe.

Japan's entry into the war terminated the nonbelligerency of the United States. The three weeks' conference, named Arcadia, that Roosevelt, Churchill, and their advisers opened in Washington, D.C., on Dec. 22, 1941, reassured the British about U.S. maintenance of the "Europe first" principle and also produced two plans: a tentative one, code-named "Sledgehammer," for the buildup of an offensive force in Great Britain, in case it should be decided to invade France; and another, code-named "Super-Gymnast," for combining a British landing behind the German forces in Libya (already planned under the code name "Gymnast") with a U.S. landing near Casablanca on the Atlantic coast of Morocco. The same conference furthermore created the machinery of the Combined Chiefs of Staff, where the British Chiefs of Staff Committee was to be linked continuously, through delegates in Washington, D.C., with the newly established U.S. Joint Chiefs of Staff Organization, so that all aspects of the war could be studied in concert. It was on Jan. 1, 1942, during the Arcadia Conference, that the Declaration of the United Nations was signed in Washington, D.C., as a collective statement of the Allies' war aims in sequel to the Atlantic Charter.

The Arcadia Conference

Meanwhile, Churchill became anxious to do something to help the embattled Soviets—who were clamouring for the United States and Britain to invade continental Europe so as to take some of the German pressure off the Eastern Front. Roosevelt was no less conscious than Churchill of the fact that the Soviet Union was bearing by far the greatest burden of the war against Germany; and this consideration inclined him to listen to the arguments of his Joint Chiefs of Staff Organization for a change of plan. After some hesitation, he sent his confidant Harry Hopkins and his army chief of staff General George C. Marshall to London in April 1942 to suggest the scrapping of "Super-Gymnast" in favour of "Bolero," namely the concentration of forces in Great Britain for a landing in Europe (perhaps at Brest or at Cherbourg) in the autumn; then "Roundup," an invasion of France by 30 U.S. and 18 British divisions, could follow in April 1943. The British agreed but soon began to doubt the practicability of mounting an amphibious invasion of France at such an early date.

Attempts to conclude an Anglo-Soviet political agreement were renewed without result, but a 20-year Anglo-Soviet alliance was signed on May 26, 1942; and, though Churchill warned Molotov not to expect an early second front in Europe, Molotov seemed gratified by what he was told about Anglo-U.S. plans.

The Anglo-Soviet Alliance and the Hyde Park Conference

Visiting Roosevelt again in the latter part of June 1942, Churchill at Hyde Park, N.Y., and in Washington, D.C., pressed for a revised and enlarged joint operation in North Africa before the end of the year, instead of a buildup for the invasion of France; but the U.S. Joint Chiefs resolutely upheld the latter plan. After further debate and disagreement, in July the U.S. Joint Chiefs yielded at

last to British obstinacy in favour of a North African enterprise: it was decided that "Torch," as this combined Anglo-U.S. operation came to be called, should begin the following autumn.

Already, on July 17, 1942, Churchill had had to notify Stalin that convoys of Allied supplies to northern Russia must be suspended because of German submarine activity on the Arctic sea route (on June 2 a convoy from Iceland had lost 23 out of 34 vessels). Consequently, it was the more awkward to inform Stalin that there would be no second front in Europe before 1943. In mid-August 1942, when Churchill went to Moscow to break the news, Stalin raged against the retreat from the plan for a second front in Europe but had to admit the military logic of "Torch."

Libya and Egypt, autumn 1941–summer 1942. In the Western Desert, a major offensive against Rommel's front was undertaken on Nov. 18, 1941, by the British 8th Army, commanded by Cunningham under the command in chief of Wavell's successor in the Middle East, General Sir Claude Auchinleck. The offensive was routed. General Neil Methuen Ritchie took Cunningham's place on November 25, still more tanks were brought up, and a fortnight's resumed pressure constrained Rommel to evacuate Cyrenaica and to retreat to Agedabia. There, however, Rommel was at last, albeit meagrely, reinforced; and, after repulsing a British attack on December 26, he prepared a counteroffensive. When the British still imagined his forces to be hopelessly crippled, he attacked on Jan. 21, 1942, and, by a series of strokes, drove the 8th Army back to the Gazala–Bir Hakeim line, just west of Tobruk.

Both sides were subsequently further reinforced. Then, on the night of May 26–27, Rommel passed around Ritchie's southern flank with his three German divisions and two Italian ones, leaving only four Italian divisions to face the Gazala line. Though at first Rommel did some damage to the British tanks as they came into action piecemeal from a weak position, he failed to break through to the coast behind Gazala. In a single day one-third of Rommel's tank force was lost; and, after another unsuccessful effort to reach the coast, he decided, on May 29, to take up a defensive position.

The Cauldron

The new German position, aptly known as the Cauldron, seemed indeed to be perilously exposed; and throughout the first days of June the British attacked it continually from the air and from the ground, imagining that Rommel's armour was caught at last. The British tanks, however, persisted in making direct assaults in small groups against the Cauldron and were beaten off with very heavy losses; and Rommel, meanwhile, secured his rear and his line of supply by overwhelming several isolated British positions to the south.

Whereas in May 1942 the British had had 700 tanks, with 200 more in reserve, against Rommel's 525, by June 10 their present armoured strength was reduced, through their wasteful tactics against the Cauldron, to 170, and most of the reserve was exhausted. Suddenly then, on June 11, Rommel struck eastward, to catch most of the remaining British armour in the converging fire of two panzer divisions. By nightfall on June 13 the British had barely 70 tanks left, and Rommel, with some 150 still fit for action, was master of the battlefield.

The British on June 14 began a precipitate retreat from the Gazala line toward the Egyptian frontier. A garrison of 33,000 men, however, with an immense quantity of material, was left behind in Tobruk—on the retention of which Churchill characteristically and most unfortunately insisted in successive telegrams from London. Rommel's prompt reduction of Tobruk, achieved on June 21, 1942, was felt by Great Britain as a national disaster second only to the loss of Singapore; and 80 percent of the transport with which Rommel chased the remnant of the 8th Army eastward consisted of captured British vehicles.

At this point Auchinleck relieved Ritchie of his command and in a realistic and soldierly way ordered a general British retreat back to the Alamein area. By June 30 the German tanks were pressing against the British positions between el-Alamein (al-'Alamayn) and the Qattara Depression, some 60 miles west of Alexandria, after an advance of more than 350 miles from Gazala. Hitler and Mussolini could expect that within a matter of days Rommel would be the master of Egypt.

Rommel halted

The ensuing First Battle of el-Alamein, which lasted throughout July 1942, marked the end of the German hopes of a rapid victory. Rommel's troops, having come so far and so fast, were exhausted; their first assaults failed to break the defense rallied by Auchinleck; and they were also subjected to disconcerting counterstrokes. At this point, the respite that Rommel had to grant to his men gave Auchinleck time to bring up reinforcements. By the end of July Rommel knew that it was he rather than Auchinleck who was now on the defensive.

Auchinleck had saved Egypt by halting Rommel's invasion, but his counterattacks had not driven it back. Early in August, when Churchill arrived in Cairo to review the situation, Auchinleck insisted on postponing the resumption of the offensive until September, so that his new forces could be properly acclimatized and trained for desert warfare. Impatient of this delay, Churchill removed Auchinleck from the command in chief in the Middle East and gave the post to General Sir Harold Alexander, while the command of the 8th Army was transferred eventually (after the sudden death of Churchill's first nominee) to General Bernard Law Montgomery. Paradoxically, Montgomery postponed the resumption of the offensive even longer than Auchinleck had desired.

While the British in the course of August raised their strength in armour at the front to some 700 tanks, Rommel received only meagre reinforcement in the shape of infantry. He had, however, about 200 gun-armed German tanks and also 240 Italian tanks (of an obsolete model). With this armament, in the night of Aug. 30–31, 1942, he launched a fresh attack, intending to capture by surprise the minefields on the southern sector of the British front and then to drive eastward with his armour for some 30 miles before wheeling north into the 8th Army's supply area on the coast. In the event, the minefields proved unexpectedly deep, and by daybreak Rommel's spearhead was only eight miles beyond them. Delayed on their eastward drive and already under attack from the air, the two German panzer divisions of the Afrika Korps had to make their wheel to the north at a much shorter distance from the breach than Rommel had planned. Their assault thus ran mainly into the position held by the British 22nd Armoured Brigade, to the southwest of the ridge 'Alam al-Halfa'. Shortage of fuel on the German side and reinforced defense on the British, together with intensification of the British bombing, spelled the defeat of the offensive, and Rommel on September 2 decided to make a gradual withdrawal.

The Germans' summer offensive in southern Russia, 1942. The German plan to launch another great summer offensive crystallized in the early months of 1942. Hitler's decision was influenced by his economists, who mistakenly told him that Germany could not continue the war unless it obtained petroleum supplies from the Caucasus. Hitler was the more responsive to such arguments because they coincided with his belief that another German offensive would so drain the Soviet Union's manpower that the U.S.S.R. would be unable to continue the war. His thinking was shared by his generals, who had been awed by the prodigality with which the Soviets squandered their troops in the fighting of 1941 and the spring of 1942. By this time at least 4,000,000 Soviet troops had been killed, wounded, or captured, while German casualties totaled only 1,150,000.

In the early summer of 1942 the German southern line ran from Orël southward east of Kursk, through Belgorod, and east of Kharkov down to the loop of the Soviet salient opposite Izyum, beyond which it veered southeastward to Taganrog, on the northern coast of the Sea of Azov. Before the Germans were ready for their principal offensive, the Red Army in May started a drive against Kharkov; but this premature effort actually served the Germans' purposes, since it not only preempted the Soviet reserves but also provoked an immediate counterstroke against its southern flank, where the Germans broke into the salient and reached the Donets River near Izyum. The Germans captured 240,000 Soviet prisoners in the encirclement that

followed. In May also the Germans drove the Soviet defenders of the Kerch Peninsula out of the Crimea; and on June 3 the Germans began an assault against Sevastopol, which, however, held out for a month.

The Germans' crossing of the Donets near Izyum on June 10, 1942, was the prelude to their summer offensive, which was launched at last on June 28: Field Marshal Maximilian von Weichs's Army Group B, from the Kursk–Belgorod sector of the front, struck toward the middle Don River opposite Voronezh, whence General Friedrich Paulus' 6th Army was to wheel southeastward against Stalingrad (Volgograd); and List's Army Group A, from the front south of Kharkov, with Kleist's 1st Panzer Army, struck toward the lower Don to take Rostov and to thrust thence northeastward against Stalingrad as well as southward into the vast oil fields of Caucasia. Army Group B swept rapidly across a 100-mile stretch of plain to the Don and captured Voronezh on July 6. The 1st Panzer Army drove 250 miles from its starting line and captured Rostov on July 23. Once his forces had reached Rostov, Hitler decided to split his troops so that they could both invade the rest of the Caucasus and take the important industrial city of Stalingrad on the Volga River, 220 miles northeast of Rostov. This decision was to have fatal consequences for the Germans, since they lacked the resources to successfully take and hold both of these objectives.

The German thrust into Caucasia

Maikop (Maykup), the great oil centre 200 miles south of Rostov, fell to Kleist's right-hand column on August 9, and Pyatigorsk, 150 miles east of Maikop, fell to his centre on the same day, while the projected thrust against Stalingrad, in the opposite direction from Rostov, was being developed. Shortage of fuel, however, slowed the pace of Kleist's subsequent southeastward progress through the Caucasian mountains; and, after forcing a passage over the Terek River near Mozdok early in September, he was halted definitively just south of that river. From the end of October 1942 the Caucasian front was stabilized; but the titanic struggle for Stalingrad, draining manpower that might have won victory for the Germans in Caucasia, was to rage on, fatefully, for three more months (see below). Already, however, it was evident that Hitler's new offensive had fallen short of its objectives, and the scapegoat this time was Halder, who was superseded by Kurt Zeitzler as chief of the army general staff.

The Allies' first decisive successes

The Solomons, Papua, Madagascar, the Aleutians, and Burma, July 1942–May 1943. On July 2, 1942, the U.S. Joint Chiefs of Staff ordered limited offensives in three stages to recapture the New Britain–New Ireland–Solomons–eastern New Guinea area: first, the seizure of Tulagi and of the Santa Cruz Islands, with adjacent positions; second, the occupation of the central and northern Solomons and of the northeast coast of New Guinea; third, the seizure of Rabaul and of other points in the Bismarck Archipelago.

Guadalcanal

On July 6 the Japanese landed troops on Guadalcanal, one of the southern Solomons, and began to construct an air base. The Allied high command, fearing further Japanese advances southeastward, sped into the area to dislodge the enemy and to obtain a base for later advances toward Japan's main base in the theatre, Rabaul. The U.S. 1st Marine Division poured ashore on August 7 and secured Guadalcanal's airfield, Tulagi's harbour, and neighbouring islands by dusk on August 8—the Pacific war's first major Allied offensive. During the night of August 8–9, Japanese cruisers and destroyers, attempting to hold Guadalcanal, sank four U.S. cruisers, themselves sustaining one cruiser sunk and one damaged and later sunk. On August 23–25, in the Battle of the Eastern Solomons, the Japanese lost a light carrier, a destroyer, and a submarine and sustained damage to a cruiser and to a seaplane carrier but sank an Allied destroyer and crippled a cruiser. On August 31 another U.S. carrier was disabled, and on September 15 Japanese submarines sank the carrier *Wasp* and damaged a battleship. Meanwhile, more than 6,000 Japanese reinforced their Guadalcanal garrison, attacking the Marines' beachhead on August 20–21 and on September 12–14. On September 18 some U.S. reinforcements arrived, and mid-October saw about 22,000 Japanese ranged against 23,000 U.S. troops. The sea battles of Cape Esperance and of the Santa Cruz Islands—in which two Japanese cruisers and two destroyers were sunk and three carriers and two destroyers damaged in return for the loss of one U.S. carrier and two destroyers, besides damage to six other Allied ships—thwarted an attempt to reinforce further the Japanese ground troops, whose attack proved a failure (October 20–29).

After October, Allied strength was built up. Another Japanese attempt at counter-reinforcement led to the naval Battle of Guadalcanal, fought on November 13–15: it cost Japan two battleships, three destroyers, one cruiser, two submarines, and 11 transports and the Allies (now under Admiral William F. Halsey) two cruisers and seven destroyers sunk and one battleship and one cruiser damaged. Only 4,000 Japanese troops out of 12,500 managed to reach land, without equipment; and on November 30 eight Japanese destroyers, attempting to land more troops, were beaten off in the Battle of Tassafaronga, losing one destroyer sunk and one crippled, at an Allied cost of one cruiser sunk and three damaged.

By Jan. 5, 1943, Guadalcanal's Allied garrison totaled 44,000, against 22,500 Japanese. The Japanese decided to evacuate the position, carrying away 12,000 men in early February in daring destroyer runs. In ground warfare Japanese losses were more than 24,000 for the Guadalcanal campaign, Allied losses about 1,600 killed and 4,250 wounded (figures that ignore the higher number of casualties from disease). On February 21, U.S. infantry began occupying the Russell Islands, to support advances on Rabaul.

Earlier, before Allied plans to secure eastern New Guinea had been implemented, the Japanese had landed near Gona on the north coast of Papua (the southeastern extremity of the great island) on July 24, 1942, in an attempt to reach Port Moresby overland, via the Kokoda Trail. Advanced Japanese units from the north, despite Australian opposition, had reached a ridge 32 miles from Port Moresby by mid-September. Then, however, they had to withdraw exhausted to Gona and to nearby Buna, where there were some 7,500 Japanese assembled by November 18. The next day U.S. infantry attacked them there. Each side was subsequently reinforced; but the Australians took Gona on December 9 and the Americans Buna village on December 14. Buna government station fell to the Allies on Jan. 2, 1943, Sanananda on January 18, and all Japanese resistance in Papua ceased on January 22.

The defense of Port Moresby

The retaking of Guadalcanal and Papua ended the Japanese drive south, and communications with Australia and New Zealand were now secure. Altogether, Papua cost Japan nearly 12,000 killed and 350 captured. Allied losses were 3,300 killed and 5,500 wounded. Allied air forces had played a particularly important role, interdicting Japanese supply lines and transporting Allied supplies and reinforcements.

Japan, having lost Guadalcanal, fought henceforth defensively, with worsening prospects. Its final effort to reinforce the Lae–Salamaua position in New Guinea from the stronghold of Rabaul was a disaster: in the Battle of the Bismarck Sea, on March 2–4, 1943, the Japanese lost four destroyers and eight transports, and only 1,000 of the 7,000 troops reached their destination. On March 25 the Japanese Army and Navy high commands agreed on a policy of strengthening the defense of strategic points and of counterattacking wherever possible, priority being given to the defense of the remaining Japanese positions in New Guinea, with secondary emphasis on the Solomon Islands. In the following three weeks, however, the Allies improved their own position in New Guinea, and Japanese intervention was confined to air attacks. Before the end of April, moreover, the Japanese Navy sustained a disaster: the guiding genius of the Japanese war effort, Yamamoto, was sent late in March to command the forces based on Rabaul but was killed in an American air ambush on a flight to Bougainville.

The Battle of the Bismarck Sea

Developments of the Allies' war against Japan also took

place outside the southwest Pacific area. British forces in the summer of 1942 invaded Vichy French-held Madagascar. A renewed British offensive in September 1942 overran the island; hostilities ceased on November 5, and a Free French administration of Madagascar took office on Jan. 8, 1943. In the North Pacific, meanwhile, the United States had decided to expel the Japanese from the Aleutians. Having landed forces on Adak in August 1942, they began air attacks against Kiska and Attu from Adak the next month and from Amchitka also in the following January, while a naval blockade prevented the Japanese from reinforcing their garrisons. Finally, U.S. troops, bypassing Kiska, invaded Attu on May 11, 1943—to kill most of the island's 2,300 defenders in three weeks of fighting. The Japanese then evacuated Kiska. Bases in the Aleutians thenceforth facilitated the Allies' bombing of the Kuril Islands.

Attu and Kiska

Burma, autumn 1942–summer 1943. On the Burmese front the Allies found they could do little to dislodge the Japanese from their occupation of that country, and what little the Allies did attempt proved abortive. Brigadier General Orde Wingate's "Chindits," which were long-range penetration groups depending on supplies from the air, crossed the Chindwin River in February 1943 and were initially successful in severing Japanese communications on the railroad between Mandalay and Myitkyina. But the Chindits soon found themselves in unfavourable terrain and in grave danger of encirclement, and so they made their way back to India.

In May 1943, however, the Allies reorganized their system of command for Southeast Asia. Vice Admiral Lord Louis Mountbatten was appointed supreme commander of the South East Asia Command (SEAC), and Stilwell was appointed deputy to Mountbatten. Stilwell at the same time was chief of staff to Chiang Kai-shek. The British–Indian forces destined for Burma meanwhile constituted the 14th Army, under Lieutenant General William Slim, whose operational control Stilwell agreed to accept. Shortly afterward, Auchinleck succeeded Wavell as commander in chief in India.

Montgomery's Battle of el-Alamein and Rommel's retreat, 1942–43. While Churchill was still chafing in London about his generals' delay in resuming the offensive in Egypt, Montgomery waited for seven weeks after 'Alam al-Halfa' in order to be sure of success. He finally chose to begin his attack in the night of Oct. 23–24, 1942, when there would be moonlight for the clearing of gaps in the German minefields.

Strengths of the rival forces in North Africa

By mid-October the British 8th Army had 230,000 men and 1,230 gun-armed tanks ready for action, while the German–Italian forces numbered only 80,000 men, with only 210 tanks of comparable quality ready; and in air support the British enjoyed a superiority of 1,500 to 350. Allied air and submarine attacks on the Axis supply lines across the Mediterranean, moreover, had prevented Rommel's army from receiving adequate replenishments of fuel, ammunition, and food; and Rommel himself, who had been ill before 'Alam al-Halfa', was convalescing in Austria.

The British launched their infantry attack at el-Alamein at 10:00 PM on Oct. 23, 1942, but found the German minefields harder to clear than they had foreseen. Two days later, however, some of those tanks were deploying six miles beyond the original front. When Rommel, ordered back to Africa by Hitler, reached the front in the evening of October 25, half of the Germans' available armour was already destroyed. Nevertheless, the impetus of the British onslaught was stopped the next day, when German antitank guns took a heavy toll of armour trying to deepen the westward penetration. In the night of October 28 Montgomery turned the offensive northward from the wedge, but this drive likewise miscarried. In the first week of their offensive the British lost four times as many tanks as the Germans but still had 800 available against the latter's remaining 90.

When Montgomery switched the British line of attack back to its original direction, early on Nov. 2, 1942, Rommel was no longer strong enough to withstand him. After expensive resistance throughout the daytime, he ordered a retreat to Fūka (Fūkah); but in the afternoon of November 3 the retreat was fatally countermanded by Hitler, who insisted that the Alamein position be held. The 36 hours wasted in obeying this long-distance instruction cost Rommel his chance of making a stand at Fūka: when he resumed his retreat, he had to race much farther back to escape successive British attempts to intercept him on the coast road by scythelike sweeps from the south. A fortnight after resuming his withdrawal from el-Alamein, Rommel was 700 miles to the west, at the traditional backstop of Agheila. As the British took their time to mount their attacks, he fell back farther by stages: after three weeks, 200 miles to Buerat (al-Bu'ayrāt); after three more weeks, in mid-January 1943, the whole distance of 350 miles past Tripoli to the Mareth Line within the frontiers of Tunisia. By that time the Axis position in Tunisia was being battered from the west, through the execution of "Torch."

Rommel on the Mareth Line

Stalingrad and the German retreat, summer 1942–February 1943. The German 4th Panzer Army, after being diverted to the south to help Kleist's attack on Rostov late in July 1942 (see above), was redirected toward Stalingrad a fortnight later. Stalingrad was a large industrial city producing armaments and tractors; it stretched for 30 miles along the banks of the Volga River. By the end of August the 4th Army's northeastward advance against the city was converging with the eastward advance of the 6th Army, under General Friedrich Paulus, with 330,000 of the German Army's finest troops. The Red Army, however, put up the most determined resistance, yielding ground only very slowly and at a high cost as the 6th Army approached Stalingrad. On August 23 a German spearhead penetrated the city's northern suburbs, and the Luftwaffe rained incendiary bombs that destroyed most of the city's wooden housing. The Soviet 62nd Army was pushed back into Stalingrad proper, where, under the command of General Vasily I. Chuikov, it made a determined stand. Meanwhile, the Germans' concentration on Stalingrad was increasingly draining reserves from their flank cover, which was already strained by having to stretch so far—400 miles on the left (north), as far as Voronezh, 400 again on the right (south), as far as the Terek River. By mid-September the Germans had pushed the Soviet forces in Stalingrad back until the latter occupied only a nine-mile-long strip of the city along the Volga, and this strip was only two or three miles wide. The Soviets had to supply their troops by barge and boat across the Volga from the other bank. At this point Stalingrad became the scene of some of the fiercest and most concentrated fighting of the war; streets, blocks, and individual buildings were fought over by many small units of troops and often changed hands again and again. The city's remaining buildings were pounded into rubble by the unrelenting close combat. The most critical moment came on October 14, when the Soviet defenders had their backs so close to the Volga that the few remaining supply crossings of the river came under German machine-gun fire. The Germans, however, were growing dispirited by heavy losses, by fatigue, and by the approach of winter.

A huge Soviet counteroffensive, planned by generals G.K. Zhukov, A.M. Vasilevsky, and Nikolay Nikolayevich Voronov, was launched on Nov. 19–20, 1942, in two spearheads, north and south of the German salient whose tip was at Stalingrad. The twin pincers of this counteroffensive struck the flanks of the German salient at points about 50 miles north and 50 miles south of Stalingrad and were designed to isolate the 250,000 remaining men of the German 6th and 4th armies in the city. The attacks quickly penetrated deep into the flanks, and by November 23 the two prongs of the attack had linked up about 60 miles west of Stalingrad; the encirclement of the two German armies in Stalingrad was complete. The German high command urged Hitler to allow Paulus and his forces to break out of the encirclement and rejoin the main German forces west of the city, but Hitler would not contemplate a retreat from the Volga River and ordered Paulus to "stand and fight." With winter setting in and food and medical supplies dwindling, Paulus' forces grew weaker. In mid-December Hitler allowed one of the most talented German commanders, Field Marshal Erich von

Manstein, to form a special army corps to rescue Paulus' forces by fighting its way eastward, but Hitler refused to let Paulus fight his way westward at the same time in order to link up with Manstein. This fatal decision doomed Paulus' forces, since the main German forces now simply lacked the reserves needed to break through the Soviet encirclement singlehandedly. Hitler exhorted the trapped German forces to fight to the death, but on Jan. 31, 1943, Paulus surrendered; 91,000 frozen, starving men (all that was left of the 6th and 4th armies) and 24 generals surrendered with him.

Besides being the greatest battle of the war, Stalingrad proved to be the turning point of the military struggle between Germany and the Soviet Union. The battle used up precious German reserves, destroyed two entire armies, and humiliated the prestigious German war machine. It also marked the increasing skill and professionalism of a group of younger Soviet generals who had emerged as capable commanders, chief among whom was Zhukov.

Meanwhile, early in January 1943, only just in time, Hitler acknowledged that the encirclement of the Germans in Stalingrad would lead to an even worse disaster unless he extricated his forces from the Caucasus. Kleist was therefore ordered to retreat, while his northern flank of 600 miles was still protected by the desperate resistance of the encircled Paulus. Kleist's forces were making their way back across the Don at Rostov when Paulus at last surrendered. Had Paulus surrendered three weeks earlier (after seven weeks of isolation), Kleist's escape would have been impossible.

German retreat from Moscow

Even west of Rostov there were threats to Kleist's line of retreat. In January, two Soviet armies, the one under General Nikolay Fyodorovich Vatutin, the other under General Filipp Ivanovich Golikov, had crossed the Don upstream from Serafimovich and were thrusting southwestward to the Donets between Kamensk and Kharkov: Vatutin's forces, having crossed the Donets at Izyum, took Lozovaya Junction on February 11, Golikov's took Kharkov five days later. Farther to the north, a third Soviet army, under General Ivan Danilovich Chernyakhovsky, had initiated a drive westward from Voronezh on February 2 and had retaken Kursk on February 8. Thus, the Germans had to retreat from all the territory they had taken in their great summer offensive in 1942. The Caucasus returned to Soviet hands.

A sudden thaw supervened to hamper the Red Army's transport of supplies and reinforcements across the swollen courses of the great rivers. With the momentum of the Soviet counteroffensive thus slowed, the Germans made good their retreat to the Dnepr along the easier routes of the Black Sea littoral and were able, before the end of February 1943, to mount a counteroffensive of their own.

The invasion of northwest Africa, November–December 1942. When the U.S. and British strategists had decided on "Torch" (Allied landings on the western coast of North Africa) late in July 1942, it remained to settle the practical details of the operation. The purpose of "Torch" was to hem Rommel's forces in between U.S. troops on the west and British troops to the east. After considerable discussion, it was finally agreed that landings, under the supreme command of Major General Dwight D. Eisenhower, should be made on November 8 at three places in the vicinity of Casablanca on the Atlantic coast of Morocco and on beaches near Oran and near Algiers itself on the Mediterranean coast of Algeria. The amphibious landings would involve a total of about 110,000 troops, most of them Americans.

The conciliation of the French on whose colonial territory the landings would be made was a more delicate matter. All of French North Africa was still loyal to the Vichy government of Marshal Pétain, with which the United States, unlike Great Britain, was still formally maintaining diplomatic relations. Thus, the French commander in chief in Algeria, General Alphonse Juin, and his counterpart in Morocco, General Charles-Auguste Noguès, were subordinate to the supreme commander of all Vichy's forces, namely Admiral Jean-François Darlan. American diplomats and generals tried to gain these officers' collaboration with the Allies in the landings, for it was vital to try to avoid a situation in which Vichy French troops put up armed resistance to the landings at the beaches.

French political complications

The U.S.-British landings at Algiers began on November 8 and were met by little French resistance. The simultaneous landings near Oran met stiffer resistance, and on November 9 the whole U.S. plan of operations was dislocated by a French counterattack on the Arzew beachhead. Around Casablanca the U.S. landings were accomplished without difficulty, but resistance developed when the invaders tried to expand their beachheads. On November 10, however, the fighting was called off; and next day the French authorities in Morocco concluded an armistice with the Americans.

The landing in Algiers, meanwhile, was complicated by the fact that Darlan himself was in the city at the time. The situation was muddled, with some French troops loyal to Pétain while others backed de Gaulle and the anti-Vichy French general whom the Allies were sponsoring in North Africa, Henri Giraud.

On Nov. 11, 1942, in reaction to the Allied landings, German and Italian forces overran southern France, the metropolitan territory hitherto under Pétain's immediate authority. This event helped induce Noguès and the other French commanders in Algeria to assent to Darlan's proposals for a working agreement with the Allies, including recognition of Giraud as military commander in chief of the French forces. Concluded on November 13, the agreement was promptly endorsed by Eisenhower. French West Africa, including Senegal, with the port of Dakar, likewise followed Darlan's lead. The Germans, however, by mining the exit from the harbour of Toulon, forestalled plans for the escape of the main French fleet from metropolitan France to North Africa: on November 27, the French crews scuttled their ships to avoid capture. On Dec. 24, 1942, Darlan was assassinated; both Royalist and Gaullist circles in North Africa had steadfastly objected to him on political grounds. Giraud thereupon took his place, for a time, as French high commissioner in North Africa.

Tunisia, November 1942–May 1943. Axis troops had begun to arrive in Tunisia as early as Nov. 9, 1942, and were reinforced in the following fortnight until they numbered about 20,000 combat troops (which were subsequently heavily reinforced by air). Thus, when the British general Kenneth Anderson, designated to command the invasion of Tunisia from the west with the Allied 1st Army, started his offensive on November 25, the defense was unexpectedly strong. By December 5 the 1st Army's advance was checked a dozen miles from Tunis and from Bizerte. Further reinforcements enabled Colonel General Jürgen von Arnim, who assumed the command in chief of the Axis defense in Tunisia on December 9, to expand his two bridgeheads in Tunisia until they were merged into one. Germany and Italy had won the race for Tunis but were henceforth to succumb to the lure of retaining their prize regardless of the greater need of conserving their strength for the defense of Europe.

After Rommel had fallen back from Libya to the Mareth Line in mid-January 1943 (see above), two German armies, Arnim's and Rommel's, were holding the north and the south of the eastern littoral both against Anderson's 1st Army attacking from the west and against Montgomery's 8th from the southeast. Rommel judged that a counterstroke should be delivered first against the Allies in the west. Accordingly, on February 14 the Axis forces delivered a major attack against U.S. forces between the Fā'iḍ Pass in the north and Gafsa in the south. West of Fā'iḍ, the 21st Panzer Division, under General Heinz Ziegler, destroyed 100 U.S. tanks and drove the Americans back 50 miles. In the Kasserine Pass, however, the Allies put up some stiffer opposition.

Rommel's frustrated counterstrokes

When on February 19 Rommel received authority to continue his attack, he was ordered to advance not against Tébessa but northward from Kasserine against Thala—where, in fact, Alexander was expecting him. Having overcome the stubborn U.S. resistance in the Kasserine Pass on February 20, the Germans entered Thala the next day, only to be expelled a few hours later by Alexander's reserve troops. His chance having been forfeited, Rommel began a gradual withdrawal on February 22.

The delays ensuing from the frustration of Rommel's stroke against the 1st Army reduced the effectiveness of his stroke against the 8th. Whereas on Feb. 26, 1943, Montgomery had had only one division facing the Mareth Line, he quadrupled his strength in the following week, massing 400 tanks and 500 antitank guns. Rommel's attack, on March 6, was brought to an early halt, and 50 German tanks were lost. A sick man and a disappointed soldier, Rommel relinquished his command.

The Allied 1st Army resumed the offensive on March 17, with attacks by the U.S. II Corps, under General George Patton, on the roads through the mountains, with the aim of cutting the Afrika Korps' line of retreat up the coast to Tunis; but these attacks were checked by the Germans in the passes. In the night of March 20–21, however, the British 8th Army launched a frontal assault on the Mareth Line, combined with an outflanking movement by the New Zealand Corps toward el-Hamma (al-Ḥāmmah) in the Germans' rear; and a few days later, seeing the frontal assault to have failed, Montgomery switched the main weight of his attack to the flank. Threatened with encirclement, the Germans decided to abandon the Mareth Line, which the 8th Army occupied on March 28; but the German defenses at el-Hamma held out long enough to enable the rest of the Afrika Korps to retreat without much loss to a new line on the Wādī al-'Akārīt, north of Gabès. The new line, however, was breached by the 8th Army on April 6; and, meanwhile, the Americans were also advancing on the Axis troops' rear from Gafsa. By the following morning the Afrika Korps was retreating rapidly northward along the littoral toward Tunis, and by April 11 it had joined hands with Arnim's forces for the defense of a 100-mile perimeter stretching around Tunis and Bizerte (Banzart).

The Mareth Line abandoned

Thanks to the rapidity of the Afrika Korps' retreat from Wādī al-'Akārīt, the German high command had an opportunity to withdraw its forces from the rump of Tunisia to Sicily, but it chose instead to defend the indefensible rump. The defenders indeed withstood the converging assaults that the 8th and 1st armies delivered against the perimeter from April 20 to April 23; but on May 6 a concentrated attack by Allied artillery, aircraft, infantry, and tanks was launched on the two-mile front of the Medjerda (Majardah) Valley leading to Tunis; and on May 7 the city fell to the leading British armoured forces, while the Americans and the French almost simultaneously captured Bizerte. At the same time, the Germans' line of retreat into the Cap Bon Peninsula was severed by an armoured division's swift turn southeastward from Tunis. A general collapse of the German resistance followed, the Allies taking more than 250,000 prisoners, including 125,000 German troops and Arnim himself. North Africa had been cleared of Axis forces and was now completely in Allied hands. Its capture insured the safety of Allied shipping and naval movements throughout the Mediterranean, and North Africa would serve as a base for future Allied operations against Italy itself.

The Atlantic, the Mediterranean, and the North Sea, 1942–45. The year 1942 was, on the whole, a favourable one for the German U-boats. First, the U.S. entry into the war entitled them to infest the U.S. coast of the North Atlantic; and it was not until the middle of the year that the Allies' introduction of the convoy system from the Caribbean northward constrained the raiders to go so far afield as the waters between Brazil and West Africa. Second, U-tankers were developed; *i.e.,* large converted U-boats equipped to provide fuel, torpedoes, and other supplies to U-boats operating in remote waters. In the course of 1942, the U-boats sank more than 6,266,000 tons of shipping; and, since in the same period their operational strength rose from 91 to 212, it seemed conceivable that they might soon score their desired target of 800,000 tons of sinkings per month.

March 1943 saw the climax of the U-boats' good fortune: their strength rose to 240; they sank in that single month 627,377 tons of shipping; and, in the greatest convoy battle of the war, when 20 of them attacked two convoys merged into one, they sank 21 ships (141,000 tons) out of 77 with the loss of only one of their own number. The anticlimax followed, thanks to five developments of the Allies' counteraction: "support groups" were reintroduced; aircraft carriers became progressively available for escorts; more and more long-range Liberator aircraft began to cover the convoys offshore; ships were equipped with a radar set of very short wavelength, the probing of which was undetectable to the U-boats; and a regular offensive against U-boats on their transit routes was launched from the air (56 were destroyed in April–May 1943). The U-boats sank 327,943 tons in April, 264,852 in May, only 95,753 in June 1943; and for the rest of the war monthly totals were less than 100,000 tons except in July and September 1943 and in March 1944.

Late in 1944 the U-boats were equipped with the snorkel breathing tube, which provided them with the necessary oxygen to recharge their batteries under water and so converted them from submersible torpedo boats into almost complete submarines virtually undetectable to radar. About the same time a new model of U-boat, with greater underwater speed and endurance, came into operation. These improvements came too late, however, because the Allies' surface and air resources for the protection of the convoys were already overwhelming.

Air warfare, 1942–43. Early in 1942 the RAF bomber command, headed by Sir Arthur Harris, began an intensification of the Allies' growing strategic air offensive against Germany. These attacks, which were aimed against factories, rail depots, dockyards, bridges, and dams and against cities and towns themselves, were intended to both destroy Germany's war industries and to deprive its civilian population of their housing, thus sapping their will to continue the war. The characteristic feature of the new program was its emphasis on area bombing, in which the centres of towns would be the points of aim for nocturnal raids.

Already in March 1942 an exceptionally destructive bombing raid, using the Germans' own incendiary method, had been made on Lübeck; and intensive attacks were also made on Essen (site of the Krupp munitions works) and other Ruhr towns. In the night of May 30–31 more than 1,000 bombers were dispatched against Cologne, where they did heavy damage to one-third of that city's built-up area. Such operations, however, became highly expensive to the bomber command, particularly because of the defense put up by the German night fighter force. Interrupted for two months during which the bombers concentrated their attention on U-boat bases on the Bay of Biscay, the air offensive against Germany was resumed in March 1943. In the following 12 months, moreover, its resources were to be increased formidably, so that by March 1944 the bomber command's average daily operational strength had risen to 974 from about 500 in 1942. These numbers helped the RAF to concentrate effectively against major industrial targets, such as those in the Ruhr. The phases of the resumed offensive were: (1) the Battle of the Ruhr, from March to July 1943, comprising 18,506 sorties and costing 872 aircraft shot down and 2,126 damaged, its most memorable operation being that of the night of May 16–17, when the Möhne Dam in the Ruhr Basin and the Eder Dam in the Weser Basin were breached, (2) the Battle of Hamburg, from July to November 1943, comprising 17,021 sorties and costing 695 bombers lost and 1,123 damaged but, nevertheless, thanks in part to the new Window antiradar and "H_2S" radar devices, achieving an unprecedented measure of devastation, since four out of its 33 major actions, with a little help from minor attacks, killed about 40,000 people and drove nearly 1,000,000 from their homes, and (3) the Battle of Berlin, from November 1943 to March 1944, comprising 20,224 sorties but costing 1,047 bombers lost and 1,682 returned damaged and achieving, on the whole, less devastation than the Battle of Hamburg.

The Ruhr, Hamburg, and Berlin bombed

The U.S. 8th Air Force, based in Great Britain, also took part in the strategic offensive against Germany from January 1943. Its bombers, Flying Fortresses and Liberators, attacked industrial targets in daylight. They proved, however, to be very vulnerable to German fighter attack whenever they went beyond the range of their own escort of fighters—that is to say, farther than the distance from Norfolk to Aachen: the raid against the important ball-

bearing factory at Schweinfurt, for instance, on Oct. 14, 1943, lost 60 out of the 291 bombers participating, and 138 of those that returned were damaged. Not until December 1943 was the P-51B (Mustang III) brought into operation with the 8th Air Force—a long-range fighter that portended a change in the balance of air power. The Germans, meanwhile, continued to increase their production of aircraft and, in particular, of their highly successful fighters.

German-occupied Europe. Hitler's racist ideology and his brutal conception of power politics caused him to pursue certain aims in those European countries conquered by the Germans in the period 1939–42. Hitler intended that those western and northern European areas in which civil administrations were installed—The Netherlands and Norway—would at some later date become part of the German Reich, or nation. Those countries left by Germany under military administration (which originally had been imposed everywhere), such as France and Serbia, would eventually be included more loosely in a German-dominated European bloc. Poland and the Soviet Union, on the other hand, were to be a colonial area for German settlement and economic exploitation.

Without regard to these distinctions, the SS, the elite corps of the Nazi Party, possessed exceptional powers throughout German-dominated Europe and in the course of time came to perform more and more executive functions, even in those countries under military administration. Similarly, the powers that Hitler gave to his chief labour commissioner, Fritz Sauckel, for the compulsory enrollment of foreign workers into the German armaments industry were soon applied to the whole of German-dominated Europe and ultimately turned 7,500,000 people into forced or slave labourers. Above all, however, there was the Final Solution of the "Jewish question" as ordered by Hitler, which meant the physical extermination of the Jewish people throughout Europe wherever German rule was in force or where German influence was decisive.

The Final Solution—that is to say the step beyond half-measures such as the concentration of Poland's Jews into overcrowded ghettoes—was introduced concurrently with Germany's preparations for the military campaign against the Soviet Union, since Hitler believed that the annihilation of the Communists entailed not only the extermination of the Soviet ruling class but also what he believed to be its "biological basis"—the millions of Jews in western Russia and the Ukraine. Accordingly, with the start of the invasion of the Soviet Union in 1941, special mobile killing squads began systematically shooting the Jewish population on conquered Soviet territory in the rear of the advancing German armies; in a few months, up to the end of 1941, they had killed about 1,400,000 people. Meanwhile, plans were made in 1941 to similarly exterminate the Jews of central and western Europe. At the Wannsee Conference of Nazi and SS chiefs in January 1942, it was agreed that those Jews would be deported and sent to camps in eastern Poland where they would be killed en masse or made to work as slave labourers until they perished. In the period from May 1942 to September 1944 more than 4,200,000 Jews were killed in such death camps as Auschwitz (Oświęcim), Treblinka, Belzec, Chełmno, Majdanek, and Sobibor. About 5,700,000 Jews died in the course of the Final Solution.

While Hitler destined the Jews in his empire to physical extermination, he regarded the Slavs, principally the Poles and the Russians, as "subhumans" who were to be subjected to continual decimation and used as a pool of cheap labour, that is to say, reduced to slavery. Poland became the training ground for this purpose. Upon the German conquest of Poland in 1939, Hitler ordered the SS to kill a large proportion of the Polish intelligentsia. A reign of terror against the nationalistic-minded Polish ruling classes began, and by the war's end a total of 3,000,000 Poles (in addition to 3,000,000 Polish Jews) had been killed. Hitler further willed that the whole mass of Slavs and Balts in the occupied portions of the Soviet Union should be indiscriminately subjected to German domination and should be economically exploited without hindrance or compassion. In the event, the Ukraine was the major area subject to economic exploitation and also became the main source of slave labour. When the German armies first entered the Ukraine in July 1941, many Ukrainians had welcomed the Germans as their liberators from Stalinist terror and collectivization. But this goodwill soon turned to resentment as the Germans requisitioned large quantities of grain from the farms, forcibly deported several million Ukrainians for work in Germany, and engaged in brutal reprisals against civilians for acts of resistance or sabotage.

These inhumane occupation policies were practiced to a greater or lesser extent in all the countries occupied by the Germans, and the result was the beginning in 1940–41 of armed, underground resistance movements in those countries. Underground resistance was especially effective in the Soviet Union because it functioned behind fronts on which the German armies were still engaged in battle with the Red Army. The Soviet Partisans, as they were called, could thus covertly receive arms, equipment, and direction from the Soviet forces at the front itself. Soviet Partisans, like the members of other nations' Resistance movements, harassed and disrupted German military and economic activities by blowing up ammunition dumps and communications and transport facilities, sabotaging factories, ambushing small German units, and gathering military intelligence for use by the Allied armies. By 1944 the Resistance organizations in the Soviet Union, Poland, Yugoslavia, France, and Greece had grown quite large and were holding down many German divisions that were badly needed at the battlefront. In eastern Europe and Yugoslavia, the Resistance came to control large tracts of land in more inaccessible areas such as forests, mountain ranges, and swamplands. Some Resistance organizations, such as the Partisans in Yugoslavia and the National Liberation Movement in Greece, were Communist ones, while others, such as the Maquis in France and the Home Army in Poland, comprised people of many different political persuasions, though they were invariably anti-Fascists.

The German occupation authorities' attempts to eradicate the Resistance in most cases merely fanned the flames, due to the Germans' use of indiscriminate reprisals against civilians. It is generally agreed that by 1944 the Germans had earned the overwhelming antipathy of most of the people in the occupied nations of Europe. It should be noted, however, that the German occupation was in general far harsher in eastern Europe and the Balkans than in western Europe. In the Soviet Union, Poland, Yugoslavia, and Greece, a process of Resistance guerrilla warfare and Nazi reprisals began in 1941 and rose to a crescendo in 1943–44 as the fury of Nazi racism resulted in a war of annihilation upon the Slavic peoples.

Casablanca and Trident, January–May 1943. To decide what should be done after victory in North Africa, Roosevelt and Churchill, with their advisers, met at Casablanca in mid-January 1943. After long argument, it was eventually agreed that Sicily should be the next Axis area to be taken, in July. For the war against Japan, it was decided that two offensive operations should be undertaken: MacArthur should move toward the Japanese base at Rabaul, on the island of New Britain; and convergent movements on Burma should be made by the British from the mainland of India and by the Americans from the sea. Politically, the Casablanca Conference owes its importance to the fact that, at its end, Roosevelt publicly announced a demand for the unconditional surrender of Germany, Italy, and Japan.

Only four months after Casablanca it became necessary to hold another Anglo-U.S. conference. In mid-May 1943, Roosevelt, Churchill, and their advisers met, in Washington, D.C., for the conference code-named Trident. There the Sicilian project was effectively confirmed, and the date May 1, 1944, was prescribed—definitively in the U.S view, provisionally in the British—for the landing of 29 divisions in France; but the question whether the conquest of Sicily should be followed, as the British proposed, by an invasion of Italy was left unsettled.

The Eastern Front, February–September 1943. The German counteroffensive of February 1943 threw back the Soviet forces that had been advancing toward the Dnepr

River on the Izyum sector of the front, and by mid-March the Germans had retaken Kharkov and Belgorod and reestablished a front on the Donets River. Hitler also authorized the German forces to fall back, in March, from their advanced positions facing Moscow to a straighter line in front of Smolensk and Orël. Finally, there was the existence of the large Soviet bulge, or salient, around Kursk, between Orël and Belgorod, which extended for about 150 miles from north to south and protruded 100 miles into the German lines. This salient irresistibly tempted Hitler and Zeitzler into undertaking a new and extremely ambitious offensive instead of remaining content to hold their newly shortened front.

Hitler concentrated all efforts on this offensive without regard to the risk that an unsuccessful attack would leave him without reserves to maintain any subsequent defense of his long front. The Germans' increasing difficulty in building up their forces with fresh drafts of men and equipment was reflected in the increased delay that year in opening the summer offensive. Three months' pause followed the close of the winter campaign.

By contrast, the Red Army had improved much since 1942, both in quality and in quantity. The flow of new equipment had greatly increased, as had the number of new divisions, and its numerical superiority over the Germans was now about 4 to 1. Better still, its leadership had improved with experience: generals and junior commanders alike had become more skilled tacticians. That could already be discerned in the summer of 1943, when the Soviets waited to let the Germans lead off and commit themselves deeply to an offensive, and so stood well-poised to exploit the Germans' loss of balance in lunging.

The German offensive against the Kursk salient was launched on July 5, 1943, and into it Hitler threw 20 infantry divisions and 17 armoured divisions having a total of about 3,000 tanks. But the German tank columns got entangled in the deep minefields that the Soviets had laid, forewarned by the long preparation of the offensive. The Germans advanced only 10–30 miles, and no large bag of Soviet prisoners was taken, since the Red Army had withdrawn their main forces from the salient before the German attack began. After a week of effort the German armoured divisions were seriously reduced by the well-prepared Soviet antitank defenses in the salient. On July 12, as the Germans began to pull out, the Soviets launched a counteroffensive upon the German positions in the salient and met with great success, taking Orël on August 5. By this time the Germans had lost 2,900 tanks and 70,000 men in the Battle of Kursk, which was the largest tank battle in history. The Soviets continued to advance steadily, taking Belgorod and then Kharkov. In September the Soviet advance was accelerated, and by the end of the month the Germans in the Ukraine had been driven back to the Dnepr.

The Southwest and South Pacific, June–October 1943. A Pacific military conference held in Washington, D.C., in March 1943 produced a new schedule of operations calling for the development of some counterattacks against the Japanese. The reduction of the threat from the large Japanese naval base at Rabaul, by encirclement if not by the capture of that stronghold, was a primary objective for MacArthur.

Between June 22 and June 30, 1943, two U.S. regiments invaded Woodlark and Kiriwina islands (northeast of the tip of Papua), whence aircraft could range over not only the Coral Sea but also the approaches to Rabaul and to the Solomons. At the same time, U.S. and Australian units advanced from Buna along the coast of New Guinea toward Lae and Salamaua, while other Australian forces simultaneously advanced from Wau in the hinterland; and in the night of June 29–30, U.S. forces secured Nassau Bay as a base for further advances against the same positions.

U.S. landings on New Georgia and on Rendova in the Solomons, however, also made in the night of June 29–30, provoked the Japanese into strong counteraction: between July 5 and July 16, in the battles of Kula Gulf and of Kolombangara, the Allies lost one cruiser and two destroyers and had three more cruisers crippled; and the Japanese, though they lost a cruiser and two destroyers, were able to land considerable reinforcements (from New Britain). Only substantial counter-reinforcement secured the New Georgia group of islands for the Allies, who, moreover, began on August 15 to extend their operation to the island of Vella Lavella also. In the last two months of the struggle, which ended with the Japanese evacuation of Vella Lavella on October 7, the Japanese sank an Allied destroyer and crippled two more but lost a further six of their own; and their attempt to defend the Solomon Islands cost them 10,000 lives, as against the Americans' 1,150 killed and 4,100 wounded.

Meanwhile, U.S. planes on August 17–18 had attacked Japanese bases at Wewak (on the New Guinea coast far to the west of Lae) and destroyed more than 200 aircraft there. On September 4 an Australian division landed near Lae, and the next day U.S. paratroops dropped at Nadzab, above Lae on the Markham River, where they were soon joined by an Australian airborne division. Salamaua fell to the Allies on September 12, Lae on September 16, and Finschhafen, on the Huon Peninsula behind Lae, on October 2. On Sept. 30, 1943, the Japanese made a new policy decision: a last defense line was to be established from western New Guinea and the Carolines to the Marianas by spring 1944, to be held at all costs, and also to be used as a base for counterattacks.

Japanese last defense line

The Allied landings in Europe and the defeat of the Axis powers

DEVELOPMENTS FROM AUTUMN 1943 TO SUMMER 1944

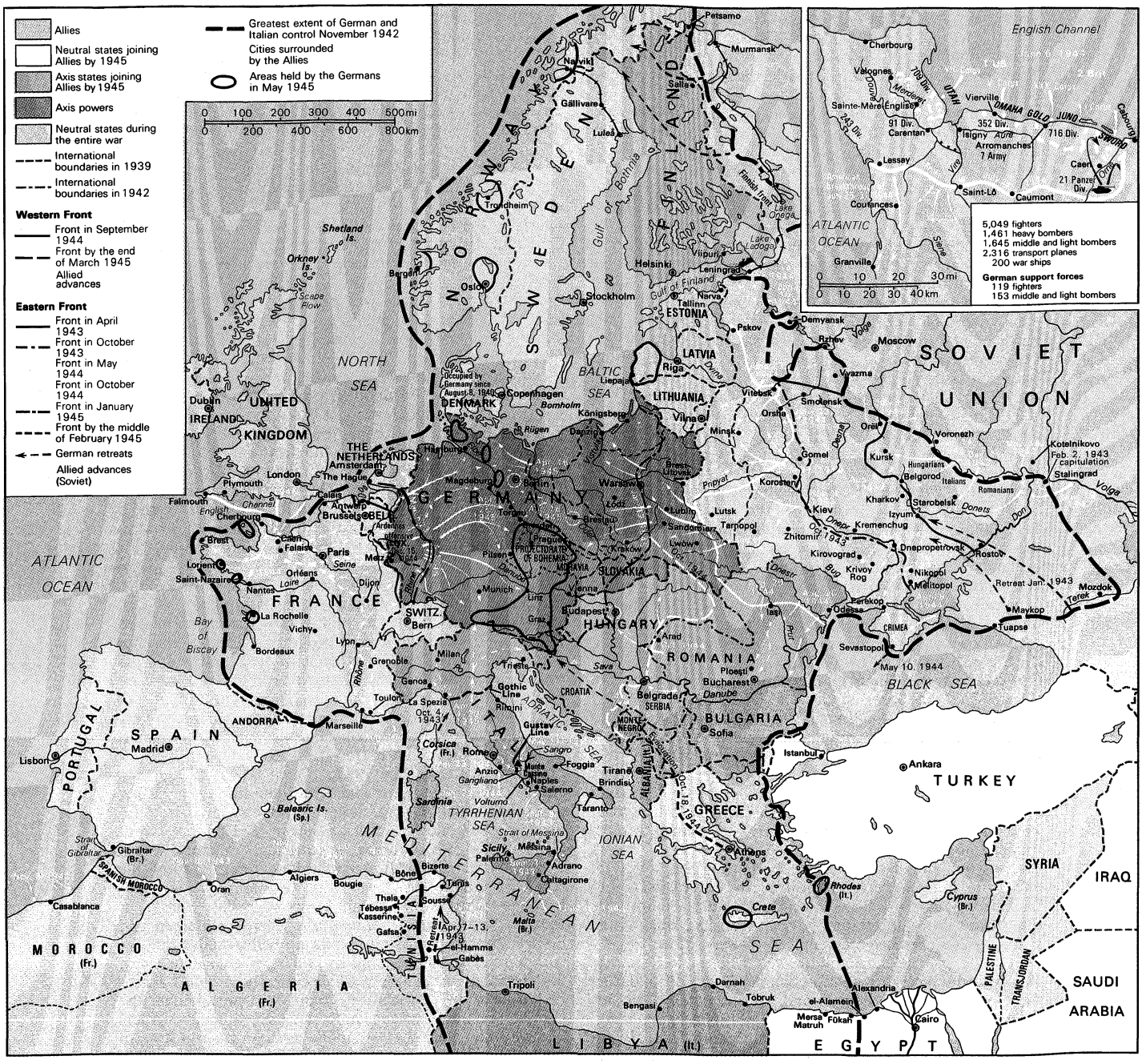
Sicily and the fall of Mussolini, July–August 1943. Hitler's greatest strategic disadvantage in opposing the Allies' imminent reentry into Europe lay in the immense stretch of Germany's conquests; from the west coast of France to the east coast of Greece. It was difficult for him to gauge where the Allies would strike next. The Allies' greatest strategic advantage lay in the wide choice of alternative objectives and in the powers of distraction they enjoyed through their superior sea power. Hitler, while always having to guard against a cross-Channel invasion from England's shores, had cause to fear that the Anglo-American armies in North Africa might land anywhere on his southern front between Spain and Greece.

Having failed to save its forces in Tunisia, the Axis had only 10 Italian divisions of various sorts and two German panzer units stationed on the island of Sicily at midsummer 1943. The Allies, meanwhile, were preparing to throw some 478,000 men into the island—150,000 of them in the first three days of the invasion. Under the supreme command of Alexander, Montgomery's British 8th Army and Patton's U.S. 7th Army were to be landed on two stretches of beach 40 miles long, 20 miles distant from one another, the British in the southeast of the island, the Americans in the south. The Allies' air superiority in the Mediterranean theatre was so great by this time—more than 4,000 aircraft against some 1,500 German and Italian ones—that the Axis bombers had been withdrawn from Sicily in June to bases in north-central Italy.

On July 10 Allied seaborne troops landed on Sicily. The coastal defenses, manned largely by Sicilians unwilling to turn their homeland into a battlefield for the Germans' sake, collapsed rapidly enough. The British forces had cleared the whole southeastern part of the island in the first three days of the invasion. The Allies' drive toward Messina then took the form of a circuitous movement by the British around Mount Etna in combination with an eastward drive by the Americans, who took Palermo, on the western half of the northern coast, on July 22. Meanwhile, the German armoured strength in Sicily had been reinforced.

After the successive disasters sustained by the Axis in Africa, many of the Italian leaders were desperately anxious to make peace with the Allies. The invasion of Sicily, representing an immediate threat to the Italian mainland, prompted them to action. On the night of July 24–25, 1943, when Mussolini revealed to the Fascist Grand Council that the Germans were thinking of evacuating the southern half of Italy, the majority of the council voted for a resolution against him, and he resigned his powers. On

The arrest of Mussolini



German and Allied movements in Europe from the end of 1942 to 1945, and (inset) the Normandy Invasion, June–July 1944.

From *Grosser Historischer Weltatlas*, vol. 3, *Neuzeit*, 2nd ed. (1967); Bayerischer Schulbuch-Verlag, Munich

July 25 the king, Victor Emmanuel III, ordered the arrest of Mussolini and entrusted Marshal Pietro Badoglio with the formation of a new government. The new government entered into secret negotiations with the Allies, despite the presence of sizable German forces in Italy.

A few days after the fall of Mussolini, Field Marshal Albert Kesselring, the German commander in chief in Italy, decided that the Axis troops in Sicily must be evacuated; the local Italian commander thought so too. While rearguard actions held up the Allies at Adrano (on the western face of Mount Etna) and at Randazzo (to the north), 40,000 Germans and 60,000 Italian troops were safely withdrawn across the Strait of Messina to the mainland, mostly in the week ending on Aug. 16, 1943—the day before the Allies' entry into Messina.

The Allies sustained about 22,800 casualties in their conquest of Sicily. The Axis powers suffered about 165,000 casualties, of whom 30,000 were Germans.

The Quadrant Conference (Quebec I). The success of the Sicilian operation and the fall of Mussolini converted the American military and political leadership into supporters of a campaign in Italy. Furthermore, Lieutenant General Sir Frederick Morgan, who after Casablanca had been designated chief of staff to the Supreme Allied Commander (COSSAC), produced a detailed and realistic plan for the long-envisaged invasion of France from Great Britain, thus enabling the U.S. strategists to calculate more precisely how much of the Allies' resources were needed for that purpose and how much could be spared for operations in the Mediterranean and for the Pacific. With regard to the Pacific, plans sponsored by Admiral Nimitz for operations against the Gilbert and Marshall islands apart from the enterprise against Rabaul were approved early in August 1943.

The new turn of strategical thought necessitated a new Anglo-U.S. conference, which took place in Quebec in mid-August 1943 and was code-named "Quadrant." After vigorous debate, the question of the timing of "Overlord" was eventually left open, but it was agreed that the strength of the assault force should exceed the original estimate by

25 percent, that the cross-Channel landing should be supported by a landing in southern France, and that a U.S. officer should be in command of "Overlord." It was also decided that a new Southeast Asia theatre of war should be organized, under British command.

The Allies' invasion of Italy and the Italian volte-face, 1943. From Sicily, the Allies had a wide choice of directions for their next offensive. Calabria, the "toe" of Italy, was the nearest and most obvious possible destination, and the "shin" was also vulnerable; and the "heel" was also very attractive. The two army corps of Montgomery's 8th Army crossed the Strait of Messina and landed on the "toe" of Italy on Sept. 3, 1943; but, though the initial resistance was practically negligible, they made only very slow progress, as the terrain, with only two good roads running up the coasts of the great Calabrian "toe" prevented the deployment of large forces. On the day of the landing, however, the Italian government at last agreed to the Allies' secret terms for a capitulation. It was understood that Italy would be treated with leniency in direct proportion to the part that it would take, as soon as possible, in the war against Germany. The capitulation was announced on September 8.

The landing at Salerno

The landing on the "shin" of Italy, at Salerno, just south of Naples, was begun on September 9, by the mixed U.S.–British 5th Army, under General Mark Clark. Transported by 700 ships, 55,000 men made the initial assault, and 115,000 more followed up. At first they were faced only by the German 16th Panzer Division; but Kesselring, though he had only eight weak divisions to defend all southern and central Italy, had had time to plan since the fall of Mussolini and had been expecting a blow at the "shin." His counterstroke made the success of the Salerno landing precarious for six days, and it was not until October 1 that the 5th Army entered Naples.

By contrast, the much smaller landing on the "heel" of Italy, which had been made on September 2 (the day preceding the invasion of the "toe"), took the Germans by surprise. Notwithstanding the paucity of its strength in men and in equipment, the expedition captured two good ports, Taranto and Brindisi, in a very short time; but it lacked the resources to advance promptly. Nearly a fortnight passed before another small force was landed at Bari, the next considerable port north of Brindisi, to push thence unopposed into Foggia.

It was the threat to their rear from the "heel" of Italy and from Foggia that had induced the Germans to fall back from their positions defending Naples against the 5th Army. When the Italian government, in pursuance of a Badoglio–Eisenhower agreement of September 29, declared war against Germany on Oct. 13, 1943, Kesselring was already receiving reinforcements and consolidating the German hold on central and northern Italy. The 5th Army was checked temporarily on the Volturno River, only 20 miles north of Naples, then more lastingly on the Garigliano River, while the 8th Army, having made its way from Calabria up the Adriatic coast, was likewise held on the Sangro River. Autumn and midwinter passed without the Allies' making any notable impression on the Germans' Gustav Line, which ran for 100 miles from the mouth of the Garigliano through Cassino and over the Apennines to the mouth of the Sangro.

The western Allies and Stalin: Cairo and Tehrān, 1943. Relations between the western Allies and the U.S.S.R. were still delicate. Besides their inability to satisfy Soviet demands for convoys of supplies and for an early invasion of France, the Americans and the British were embarrassed by the discrepancy between their political war aims and Stalin's.

The Polish problem

The longest-standing difference was about Poland. While Poles were still fighting on the Allies' side and acknowledging the authority of General Władysław Sikorski's London-based Polish government in exile, Stalin was trying to get the Allies to consent to the U.S.S.R.'s retention, after the war, of all the territory taken from Poland by virtue of the German–Soviet pacts of 1939. On Jan. 16, 1943, the Soviet government announced that Poles from the border territories in dispute were being treated as Soviet citizens and drafted into the Red Army. On April 25, the Soviet government severed relations with the London Poles, and Moscow subsequently began to build up its own puppet government for postwar Poland.

Besides the quarrel over Poland, the western Allies and the U.S.S.R. were also at variance with regard to the postwar fate of other European states still under German domination; but the Americans and the British were really more interested in maintaining the Soviet war effort against Germany than in insisting, at the risk of offense to Stalin, on the detailed application of their own loudly but vaguely enunciated war aims.

Sextant (Cairo I), Tehrān, and Cairo II

Sextant, the conference of Nov. 22–27, 1943, for which Churchill, Roosevelt, and Chiang Kai-shek met in Cairo, was, on Roosevelt's insistence, devoted mainly to discussing plans for a British–U.S.–Chinese operation in northern Burma. Little was produced by Sextant except the Cairo Declaration, published on December 1, a further statement of war aims. It prescribed inter alia that Japan was to surrender all Pacific islands acquired since 1914, to retrocede Manchuria, Formosa, and the Pescadores to China, and to give up all other territory "taken by violence and greed"; and, in addition, it was stipulated that Korea was in due course to become independent.

From Cairo, Roosevelt and Churchill went to Tehrān, to meet Stalin at the Eureka conference of November 28–December 1. Stalin renewed the Soviet promise of military intervention against Japan, but he primarily wanted an assurance that "Overlord" (the invasion of France) would indeed take place in 1944. Reassured about this by Roosevelt, he declared that the Red Army would attack simultaneously on the Eastern Front. On the political plane, Stalin now demanded the Baltic coast of East Prussia for the U.S.S.R. as well as the territories annexed in 1939–40. The main communique of the conference was accompanied by a joint declaration guaranteeing the postwar restoration of Iran. Returning to Cairo, Roosevelt and Churchill spent six more days, December 2–7, in staff talks to compose their differences on strategy. They finally agreed that "Overlord" (with Eisenhower in command) should have first claim on resources.

German strategy, from 1943. From late 1942 German strategy, every feature of which was determined by Hitler, was solely aimed at protecting the still very large area under German control—most of Europe and part of North Africa—against a future Soviet onslaught on the Eastern Front and against future Anglo-U.S. offensives on the southern and western fronts. The Germans' vague hopes that the Allies would shrink from such costly tasks or that the "unnatural" coalition of western capitalism and Soviet Communism would break up before achieving victory were disappointed; and so Hitler, in accordance with his dictum that "Germany shall either be a world power or not be at all," consciously resolved to preside over the downfall of the German nation. He gave inflexible orders whereby whole armies were made to stand their ground in tactically hopeless positions and were forbidden to surrender under any circumstances. The initial success of this strategy in preventing a German rout during the Soviet winter counteroffensive of 1941–42 had blinded Hitler to its impracticability in the very different military circumstances on the Eastern Front by 1943, by which time the Germans simply lacked sufficient numbers of troops to defend an extremely long front against much more numerous Soviet forces. (By December 1943 the 3,000,000 German troops there were opposed by about 5,500,000 Soviet troops.)

The strategy of keeping his armies stationary was made easier for Hitler by the complete ascendancy he had achieved over his generals, who disputed with Hitler only at the risk of losing their commands or worse. Frequent changes were made in the command of the various army groups and armies, with the result that during 1943–44 most of the talented commanders who had been associated with Germany's past successes were removed, and everyone who was suspected of a critical attitude at headquarters was silenced.

From late 1943 on, Hitler's strategy, which from a political standpoint remains inexplicable to most Western historians, was to strengthen the German forces in west-

ern Europe at the expense of those on the Eastern Front. In view of the danger of the great Anglo-U.S. invasion of western Europe that seemed imminent by early 1944, the loss of some part of his eastern conquests evidently seemed to Hitler to be less serious. Hitler continued to insist on the primacy of the war in the west after the start of the Allied invasion of northern France in June 1944, and while his armies made strenuous efforts to contain the Allied bridgehead in Normandy for the next two months, Hitler accepted the annihilation of the German Army Group Centre on the Eastern Front by the Soviet summer offensive (from June 1944), which brought the Red Army in a few weeks' time to the Vistula River and the borders of East Prussia. But the Western Front likewise crumbled in a few weeks, whereupon the Allies advanced to Germany's western borders. Then, still adhering to his guiding principle, Hitler assembled on the Western Front all that was left of his forces there and tried to drive the British and Americans back in what became known as the Battle of the Bulge. This campaign had some successes but meant that Germany's last battleworthy units were used on the Western Front while the Red Army, heavily outnumbering the remaining German troops in the east, resumed its drive on the eastern frontiers of Germany and reached the Oder River by the end of January 1945.

The Eastern Front, October 1943–April 1944. By the end of the first week of October 1943, the Red Army had established several bridgeheads on the right bank of the Dnepr River. Then, while General N.F. Vatutin's drive against Kiev was engaging the Germans' attention, General Ivan Stepanovich Konev suddenly pushed so far forward from the Kremenchug bridgehead (more than halfway downstream between Kiev and Dnepropetrovsk) that the German forces within the great bend of the Dnepr to the south would have been isolated if Manstein had not stemmed the Soviet advance just in time to extricate them. By early November the Red Army had reached the mouth of the Dnepr also, and the Germans in the Crimea were isolated. Kiev, too, fell to Vatutin on November 6, Zhitomir, 80 miles to the west, and Korosten, north of Zhitomir, in the next 12 days. Farther north, however, the Germans, who had already fallen back from Smolensk to a line covering the upper Dnepr, repelled with little difficulty five rather predictable Soviet thrusts toward Minsk in the last quarter of 1943.

The Soviet advance into Poland

Vatutin's forces from the Zhitomir–Korosten sector advanced westward across the prewar Polish frontier on Jan. 4, 1944; and, though another German flank attack, by troops drawn from adjacent fronts, slowed them down, they had reached Lutsk, 100 miles farther west, a month later. Vatutin's left wing, meanwhile, wheeled southward to converge with Konev's right, so that 10 German divisions were encircled near Korsun, on the Dnepr line south of Kiev. Vainly trying to save those 10 divisions, the Germans had to abandon Nikopol, in the Dnepr bend far to the south, with its valuable manganese mines.

March 1944 saw a triple thrust by the Red Army: Zhukov, succeeding to Vatutin's command, drove southwest toward Tarnopol, to outflank the Germans on the upper stretches of the southern Bug River. General Rodion Yakovlevich Malinovsky, in the south, advanced across the mouth of the latter river from that of the Dnepr; and between them Konev, striking over the central stretch of the Bug, reached the Dnestr, 70 miles ahead, and succeeded in crossing it. When Zhukov had crossed the upper Prut River and Konev was threatening Iaşi on the Moldavian stretch of the river, the Carpathian Mountains were the only natural barrier remaining between the Red Army and the Hungarian Plain. German troops occupied Hungary on March 20, since Hitler suspected that the Hungarian regent, Admiral Miklós Horthy, might not resist the Red Army to the utmost.

The German occupation of Hungary

A German counterstroke from the Lwów area of southern Poland against Zhukov's extended flank early in April not only put an end to the latter's overhasty pressure on the Tatar (Yablonitsky) Pass through the Carpathians but also made possible the withdrawal of some of the German forces endangered by the Red Army's March operation. Konev, too, was halted in front of Iaşi; but his left swung southward down the Dnestr to converge with Malinovsky's drive on Odessa. That great port fell to the Red Army on April 10. On May 9 the Germans in the Crimea abandoned Sevastopol, caught as they were between Soviet pincers from the mainland north of the isthmus and from the east across the Strait of Kerch.

At the northern end of the Eastern Front, a Soviet offensive in January 1944 had been followed by an orderly German retreat from the fringes of the long-besieged Leningrad area to a shorter line exploiting the great lakes farther to the south. The retreat was beneficial to the Germans but sacrificed their land link with the Finns, who now found themselves no better off than they had been in 1939–40. Finland in February 1944 sought an armistice from the U.S.S.R., but the latter's terms proved unacceptable.

The war in the Pacific, October 1943–August 1944. Considering that it might be necessary for them to invade Japan proper, the Allies drew up new plans in mid-1943. The main offensive, it was decided, should be from the south and from the southeast, through the Philippines and through Micronesia (rather than from the Aleutians in the North Pacific or from the Asian mainland). While occupation of the Philippines would disrupt Japanese communications with the East Indian isles west of New Guinea and with Malaya, the conquest of Micronesia, from the Gilberts by way of the Marshalls and Carolines to the Marianas, would not only offer the possibility of drawing the Japanese into a naval showdown but also win bases for heavy air raids on the Japanese mainland prior to invasion.

For the approach to the Philippines, it was prerequisite, on the one hand, to complete the encirclement of Rabaul, thereby nullifying the threat from the Japanese positions in the Solomon Islands and in the Bismarck Archipelago (New Britain, New Ireland, etc.) and, on the other, to reduce the Japanese hold on western New Guinea. Great emphasis, however, was put on the advance across the central Pacific through Micronesia, to be begun via the Gilberts.

The encirclement of Rabaul. Allied moves to isolate the large Japanese garrison on Rabaul proceeded by land and air. The encirclement of Rabaul by land began during October and November 1943 with the capture by New Zealand troops of the Treasury Islands in the Solomons and was accompanied on November 1 by a U.S. landing at Empress Augusta Bay on the west of Bougainville. U.S. reinforcements subsequently repulsed Japanese counterattacks in December, when they sank two destroyers, and in March 1944, when they killed almost 6,000 men. What remained of the Japanese garrison on Bougainville was no longer capable of fighting, though it did not surrender until the end of the war.

Continuing the approach to Rabaul, U.S. troops landed on December 15 at Arawe on the southwestern coast of New Britain, thereby distracting Japanese attention from Cape Gloucester, on the northwestern coast, where a major landing was made on December 26. By Jan. 16, 1944, the airstrip at Cape Gloucester had been captured and defense lines set up. Talasea, halfway to Rabaul, fell in March 1944. The conquest of western New Britain secured Allied control of the Vitiaz and Dampier straits between that island and New Guinea.

By constructing air bases on each island that they captured, the Allies systematically blocked any westward movement that the Japanese might have made: New Zealand troops took the Green Islands southeast of New Guinea on February 15; and U.S. forces invaded Los Negros in the Admiralty Islands on February 29 and captured Manus on March 9.

With the fall of the Emirau Islands on March 20, the Allies' stranglehold on Rabaul and Kavieng was practically complete, so that they could thenceforth disregard the 100,000 Japanese immobilized there.

Western New Guinea. Before they could push northward to the Philippines, the Allies had to subdue Japanese-held western New Guinea. U.S. troops took Saidor, on the Huon Peninsula, on Jan. 2, 1944, and established an air base there; and the Australians took Sio, to the east of

Saidor, on January 16. Then reinforcements were landed at Mindiri, west of Saidor, on March 5, and Australian infantry began to move westward up the coast, to take Bogadjim, Madang, and Alexishafen.

Bypassing Hansa Bay (which was eventually captured on June 15) and Wewak, whither the Japanese had retreated, the Allies, on April 22, 1944, made two simultaneous landings at Hollandia: having in the past weeks already destroyed 300 Japanese planes, they captured the airfields there in four days' time. In the following months Hollandia was converted into a major base and command post for the Southwest Pacific area. The Allies also took Aitape, on the coast east of Hollandia, and held it against counterattacks by more than 200,000 Wewak-based Japanese during July and August. Biak, the isle guarding the entrance to Geelvink Bay, west of Hollandia, was invaded by U.S. troops on May 27, 1944; but the Japanese defense of it was maintained until early August. Though westernmost New Guinea fell likewise to the Allies in August 1944, the Japanese garrison at Wewak held out until May 10, 1945.

Hollandia taken by the Allies

The central Pacific. Though the U.S. Joint Chiefs of Staff envisaged no major offensive westward across the Pacific toward Formosa until mid-1944, they nevertheless decided to launch a limited offensive in the central Pacific in 1943, hoping thereby both to speed the pace of the war and to draw the Japanese away from other areas. Accordingly, Nimitz' central Pacific forces invaded the Gilberts on Nov. 23, 1943. Makin fell easily, but well-fortified Japanese defenses on Tarawa cost the U.S. Marines 1,000 killed and 2,300 wounded. Japanese losses in the Gilberts totaled about 8,500 men.

Having been forced to cede the Gilberts, the Japanese elected next to defend the Marshalls, in order both to absorb Allied forces and to strain the latters' extended lines of supply. Nimitz subjected Kwajalein Atoll, which he chose first to attack, to so heavy a preliminary bombardment that the U.S. infantry could land on it on Jan. 31, 1944; and U.S. forces moved on to Eniwetok on February 17.

In support of the landings on the Marshalls, the U.S. fleet on Feb. 17, 1944, started a series of day and night attacks against the Japanese base at Truk in the Caroline Islands, where they destroyed some 300 aircraft and 200,000 tons of merchant shipping. Henceforth, the Allies could confidently ignore Truk and bypass it.

The Allies' next objective, for which they required more than 500 ships and 125,000 troops, was to reduce the Mariana Islands, lying 1,000 miles from Eniwetok and 3,500 miles from Pearl Harbor. Against this threat, after the destruction at Truk, the Japanese hastily drew up a new defense plan, "Operation A," relying on their remaining 1,055 land-based aircraft in the Marianas, in the Carolines, and in western New Guinea and on timely and decisive intervention by a sea force, which should include nine aircraft carriers with 450 aircraft. But in the spring of 1944 the Japanese air strength was still further depleted, and, moreover, on March 31 the sponsor of the plan, Admiral Koga Mineichi (Yamamoto's successor), and his staff were killed in an air disaster. When, on June 15, two U.S. Marine divisions went ashore on Saipan Island in the Marianas, the 30,000 Japanese defenders put up so fierce a resistance that an army division was needed to reinforce the Marines. Using the same defensive tactics as on other small islands, the Japanese had fortified themselves in underground caves and bunkers that afforded protection from American artillery and naval bombardment. Notwithstanding this, the Japanese defenders were gradually compressed into smaller and smaller pockets, and they themselves ended most organized resistance with a suicidal counterattack on July 7, the largest of its kind during the war.

Saipan taken

The loss of Saipan was such a disaster for Japan that when the news was announced in Tokyo the prime minister, Tōjō Hideki, and his entire Cabinet resigned. To realists in the Japanese high command, the loss of the Marianas spelled the ultimate loss of the war, but no one dared say so. Tōjō's Cabinet was succeeded by that of General Koiso Kuniaki, which was pledged to carrying on the fight with renewed vigour.

Air power enthusiasts have called the conquest of Saipan "the turning point of the war in the Pacific," for it enabled the United States to establish air bases there for the big B-29 bombers, which had been developed for the specific purpose of bombing Japan. The first flight of 100 B-29s took off from Saipan on Nov. 24, 1944, and bombed Tokyo, the first bombing raid on the Japanese capital since 1942.

While the Japanese were still resisting on Saipan, the Japanese Combined Fleet, under Admiral Ozawa Jisaburō, was approaching from Philippine and East Indian anchorages, in accordance with "Operation A," to challenge the U.S. 5th Fleet, under Admiral Raymond Spruance. Ozawa, with only nine aircraft carriers against 15 for the United States, was obviously inferior in naval power, but he counted heavily on help from land-based aircraft on Guam, Rota, and Yap. The encounter, which took place west of the Marianas and is known as the Battle of the Philippine Sea, has been called the greatest carrier battle of the war. It began on June 19 when Ozawa sent 430 planes in four waves against Spruance's ships. The result was a disaster for the Japanese. U.S. airmen shot down more than 300 planes and sank two carriers, and as the Japanese fleet retreated northward toward Okinawa it lost another carrier and almost 100 more planes. The United States lost about 130 planes. The hasty and incomplete training of the Japanese pilots and the inadequate armour plating of their planes were decisive factors in the numerous aerial combats of this battle, which was ultimately of more strategic importance than the fall of Saipan. Nimitz' forces could thereafter occupy other major islands in the Marianas: Guam on July 21 and Tinian on July 24. The Marianas cost the Japanese 46,000 killed or captured, the Americans only 4,750 killed.

The Battle of the Philippine Sea

The Burmese frontier and China, November 1943–summer 1944. For the dry season of 1943–44 both the Japanese and the Allies were resolved on offensives in Southeast Asia. On the Japanese side, Lieutenant General Kawabe Masakazu planned a major Japanese advance across the Chindwin River, on the central front, in order to occupy the plain of Imphāl and to establish a firm defensive line in eastern Assam. The Allies, for their part, planned a number of thrusts into Burma: Stilwell's NCAC forces, including his three Chinese divisions and "Merrill's Marauders" (U.S. troops trained by Wingate on Chindit lines), were to advance against Mogaung and Myitkyina; while Slim's 14th Army was to launch its XV Corps southeastward into Arakan and its IV Corps eastward to the Chindwin. Because the Japanese had habitually got the better of advanced British forces by outflanking them, Slim formulated a new tactic to ensure that his units would stand against attack in the forthcoming campaign, even if they should be isolated: they were to know that, when ordered to stand, they could certainly count both on supplies from the air and on his use of reserve troops to turn the situation against the Japanese attackers.

On the southern wing of the Burmese front, the XV Corps's Arakan operation, launched in November 1943, had achieved most of its objectives by the end of January 1944. When the Japanese counterattack surrounded one Indian division and part of another, Slim's new tactic was brought into play, and the Japanese found themselves crushed between the encircled Indians and the relieving forces.

The Japanese crossing of the Chindwin into Assam, on the central Burmese front, when the fighting in Arakan was dying down, played into Slim's hands, since he could now profit from the Allies' superiority in aircraft and in tanks. The Japanese were able to approach Imphāl and to surround Kohīma, but the British forces protecting these towns were reinforced with several Indian divisions that were taken from the now-secure Arakan front. With air support, Slim's reinforced forces now defended Imphāl against multiple Japanese thrusts and outflanking movements until, in mid-May 1944, he was able to launch two of his divisions into an offensive eastward, while still containing the last bold effort of the Japanese to capture Imphāl. By June 22 the 14th Army had averted the Japanese menace to Assam and won the initiative for its own advance into Burma. The Battle of Imphāl–Kohīma

cost the British and Indian forces 17,587 casualties (12,600 of them sustained at Imphāl), the Japanese forces 30,500 dead (including 8,400 from disease) and 30,000 wounded.

On the northern Burmese front, Stilwell's forces were already approaching Mogaung and Myitkyina before the southern crisis of Imphāl–Kohīma; and the subsidiary Chindit operation against Indaw was going well ahead when, on March 24, 1944, Wingate himself was killed in an air crash. Meanwhile, Chiang Kai-shek was constrained by U.S. threats of a suspension of lend-lease to finally authorize some action by the 12 divisions of his Yunnan Army, which on May 12, 1944, with air support, began to cross the Salween River westward in the direction of Myitkyina, Bhamo, and Lashio. Myitkyina airfield was taken by Stilwell's forces, with "Merrill's Marauders," on May 17, Mogaung was taken by the Chindits on June 26, and finally Myitkyina itself was taken by Stilwell's Chinese divisions on August 3. All of northwest and much of northern Burma was now in Allied hands.

In China proper, a Japanese attack toward Ch'ang-sha, begun on May 27, won control not only of a further stretch of the north–south axis of the Peking–Han-K'ou railroad but also of several of the airfields from which the Americans had been bombing the Japanese in China and were intending to bomb them in Japan.

The Italian front, 1944. The Allies' northward advance up the Italian peninsula to Rome was still blocked by Kesselring's Gustav Line, which was hinged on Monte Cassino. To bypass that line, the Allies landed some 50,000 seaborne troops, with 5,000 vehicles, at Anzio, only 33 miles south of Rome, on Jan. 22, 1944. The landing surprised the Germans and met, at first, with very little opposition; but, instead of driving on over the Alban Hills to Rome at once, the force at Anzio spent so much time consolidating its position there that Kesselring was able, with his reserves, to develop a powerful counteroffensive against it on February 3. The beachhead was thereby reduced to a very shallow dimension, while the defenses at Monte Cassino held out unimpaired against a new assault by Clark's 5th Army.

For a final effort against the Gustav Line, Alexander decided to shift most of the 8th Army, now commanded by Major General Sir Oliver Leese, from the Adriatic flank of the peninsula to the west, where it was to strengthen the 5th Army's pressure around Monte Cassino and on the approaches to the valley of the Liri (headstream of the Garigliano). The combined attack, which was started in the night of May 11–12, 1944, succeeded in breaching the German defenses at a number of points between Cassino and the coast. Thanks to this victory, the Americans could push forward up the coast, while the British entered the valley and outflanked Monte Cassino, which fell to a Polish corps of the 8th Army on May 18. Five days later, the Allies' force at Anzio struck out against the investing Germans (whose strength had been diminished in order to reinforce the Gustav Line); and by May 26 it had achieved a breakthrough. When the 8th Army's Canadian Corps penetrated the last German defenses in the Liri Valley, the whole Gustav Line began to collapse.

Battle of Monte Cassino

Concentrating all available strength on his left wing, Alexander pressed up from the south to effect a junction with the troops thrusting northward from Anzio. The Germans in the Alban Hills could not withstand the massive attack. On June 5, 1944, the Allies entered Rome. The propaganda value of their occupying the Eternal City, Mussolini's former capital, was offset, however, by an unforeseen strategical reality: Kesselring's forces retreated not in the expected rout but gradually, to the line of the Arno River; Florence, 160 miles north of Rome, did not fall to the Allies until August 13; and by that time the Germans had made ready yet another chain of defenses, the Gothic Line, running from the Tyrrhenian coast midway between Pisa and La Spezia, over the Apennines in a reversed S curve, to the Adriatic coast between Pesaro and Rimini.

Alexander might have made more headway against Kesselring's new front if some of his forces had not been subtracted, in August 1944, for the American-sponsored but eventually unnecessary invasion of southern France ("Operation Anvil," finally renamed "Dragoon" [see below]). As it was, the 8th Army, switched back from the west to the Adriatic coast, achieved only an indecisive breakthrough toward Rimini. After this September offensive, the autumn rains set in, to make even more difficult Alexander's indirect movements, against Kesselring's resolute opposition, toward the mouth of the Po River.

Alexander's offensive weakened

DEVELOPMENTS FROM SUMMER 1944 TO AUTUMN 1945

The Allied invasions of western Europe, June–November 1944. The German Army high command had long been expecting an Allied invasion of northern France but had no means of knowing where precisely the stroke would come: while Rundstedt, commander in chief in the west, thought that the landings would be made between Calais and Dieppe (at the narrowest width of the Channel between England and France), Hitler prophetically indicated the central and more westerly stretches of the coast of Normandy as the site of the attack; and Rommel, who was in charge of the forces on France's Channel coast, finally came around to Hitler's opinion. The fortifications of those stretches were consequently improved, but Rundstedt and Rommel still took different views about the way in which the invasion should be met: while Rundstedt recommended a massive counterattack on the invaders after their landing, Rommel, fearing that Allied air supremacy might interfere fatally with the adequate massing of the German forces for such a counterattack, advocated instead immediate action on the beaches against any attempted landing. The Germans had 59 divisions spread over western Europe from the Low Countries to the Atlantic and Mediterranean coasts of France; but approximately half of this number was static, and the remainder included only 10 armoured or motorized divisions.

Postponed from May, the western Allies' "Operation Overlord," their long-debated invasion of northern France, took place on June 6, 1944—the war's most celebrated D-Day—when 156,000 men were landed on the beaches of Normandy between the Orne estuary and the southeastern end of the Cotentin Peninsula: 83,000 British and Canadian troops on the eastern beaches, 73,000 Americans on the western. Under Eisenhower's supreme direction and Montgomery's immediate command, the invading forces initially comprised the Canadian 1st Army (Lieutenant General Henry Duncan Graham Crerar); the British 2nd Army (Lieutenant General Sir Miles Dempsey); and the British 1st and 6th airborne divisions, the U.S. 1st Army, and the U.S. 82nd and 101st airborne divisions (all under Lieutenant General Omar N. Bradley).

Normandy invaded

By 9:00 AM on D-Day the coastal defenses were generally breached, but Caen, which had been scheduled to fall on D-Day and was the hinge of an Allied advance, held out until July 9, the one panzer division already available there on June 6 having been joined the next day by a second. Though the heavy fighting at Caen attracted most of the German reserves, the U.S. forces in the westernmost sector of the front likewise met a very stubborn resistance. But when they had taken the port of Cherbourg on June 26 and proceeded to clear the rest of the Cotentin, they could turn southward to take Saint-Lô on July 18.

The Allies could not have made such rapid progress in northern France if their air forces had not been able to interfere decisively with the movement of the German reserves. Allied aircraft destroyed most of the bridges over the Seine River to the east and over the Loire to the south. The German reserves thus had to make long detours in order to reach the Normandy battle zone and were so constantly harassed on the march by Allied strafing that they suffered endless delays and only arrived in driblets. And even where reserves could have been brought up, their movement was sometimes inhibited by hesitation and dissension on the Germans' own side. Hitler, though he had rightly predicted the zone of the Allies' landings, came to mistakenly believe, after D-Day, that a second and larger invasion was to be attempted east of the Seine and so was reluctant to allow reserves to be moved westward over that river. He also forbade the German forces already engaged in Normandy to retreat in time to make an orderly withdrawal to new defenses.

Rundstedt, meanwhile, was slow in obtaining Hitler's

authority for the movement of the general reserve's SS panzer corps from its position north of Paris to the front; and Rommel, though he made prompt use of the forces at hand, had been absent from his headquarters on D-Day itself, when a forecast of rough weather had seemed to make a cross-Channel invasion unlikely. Subsequently, Rundstedt's urgent plea for permission to retreat provoked Hitler, on July 3, to appoint Kluge as commander in chief in the west in Rundstedt's place; and Rommel was badly hurt on July 17, when his car crashed under attack from Allied planes.

There was something else, besides the progress of the Allies, to demoralize the German commanders—the failure and the aftermath of a conspiracy against Hitler. Alarmed at the calamitous course of events and disgusted by the crimes of the Nazi regime, certain conservative but anti-Nazi civilian dignitaries and military officers had formed themselves into a secret opposition, with Karl Friedrich Goerdeler (a former chief mayor of Leipzig) and Colonel General Ludwig Beck (a former chief of the army general staff) among its leaders. From 1943 this opposition canvased the indispensable support of the active military authorities with some notable success: General Friedrich Olbricht (chief of the General Army Office) and several of the serving commanders, including Rommel and Kluge, became implicated to various extents. Apart from General Henning von Tresckow, however, the group's most dynamic member was Colonel Graf Claus von Stauffenberg, who as chief of staff to the chief of the army reserve from July 1, 1944, had access to Hitler. Finally, it was decided to kill Hitler and to use the army reserve for a coup d'état in Berlin, where a new regime under Beck and Goerdeler should be set up. On July 20, therefore, Stauffenberg left a bomb concealed in a briefcase in the room where Hitler was conferring at his headquarters in East Prussia. The bomb duly exploded; but Hitler survived, and the coup in Berlin miscarried. The Nazi reaction was savage: besides 200 immediately implicated conspirators, 5,000 people who were more remotely linked with the plot or were altogether unconnected with it were put to death. Kluge committed suicide on August 17, Rommel on October 14. Fear permeated and paralyzed the German high command in the weeks that followed.

The plot against Hitler

On July 31, 1944, the Americans on the Allies' right, newly supported by the landing of the U.S. 3rd Army under Patton, broke through the German defenses at Avranches, the gateway from Normandy into Brittany. On August 7 a desperate counterattack by four panzer divisions from Mortain, east of Avranches, failed to seal the breach, and American tanks poured southward through the gap and flooded the open country beyond. Though some of the U.S. forces were then swung southwestward in the hope of seizing the Breton ports in pursuance of the original prescription of "Overlord" and though some went on in more southerly directions toward the crossings of the Loire, others were wheeled eastward—to trap, in the Falaise "pocket," a large part of the German forces retreating southward from the pressure of the Allies' left at Caen. The Americans' wide eastward flanking maneuver after the breakout speedily produced a general collapse of the German position in northern France.

Meanwhile, more and more Allied troops were being landed in Normandy. On August 1, two army groups were constituted: the 21st (comprising the British and Canadian armies) under Montgomery; and the 12th (for the Americans) under Bradley. By the middle of August an eastward wheel wider than that which had cut off the Falaise pocket had brought the Americans to Argentan, southeast of Falaise and level with the British and Canadian advance on the left (north) of the Allies' front, so that a concerted drive eastward could now be launched; and on August 19 a U.S. division successfully crossed the Seine at Mantes-Gassicourt. Already on August 17 the Americans on the Loire had taken Orléans. The clandestine French Resistance in Paris rose against the Germans on August 19; and a French division under General Jacques Leclerc, pressing forward from Normandy, received the surrender of the German forces there and liberated the city on August 25.

The German forces would have had ample time to pull back to the Seine River and to form a strong defensive barrier line there had it not been for Hitler's stubbornly stupid orders that there should be no withdrawal. It was his folly that enabled the Allies to liberate France so quickly. The bulk of the German armoured forces and many infantry divisions were thrown into the Normandy battle and kept there by Hitler's "no withdrawal" orders until they collapsed and a large part of them were trapped. The fragments were incapable of further resistance, and their retreat (which was largely on foot) was soon outstripped by the British and American mechanized columns. More than 200,000 German troops were taken prisoner in France, and 1,200 German tanks had been destroyed in the fighting. When the Allies approached the German border at the beginning of September, after a sweeping drive from Normandy, there was no organized resistance to stop them from driving on into the heart of Germany.

Meanwhile, "Operation Dragoon" (formerly "Anvil") was launched on Aug. 15, 1944, when the U.S. 7th Army and the French 1st Army landed on the French Riviera, where there were only four German divisions to oppose them. While the Americans drove first into the Alps to take Grenoble, the French took Marseille on August 23 and then advanced eastward through France up the Rhône Valley, to be rejoined by the Americans north of Lyon early in September. Both armies then moved swiftly northeastward into Alsace.

Southern France invaded

In the north, however, some discord had arisen among the Allied commanders after the crossing of the Seine. Whereas Montgomery wanted to concentrate on a single thrust northeastward through Belgium into the heavily industrialized Ruhr Valley (an area vital to Germany's war effort), the U.S. generals argued for continuing to advance eastward through France on a broad front, in accordance with the pre-invasion plan. Eisenhower, by way of compromise, decided on August 23 that Montgomery's drive into Belgium should have the prior claim on resources until Antwerp should have been captured but that thereafter the pre-invasion plan should be resumed.

Consequently, Montgomery's 2nd Army began its advance on August 29, entered Brussels on September 3, took Antwerp, with its docks intact, on September 4, and went on, three days later, to force its way across the Albert Canal. The U.S. 1st Army, meanwhile, supporting Montgomery on the right, had taken Namur on the day of the capture of Antwerp and was nearing Aachen. Far to the south, however, Patton's U.S. 3rd Army, having raced forward to take Verdun on August 31, was already beginning to cross the Moselle River near Metz on September 5, with the obvious possibility of achieving a breakthrough into Germany's economically important Saarland. Eisenhower, therefore, could no longer devote a preponderance of supplies to Montgomery at Patton's expense.

Montgomery nevertheless attempted a thrust to cross the Rhine River at Arnhem, the British 1st Airborne Division being dropped ahead there to clear the way for the 2nd Army; but the Germans were just able to check the thrust, thus isolating the parachutists, many of whom were taken prisoner. By this time, indeed, the German defense was rapidly stiffening as the Allies approached the German frontiers: the U.S. 1st Army spent a month grinding down the defenses of Aachen, which fell at last on October 20 (the first city of prewar Germany to be captured by the western Allies); and the 1st Canadian Army, on the left of the British 2nd, did not clear the Schelde estuary west of Antwerp, including Walcheren Island, until early November. Likewise, Patton's 3rd Army was held up before Metz.

The Allies' amazing advance of 350 miles in a few weeks was thus brought to a halt. In early September the U.S. and British forces had had a combined superiority of 20 to 1 in tanks and 25 to 1 in aircraft over the Germans, but by November 1944 the Germans still held both the Ruhr Valley and the Saarland, after having been so near collapse in the west in early September that one or the other of those prizes could have easily been taken by the Allies. The root of the Allied armies' sluggishness in September was that none of their top planners had foreseen such a complete collapse of the Germans as occurred in August

1944. They were therefore not prepared, mentally or materially, to exploit it by a rapid offensive into Germany itself. The Germans thus obtained time to build up their defending forces in the west, with serious consequences both for occupied Europe and the postwar political situation of the Continent.

The Eastern Front, June–December 1944. After a successful offensive against the Finns on the Karelian Isthmus had culminated in the capture of Viipuri (Vyborg) on June 20, 1944, the Red Army on June 23 began a major onslaught on the Germans' front in Belorussia. The attackers' right wing took the bastion town of Vitebsk (Vitebskaya) and then wheeled southward across the highway from Orsha to Minsk; their left wing, under General Konstantin Konstantinovich Rokossovsky, broke through just north of the Pripet Marshes and then drove forward for 150 miles in a week, severing the highway farther to the west, between Minsk and Warsaw. Minsk itself fell to the Red Army on July 3; and, though the Germans extricated a large part of their forces from the Soviet enveloping movement, the Soviet tanks raced ahead, bypassing any attempts to block their path, and were deep into Lithuania and northeastern Poland by mid-July. Then the Soviet forces south of the Pripet Marshes struck too, capturing Lwów and pushing across the San River. This increase of pressure on the Germans enabled Rokossovsky's mobile columns to thrust still farther westward: they reached the Vistula River, and one of them, on July 31, even penetrated the suburbs of Warsaw. The Polish underground in Warsaw thereupon rose in revolt against the Germans and briefly gained control of the city. But three SS armoured divisions arrived to suppress the revolt in Warsaw, and the Soviet Red Army stood idly by across the Vistula while the Germans crushed the insurrection. Although the Soviet halt outside Warsaw was a purposeful move, it is true that the unprecedented length and speed of the Red Army's advance—450 miles in five weeks—had overstrained the Soviet communications. The halt on the Vistula was to last six months.

Romania's change of side

On August 20, however, two Soviet thrusts were launched in another direction—against the German salient in Bessarabia. A new government came to power in Romania on August 23 and not only suspended hostilities against the U.S.S.R. but also, on August 25, declared war against Germany. This long-premeditated volte-face opened the way for three great wheeling movements by the Red Army's left wing through the vast spaces of southeastern and central Europe: southwestward across Bulgaria, where they met no opposition; westward up the Danube Valley and over the Yugoslav frontier; and northwestward through the Carpathians into Transylvania. The Germans could only try to hold the threatened centres of communication long enough for the withdrawal of their forces from Greece and from southern Yugoslavia. Belgrade fell to a concerted action by the Red Army and Tito's Partisan forces on Oct. 20, 1944; and a rapid drive from the Transylvanian sector into the Hungarian Plain brought Soviet forces up to the suburbs of Budapest on November 4. Budapest, however, was stubbornly defended: by the end of the year, it was enveloped but still holding out.

At the northern end of the Eastern Front, Finland had capitulated early in September, and the following weeks saw a series of scythelike strokes by the Red Army against the German forces remaining in Estonia, Latvia, and Lithuania. By mid-October the remnants of those forces were cornered in Courland, but the subsequent Soviet attempt to break through from Lithuania into East Prussia was repelled.

Air warfare, 1944. The Allies' strategic air offensive against Germany began to attain its maximum effectiveness in the opening months of 1944. Both the U.S. air forces concerned, namely, the 8th in England and the 15th in Italy, were increased in numbers and improved in technical proficiency. By the end of 1943 the 8th Bomber Command alone could mount attacks of 700 planes, and early in 1944 regular 1,000-bomber attacks became possible. Even more important was the arrival in Europe of effective long-range fighters, chief of which, the P-51 Mustang, was capable of operating at maximum bomber range. The U.S. fighters could now get the better of the Luftwaffe in the air over Germany, so that whereas 9.1 percent of bombers going out had been lost and 45.6 percent damaged in October 1943, the corresponding figures were only 3.5 percent and 29.9 percent in February 1944, though in that very month a massive and very difficult attack at extreme range had been made on the German aircraft industry. Carl Spaatz, commanding general of the U.S. Strategic Air Forces in Europe, in May 1944 initiated an offensive against Germany's synthetic-oil production—an offensive that was to become more and more harmful to the German war effort after the loss of Romania's oil fields to the Soviet Union. Meanwhile, the Luftwaffe's resistance dwindled almost to nothing as its fighter plane production dropped and most of its remaining trained pilots died in aerial combat.

The RAF Bomber Command launched nearly 10,000 sorties in March 1944 and dropped some 27,500 tons of bombs, about 70 percent of this effort being concentrated on Germany; but in the following months its offensive was largely diverted to the intensive preparation and, later, to the support of the Allied landings in France. Nevertheless, it joined usefully in the U.S. offensive against German oil production, continued to play its part in the Battle of the Atlantic, and also assumed the task of bombing the launching ramps of the Germans' "V" missiles. By early 1945, the unending Allied bombing and strafing raids on bridges, roads, rail facilities, locomotives, and supply columns had paralyzed the German transportation system.

Germany's "V" weapons

The "V" missiles, flying bombs and long-range rockets, were the new weapons on which Hitler had vainly been counting to reduce Great Britain to readiness for peace. His faith in them had indeed been a major motive for his insistence on holding the sites, in northernmost France, from which they were initially to be aimed at London. The V1 missiles were first launched on June 13, 1944, mostly from sites in the Pas-de-Calais; the V2 missiles were launched a few months later, on September 8, from sites in The Netherlands (after the Allies' occupation of the Pas-de-Calais on their way to Belgium). The V2 offensive was maintained until March 1945.

Allied policy and strategy: Octagon (Quebec II) and Moscow, 1944. The progress of the Soviet armies toward central and southeastern Europe made it all the more urgent for the western Allies to come to terms with Stalin about the fate of the "liberated" countries of eastern Europe. London had already proposed to Moscow in May 1944 that Romania and Bulgaria should be zones for Soviet military operation, Yugoslavia and Greece—whose royalist governments in exile were under British protection—for British; and Roosevelt had approved this proposition in June.

The Soviet Union had in February 1944 sent a military mission to Tito's Communist Partisans in Yugoslavia (the Partisans had become the sole Yugoslavian recipients, since the Tehrān Conference, of western aid, though their royalist rivals, the Chetniks, were not publicly disavowed by Churchill until May 25). Along with this, a would-be government of Greece had been set up in March by the EAM (National Liberation Front), which was a Communist movement controlling a military organization, the ELAS (National Popular Liberation Army), in opposition to the EDES (Greek Democratic National Army), which was loyal to the British-backed government in exile. The Polish question, moreover, was still unresolved, and in July the Soviets established, at Lublin, a Committee of National Liberation independent of the London Poles. In Romania, despite the government's change of side in August, the Soviets proceeded to disband the Romanian Army; and early in September they declared war on Bulgaria, invaded that country, and sponsored a Communist revolution there.

"Octagon" and the Moscow Conference

With this background, Churchill and Roosevelt met again for their second Quebec Conference, code-named "Octagon," which lasted from September 11 to 16. The most important decision made at the conference was that Roosevelt and Churchill together approved the European Advisory Commission's scheme for the division of defeated Germany into U.S., British, and Soviet zones of

occupation (the southwest, the northwest, and the east, respectively) and also the radical plan elaborated by the U.S. secretary of the treasury, Henry Morgenthau, Jr., for turning Germany "into a country primarily agricultural and pastoral" without "war-making industries." The Morgenthau Plan, however, was subsequently revoked.

The next conference of the Allies was held in Moscow Oct. 9–20, 1944, between Churchill and Stalin, with U.S. ambassador W. Averell Harriman also present at most of their talks. Disagreement persisted over Poland. Stalin, however, consented readily to Churchill's provisional suggestion for zones of influence in southeastern Europe: the U.S.S.R. should be preponderant in Romania and in Bulgaria, the western powers in Greece, and western and Soviet influences should counterbalance one another evenly in Yugoslavia and in Hungary. The timing of the next western and Soviet offensives against Germany was also agreed, and some accord was reached about the scale of the eventual Soviet participation in the war against Japan.

The Philippines and Borneo, from September 1944. On July 27–28, 1944, Roosevelt had approved MacArthur's argument that the next objective in the Pacific theatre of the war should be the Philippine Archipelago (which was comparatively near to the already conquered New Guinea). The initial steps toward the Philippines were taken almost simultaneously, in mid-September 1944: MacArthur's forces from New Guinea seized Morotai, the northeasternmost isle of the Moluccas, which was on the direct route to Mindanao, southernmost landmass of the Philippines; and Nimitz' fleet from the east landed troops in the Palau Islands.

Already by mid-September the Americans had discovered that the Japanese forces were unexpectedly weak not only on Mindanao but also on Leyte, the smaller island north of the Surigao Strait. With this knowledge they decided to bypass Mindanao and to begin their invasion of the Philippines on Leyte. On Oct. 17–18, 1944, American forces seized offshore islets in Leyte Gulf, and on October 20 they landed four divisions on the east coast of Leyte.

The threat to Leyte was the signal for the Japanese to put into effect their recently formulated plan "Sho-Go" ("Operation Victory"), whereby the Allies' next attempts at invasion were to be countered by concerted air attacks. Though in the case of Leyte the Japanese Army and Navy air forces in the immediate theatre numbered only 212 planes, it was hoped that the dispatch of four carriers under Vice Admiral Ozawa, with 106 planes, southward from Japanese waters would lure the U.S. aircraft carriers away from Leyte Gulf and that the suicidal "kamikaze" tactics of the Japanese airmen would save the situation. (Kamikaze pilots deliberately crashed their bomb-armed planes into enemy ships.) At the same time, however, a Japanese naval force from Singapore was to sail to Brunei Bay and there split itself into two groups that would converge on Leyte Gulf from the north and from the southwest: the stronger group, under Vice Admiral Kurita Takeo, would enter the Pacific through the San Bernardino Strait between the Philippine islands of Samar and Luzon; the other, under Vice Admiral Nishimura Teiji, would pass through the Surigao Strait.

The Battle of Leyte Gulf

Kurita's fleet (five battleships, 12 cruisers, 15 destroyers) lost two of its heavy cruisers to U.S. submarine attack on October 23, when it was off Palawan; and one of the mightiest of Japan's battleships, the *Musashi,* was sunk by aerial attack the next day. On October 25, however, Kurita made his way unopposed through the San Bernardino Strait, since the commander of the U.S. 3rd Fleet, Admiral Halsey, had diverted his main strength toward the bait dangled by Ozawa farther to the north. Three groups of U.S. escort carriers, met by Kurita on his way toward Leyte Gulf, suffered heavy damage; but, meanwhile, Nishimura's fleet (two battleships, one heavy cruiser, four destroyers) had been detected on its way to the Surigao Strait and, on its entry into Leyte Gulf in the early hours of October 25, had been practically annihilated by the U.S. 7th Fleet. Kurita consequently turned back from his rendezvous in Leyte Gulf; and the Japanese defeat in the war's greatest naval confrontation was sealed by Ozawa's losses to Halsey: all of his four carriers, together with a light cruiser and two destroyers. The Japanese Navy's "Sho-Go" as it transpired in the Battle of Leyte Gulf had not only failed to inflict serious damage on the Americans but had resulted in serious losses for the Japanese. These losses amounted to three battleships, one large aircraft carrier, three light carriers, six heavy cruisers, four light cruisers, and 11 destroyers, while the United States lost only one light carrier, two escort carriers, and three destroyers. The battle reduced the Japanese Navy to vestigial strength and cleared the way for the U.S. occupation of the Philippines.

Defeat in the gulf, however, did not prevent the Japanese from landing reinforcements on the west coast of Leyte. They put up so stubborn a resistance that the Americans themselves had to be reinforced before Ormoc fell on Dec. 10, 1944; it was not before December 25 that the Americans could claim control of all Leyte—though there was still some mopping up to be done. Altogether, the defense of Leyte cost the Japanese some 75,000 combatants killed or taken prisoner.

From Leyte the Americans proceeded first, on December 15, to the invasion of Mindoro, the largest of the islands immediately south of Luzon. Kamikaze counterattacks made this conquest more costly; and they were to be continued after the Americans had surprised the Japanese by landing, on Jan. 9, 1945, at Lingayen Gulf on the west coast of Luzon itself, the most important island of the Philippines. The local Japanese commander, Lieutenant General Yamashita Tomoyuki, with no hope of reinforcement, opted for tying the enemy forces down as long as possible by a static defense in three mountainous sectors—west, northwest, and east of the Central Plains behind Manila.

The retaking of Manila

Manila itself was also strongly defended by the Japanese. One U.S. corps, however, was approaching it from Lingayen over the Central Plains; a second corps was landed at Subic Bay, at the northern end of the Bataan Peninsula, on Jan. 29, 1945, to make contact with the former corps at Dinalupihan a week later; and troops made an amphibious landing at Nasugbu, south of Manila Bay, on January 31. Manila was then invested, and during the siege the bay was cleared by the occupation of the southern tip of Bataan Peninsula on February 15 and by the reduction of Corregidor Island in the following fortnight. On March 3 Manila fell at last to the Americans.

The Japanese resistance on Luzon continued in the mountains, and east of Manila it went on until mid-June 1945. Mindanao, meanwhile, was likewise being reduced. A U.S. division landed at Zamboanga, on the southwestern peninsula, on March 10, 1945, and a corps began the occupation of the core of the island on April 17.

The last phase of the U.S. campaign in the Philippines coincided with the opening of the reconquest of Borneo from the Japanese, chiefly by Australian forces. Tarakan Island, off the northeast coast, was invaded on May 1; Brunei on the northwest coast was invaded on June 10; and Balikpapan, on the east coast far to the south of Tarakan, was attacked on July 1. The subsequent collapse of the Japanese defenses around Balikpapan deprived Japan of the oil supplies of southern Borneo.

Burma and China, October 1944–May 1945. Chiang Kai-shek's demand for the recall of the talented but abrasive Stilwell was satisfied in October 1944, and some reorganization of the Allies' commands in Southeast Asia followed. While Lieutenant General Daniel Sultan took Stilwell's place, Major General A.C. Wedemeyer became commander of U.S. forces in the China theatre and Sir Oliver Leese commander of the land forces under Mountbatten.

On the northern wing of the Burma front, a three-pronged drive by NCAC forces southward from Myitkyina to the Irrawaddy River had been planned by Stilwell. Launched under Sultan, the triple drive was at first only partially successful: the right took Indaw and Katha early in December and effected a junction with Slim's British 14th Army, and the centre reached Shwegu, across the river; but the left, though it took Bhamo, was checked 60 miles west of Wan-t'ing. Sultan thereupon decided to push farther southward, both on the right against Kyaukme,

The Burma Road reopened

on the Burma Road northeast of Mandalay, and on the left against Wan-t'ing. Threatened with envelopment, the Japanese fell back from Wan-t'ing, which Sultan's troops promptly occupied. Convoys up the Burma Road from Wan-t'ing to K'un-ming were resumed on Jan. 18, 1945.

For central Burma, meanwhile, Slim had thought, after his victory at Imphāl, that he must immediately seize the crossings of the Chindwin River at Sittaung and at Kalewa and then advance southward against Mandalay itself. He did indeed effect the Chindwin crossings, but in mid-December 1944 he saw that the Japanese were in any case going to withdraw altogether to the left bank of the Irrawaddy. Thereupon, he changed his plan: his objective should rather be Meiktila, which lay east of the Irrawaddy and was a vital centre of Japanese communications between Mandalay and Rangoon to the south. To conceal his new intention, he allowed one of the corps already directed against Mandalay to continue its eastward advance, but the other corps was surreptitiously moved over a circuitous route of 300 miles southward to Pakokku, which lay south of the Chindwin–Irrawaddy confluence and northwest of Meiktila. While the crossing of the Irrawaddy by the former corps on both sides of Mandalay distracted the attention of the Japanese, the latter corps took Meiktila on March 3, 1945, and held it against fierce counterattacks. Mandalay fell 10 days later, and the whole area was under the 14th Army's control by the end of the month. When the action was over, two Japanese armies had lost one-third of their fighting strength.

It remained for Slim to capture the Burmese capital, Rangoon. Allied ground forces advanced on Rangoon along two routes from the north: one corps, having moved down the Sittang Valley east of the Irrawaddy, took Pegu; the other, moving down the river, took Prome (Pye). The monsoon, however, was imminent, and to forestall it a small combined operation was undertaken: parachute troops were dropped at Elephant Point, on the coast south of Rangoon, on May 1, 1945; and an Indian division, landing at Rangoon itself the next day, took the city without opposition, just when the monsoon rains were beginning to fall. The recapture of Burma was essentially complete with the taking of Rangoon.

Rangoon liberated

The German offensive in the west, winter 1944–45. Hitler still hoped to drive the Allies back and still adhered to his principle of concentrating on the war in the west. Late in 1944, therefore, he assembled on the Western Front all the manpower that had become available as a consequence of his second "total mobilization": a decree of October 18 had raised a *Volkssturm,* or "home guard," for the defense of the Third Reich, conscripting all able-bodied men between the ages of 16 and 60 years.

In mid-November all six Allied armies on the Western Front had launched a general offensive; but, though the French 1st Army and the U.S. 7th had reached the Rhine River in Alsace, there were only small gains on other sectors of the front. Meanwhile, the German defense was being continuously strengthened with hastily shifted reserves and with freshly raised forces, besides the troops that had managed to make their way back from France. The German buildup along the front was by now progressing faster than that of the Allies, despite Germany's great inferiority of material resources. In mid-December 1944 the Germans gave the Allied armies a shock by launching a sizable counteroffensive. The Germans amassed 24 divisions for the attack. Under the overall command of the reinstated Rundstedt, this attack was to be delivered through the wooded hill country of the Ardennes against the weakest sector of the U.S.-manned front, between Monschau (southwest of Aachen) and Echternach (northwest of Trier). While the 5th Panzer Army on the left, under the talented commander General Hasso von Manteuffel, with its own left flank covered by the German 7th Army, was to wheel northwestward after the breakthrough and to cross the Meuse River of Namur in a drive on Brussels, the 6th Panzer Army on the right, under SS General Sepp Dietrich, was to wheel more sharply northward against the Allies' important supply port of Antwerp. Thus, it was hoped, the British and Canadian forces at the northern end of the front could be cut off from their supplies and crushed, while the U.S. forces to the south were held off by the German left.

The offensive was prepared with skill and secrecy and was launched on Dec. 16, 1944, at a time when mist and rain would minimize the effectiveness of counteraction from the air. The leading wedge of the attack by eight German armoured divisions along a 75-mile front took the Allies by surprise; and the 5th Panzer Army, which achieved the deeper penetration, reached points within 20 miles of the crossings of the Meuse River at Givet and at Dinant. U.S. detachments, however, stood firm, albeit outflanked, at Bastogne and at other bottlenecks in the Ardennes; and there followed what is popularly remembered as the Battle of the Bulge. By December 24 the German drive had narrowed but deepened, having penetrated about 65 miles into the Allied lines along a 20-mile front. But by this time the Allies had begun to respond. Montgomery, who had taken charge of the situation in the north, swung his reserves southward to forestall the Germans on the Meuse. Bradley, commanding the Allied forces south of the German wedge, sent his 3rd Army under Patton to the relief of Bastogne, which was accomplished on December 26. The weather cleared, and as many as 5,000 Allied aircraft began to bomb and strafe the German forces and their supply system. During Jan. 8–16, 1945, the German attackers were compelled to withdraw, lest the salient that they had driven into the Allied front be cut off in its turn. Though their abortive offensive inflicted much damage and upset the Allies' plans, the Germans spent too much of their strength on it and thereby forfeited whatever chance they had had of maintaining prolonged resistance later. The Germans sustained 120,000 casualties and the Americans sustained about 75,000 in the Battle of the Bulge.

The Battle of the Bulge

The Soviet advance to the Oder, January–February 1945. At the end of 1944 the Germans still held the western half of Poland, and their front was still 200 miles east of where it had been at the start of the war in 1939. The Germans had checked the Soviets' summer offensive and had established a firm line along the Narew and Vistula rivers southward to the Carpathians, and in October they repelled the Red Army's attempted thrust into East Prussia. Meanwhile, however, the Soviet left, moving up from the eastern Balkans, had been gradually pushing around through Hungary and Yugoslavia in a vast flanking movement; and the absorption of German forces in opposing this side-door approach detracted considerably from the Germans' capacity to maintain their main Eastern and Western fronts.

The Soviet high command was now ready to exploit the fundamental weaknesses of the German situation. Abundant supplies for their armies had been accumulated at the railheads. The mounting stream of American-supplied trucks had by this time enabled the Soviets to motorize a much larger proportion of their infantry brigades and thus, with the increasing production of their own tanks, to multiply the number of armoured and mobile corps for a successful breakthrough.

Before the end of December ominous reports were received by Guderian—who, in this desperately late period of the war, had been made chief of the German general staff. German Army intelligence reported that 225 Soviet infantry divisions and 22 armoured corps had been identified on the front between the Baltic and the Carpathians, assembled to attack. But when Guderian presented the report of these massive Soviet offensive preparations, Hitler refused to believe it, exclaiming: "It's the biggest imposture since Genghis Khan! Who is responsible for producing all this rubbish?"

If Hitler had been willing to stop the Ardennes counteroffensive in the west, troops could have been transferred to the Eastern Front; but he refused to do so. At the same time he refused Guderian's renewed request that the 30 German divisions now isolated in Courland (on the Baltic seacoast in Lithuania) should be evacuated by sea and brought back to reinforce the gateways into Germany. As a consequence, Guderian was left with a mobile reserve of only 12 armoured divisions to back up the 50 weak infantry divisions stretched out over the 700 miles of the main front.

The Soviet offensive opened on Jan. 12, 1945, when Konev's armies were launched against the German front in southern Poland, starting from their bridgehead over the Vistula River near Sandomierz. After it had pierced the German defense and produced a flanking menace to the central sector, Zhukov's armies in the centre of the front bounded forward from their bridgeheads nearer Warsaw. That same day, January 14, Rokossovsky's armies also joined in the offensive, striking from the Narew River north of Warsaw and breaking through the defenses covering this flank approach to East Prussia. The breach in the German front was now 200 miles wide.

On Jan. 17, 1945, Warsaw was captured by Zhukov, after it had been surrounded; and on January 19 his armoured spearheads drove into Łódź. That same day Konev's spearheads reached the Silesian frontier of prewar Germany. Thus, at the end of the first week the offensive had been carried 100 miles deep and was 400 miles wide—far too wide to be filled by such scanty reinforcements as were belatedly provided.

The crisis made Hitler renounce any idea of pursuing his offensive in the west; but, despite Guderian's advice, he switched the 6th Panzer Army not to Poland but to Hungary in an attempt to relieve Budapest. The Soviets could thus continue their advance through Poland for two more weeks. While Konev's spearheads crossed the Oder River in the vicinity of Breslau (Wrocław) and thus cut Silesia's important mineral resources off from Germany, Zhukov made a sweeping advance in the centre by driving forward from Warsaw, past Poznań, Bydgoszcz, and Toruń, to the frontiers of Brandenburg and of Pomerania. At the same time Rokossovsky pushed on, through Allenstein (Olsztyn), to the Gulf of Danzig, thus cutting off the 25 German divisions in East Prussia. To defend the yawning gap in the centre of the front, Hitler created a new army group and put Heinrich Himmler in command of it with a staff of favoured SS officers. Their fumbling helped to clear the path for Zhukov, whose mechanized forces by Jan. 31, 1945, were at Küstrin, on the lower Oder, only 40 miles from Berlin.

The Soviet drive into Germany

Zhukov's advance now came to a halt. Konev, however, could still make a northwesterly sweep down the left bank of the middle Oder, reaching Sommerfeld, 80 miles from Berlin, on February 13, and the Neisse River two days later. The Germans' defense benefited from being driven back to the straight and shortened line formed by the Oder and Neisse rivers. This front, extending from the Baltic coast to the Bohemian frontier, was less than 200 miles long. The menace of the Soviets' imminent approach to Berlin led Hitler to decide that most of his fresh drafts of troops must be sent to reinforce the Oder; the way was thus eased for the crossing of the Rhine River by the American and British armies.

On Feb. 13, 1945, the Soviets took Budapest, the defense of which had entailed the Germans' loss of Silesia.

Yalta. Roosevelt's last meeting with Stalin and Churchill took place at Yalta, in the Crimea, Feb. 4–11, 1945. The conference is chiefly remembered for its treatment of the Polish problem: the western Allied leaders, abandoning their support of the Polish government in London, agreed that the Lublin committee—already recognized as the provisional government of Poland by the Soviet masters of the country—should be the nucleus of a provisional government of national unity, pending free elections. But while they also agreed that Poland should be compensated in the west for the eastern territories that the U.S.S.R. had seized in 1939, they declined to approve the Oder–Neisse line as a frontier between Poland and Germany, considering that it would put too many Germans under Polish rule. For the rest of "liberated Europe" the western Allied leaders obtained nothing more substantial from Stalin than a declaration prescribing support for "democratic elements" and "free elections" to produce "governments responsive to the will of the people."

For Germany the conference affirmed the project for dividing the country into occupation zones, with the difference that the U.S. zone was to be reduced in order to provide a fourth zone, for the French to occupy. Roosevelt and Churchill, however, had already discarded the Morgenthau Plan for the postwar treatment of Germany; and Yalta found no comprehensive formula to replace it. The three leaders simply pledged themselves to furnish the defeated Germans with the necessities for survival; to "eliminate or control" all German industry that could be used for armaments; to bring major war criminals to trial; and to set up a commission in Moscow for the purpose of determining what reparation Germany should pay.

The German collapse, spring 1945. Before their ground forces were ready for the final assault on Germany, the western Allies intensified their aerial bombardment. This offensive culminated in a series of five attacks on Dresden, launched by the RAF with 800 aircraft in the night of Feb. 13–14, 1945, and continued by the U.S. 8th Air Force with 400 aircraft in daylight on February 14, with 200 on February 15, with 400 again on March 2, and, finally, with 572 on April 17. The motive of these raids was allegedly to promote the Soviet advance by destroying a centre of communications important to the German defense of the Eastern Front; but, in fact, the raids achieved nothing to help the Red Army militarily and succeeded in obliterating the greater part of one of the most beautiful cities of Europe and in killing at least 35,000 people and perhaps 135,000.

The destruction of Dresden

The main strength of the ground forces being built up meanwhile for the crossing of the Rhine was allotted to Montgomery's armies on the northern sector of the front. Meanwhile, some of the U.S. generals sought to demonstrate the abilities of their own less generously supplied forces. Thus, Patton's 3rd Army reached the Rhine at Coblenz (Koblenz) early in March, and, farther downstream, General Courtney H. Hodges' 1st Army seized the bridge over the Rhine at Remagen south of Bonn and actually crossed the river, while, still farther downstream, Lieutenant General William H. Simpson's 9th Army reached the Rhine near Düsseldorf. All three armies were ordered to mark time until Montgomery's grand assault was ready; but, meanwhile, they cleared the west bank of the river, and eventually, in the night of March 22–23, the 3rd Army crossed the Rhine at Oppenheim, between Mainz and Mannheim, almost unopposed.

At last, in the night of March 23–24, Montgomery's attack by 25 divisions was launched across a stretch—30 miles long—of the Rhine near Wesel after a stupendous bombardment by more than 3,000 guns and waves of attacks by bombers. Resistance was generally slight; but Montgomery would not sanction a further advance until his bridgeheads were consolidated into a salient 20 miles deep. Then the Canadian 1st Army, on the left, drove ahead through The Netherlands, the British 2nd went northeastward to Lübeck and to Wismar on the Baltic, and the U.S. armies swept forward across Germany, fanning out to reach an arc that stretched from Magdeburg (9th Army) through Leipzig (1st) to the borders of Czechoslovakia (3rd) and of Austria (7th and French 1st).

Guderian had tried to shift Germany's forces eastward to hold the Red Army off; but Hitler, despite his anxiety for Berlin, still wished to commit the 11th and 12th armies—formed from his last reserves—to driving the western Allies back over the Rhine and, on March 28, replaced Guderian with General Hans Krebs as chief of the general staff.

The dominant desire of the Germans now, both troops and civilians, was to see the British and American armies sweep eastward as rapidly as possible to reach Berlin and occupy as much of the country as possible before the Soviets overcame the Oder line. Few of them were inclined to assist Hitler's purpose of obstruction by self-destruction. On March 19 (the eve of the Rhine crossing), Hitler had issued an order declaring that "the battle should be conducted without consideration for our own population." His regional commissioners were instructed to destroy "all industrial plants, all the main electricity works, waterworks, gas works" together with "all food and clothing stores" in order to create "a desert" in the Allies' path. When his minister of war production, Albert Speer, protested against this drastic order, Hitler retorted: "If the war is lost, the German nation will also perish. So there is no need to consider what the people require for continued existence." Appalled at such callousness, Speer was shaken

out of his loyalty to Hitler: he went behind Hitler's back to the army and industrial chiefs and persuaded them, without much difficulty, to evade executing Hitler's decree. The Americans and the British, driving eastward from the Rhine, met little opposition and reached the Elbe River 60 miles from Berlin, on April 11. There they halted.

On the Eastern Front, Zhukov enlarged his bridgehead across the Oder early in March. On their far left the Soviets reached Vienna on April 6; and on the right they took Königsberg on April 9. Then, on April 16, Zhukov resumed the offensive in conjunction with Konev, who forced the crossings of the Neisse; this time the Soviets burst out of their bridgeheads, and within a week they were driving into the suburbs of Berlin. Hitler chose to stay in his threatened capital, counting on some miracle to bring salvation and clutching at such straws as the news of the death of Roosevelt on April 12. By April 25 the armies of Zhukov and Konev had completely encircled Berlin, and on the same day they linked up with the Americans on the Elbe River.

Isolated and reduced to despair, Hitler married his mistress, Eva Braun, during the night of April 28–29, and on April 30 he committed suicide with her in the ruins of the Chancellery, as the advancing Soviet troops were less than a half mile from his bunker complex; their bodies were hurriedly cremated in the garden. The "strategy" of Hitler's successor, Dönitz, was one of capitulation and of saving as many as possible of the westward-fleeing civilians and of his German troops from Soviet hands. During the interval of surrender, 1,800,000 German troops (55 percent of the Army of the East) were transferred into the British–U.S. area of control.

Hitler's suicide and the surrender of the Germans

On the Italian front, the Allied armies had long been frustrated by the depletion of their forces for the sake of other enterprises; but early in 1945 four German divisions were transferred from Kesselring's command to the Western Front, and in April the thin German defenses in Italy were broken by an Allied attack. A surrender document that had been signed on April 29 (while Hitler was still alive) finally brought the fighting to a conclusion on May 2.

The surrender of the German forces in northwestern Europe was signed at Montgomery's headquarters on Lüneburg Heath on May 4; and a further document, covering all the German forces, was signed with more ceremony at Eisenhower's headquarters at Reims, in the presence of Soviet as well as U.S., British, and French delegations. At midnight on May 8, 1945, the war in Europe was officially over.

Potsdam. The last inter-Allied conference of World War II, code-named "Terminal," was held at the suburb of Potsdam, outside ruined Berlin, from July 17 to Aug. 2, 1945. It was attended by the Soviet, U.S., and British heads of government and foreign ministers: respectively, Stalin and Molotov; President Harry S. Truman (Roosevelt's successor) and James F. Byrnes; and Churchill and Eden, the last-named pair being replaced by Clement Attlee and Ernest Bevin after Great Britain's change of government following a general election.

Operations against Japan were discussed, and the successful testing of an atomic bomb in the United States was divulged to Stalin. Pending the Soviet entry into the war against Japan, a declaration was issued on July 26 calling on Japan to surrender unconditionally and forecasting the territorial spoliation of the empire and the military occupation of Japan proper as well as the prosecution of war criminals, yet still promising that the Japanese people would not be enslaved or the nation destroyed.

Time was spent discussing the peace settlement and its procedure. Stalin induced Truman and Attlee to consent provisionally to the Soviet Union's demands that it should take one-third of Germany's naval and merchant fleet; have the right to exact reparations from its occupied zones of Germany and of Austria and also from Finland, Hungary, Romania, and even Bulgaria; and should furthermore receive a percentage of reparation from the western-occupied zones. The total amounts of all these exactions were, however, to be determined at a later date.

There was a profound disagreement at the conference about the Balkan areas occupied by the Red Army in which representatives of the western powers were allowed little say, and about the area east of the Oder–Neisse line, all of which the Soviets had arbitrarily put under Polish administration. The western statesmen protested at these lone-handed arrangements but perforce accepted them.

The end of the Japanese war, February–September 1945. While the campaign for the Philippines was still in progress, U.S. forces were making great steps in the direct advance toward their final objective, the Japanese homeland. Aerial bombardment was, of course, the prerequisite of the projected invasion of Japan—which was to begin, it was imagined, with landings on Kyushu, the southernmost of the major Japanese islands.

Iwo Jima and the bombing of Tokyo. With U.S. forces firmly established in the Mariana Islands, the steady long-range bombing of Japan by B-29s under the command of General Curtis E. LeMay continued throughout the closing months of 1944 and into 1945. But it was still 1,500 miles from Saipan to Tokyo, a long flight even for the B-29s. Strategic planners therefore fixed their attention on the little volcanic island of Iwo Jima in the Bonin Islands, which lay about halfway between the Marianas and Japan. If Iwo Jima could be eliminated as a Japanese base, the island could then be immensely valuable as a base for U.S. fighter planes defending the big bombers.

The Japanese were determined to hold Iwo Jima. As they had done on other Pacific islands, they had created underground defenses there, making the best possible use of natural caves and the rough, rocky terrain. The number of Japanese defenders on the island, under command of Lieutenant General Kuribayashi Tadamichi, was more than 20,000.

Day after day before the actual landing the island was subjected to intense bombardment by naval guns, by rockets, and by air strikes using napalm bombs. But the results fell far short of expectations. The Japanese were so well protected that no amount of conventional bombing or shelling could knock them out. U.S. Marines landed on Iwo Jima on Feb. 19, 1945, and encountered an obstinate resistance. Meanwhile, kamikaze counterattacks from the air sank the light carrier *Bismarck Sea* and damaged other ships; and, though the U.S. flag was planted on Mount Suribachi on February 23, the isle was not finally secured until March 16. Iwo Jima had cost the lives of 6,000 Marines, as well as the lives of nearly all the Japanese defenders; but in the next five months more than 2,000 B-29 bombers were able to land on it.

Meanwhile, a new tactic had been found for the bombing of Japan from bases in the Marianas. Instead of high-altitude strikes in daylight, which had failed to do much damage to the industrial centres attacked, low-level strikes at night, using napalm firebombs, were tried, with startling success. The first, in the night of March 9–10, 1945, against Tokyo, destroyed about 25 percent of the city's buildings (most of them flimsily built of wood and plaster), killed more than 80,000 people, and made 1,000,000 homeless. This result indicated that Japan might be defeated without a massive invasion by ground troops, and so similar bombing raids on such major cities as Nagoya, Ōsaka, Kōbe, Yokohama, and Toyama followed. Japan literally was being bombed out of the war.

Low-level incendiary bombing by night

Okinawa. Plans for invasion, however, were not immediately discarded. Okinawa, largest of the Ryukyu Islands strung out northeastward from Taiwan, had been regarded as the last stepping-stone to be taken toward Kyushu, which was only 350 miles away from it. It had therefore been subjected to a series of air raids from October 1944, culminating in March 1945 in an attack that destroyed hundreds of Japanese planes; but there were still at least 75,000 Japanese troops on the island, commanded by Lieutenant General Ushijima Mitsuru. The invasion of Okinawa was, in fact, to be the largest amphibious operation mounted by the Americans in the Pacific war.

Under the overall command of Nimitz, with Admiral Raymond Spruance in charge of the actual landings and with Lieutenant General Simon Bolivar Buckner commanding the ground forces, the operation began with the occupation of the Kerama Islets, 15 miles west of Okinawa, on March 26, 1945. Five days later a landing was made on

Keise-Jima, whence artillery fire could be brought to bear on Okinawa itself. Then, on April 1, some 60,000 U.S. troops landed on the central stretch of Okinawa's west coast, seizing two nearby airfields and advancing to cut the island's narrow waist. Koiso's government in Tokyo resigned on April 5, and the U.S.S.R. on the same day refused to renew its treaty of nonaggression with Japan.

The first major counterattack on Okinawa by the Japanese, begun on April 6, involved not only 355 kamikaze air raids but also the *Yamato,* the greatest battleship in the world (72,000 tons, with nine 18.1-inch [460-millimetre] guns), which was sent out on a suicidal mission with only enough fuel for the single outward voyage and without sufficient air cover. The Japanese hoped the *Yamato* might finish off the Allied fleet after the latter had been weakened by kamikaze attacks. In the event, the *Yamato* was hit repeatedly by bombs and torpedoes and was sunk on April 7. Equally suicidal was a new Japanese weapon, *baka,* which claimed its first victim, the U.S. destroyer *Abele,* off Okinawa on April 12. *Baka* was a rocket-powered glider crammed with explosives which was towed into range by a bomber and was then released to be guided by its solitary pilot into the chosen target for their mutual destruction.

The U.S. ground forces invading Okinawa met little opposition on the beaches because Ushijima had decided to offer his main resistance inland, out of range of the enemy's naval guns. In the southern half of the island this resistance was bitterest: it lasted until June 21, and Ushijima killed himself the next day. The campaign for Okinawa was ended officially on July 2. For U.S. troops it had been the longest and bloodiest Pacific campaign since Guadalcanal in 1942. Taking the island had cost the Americans 12,000 dead and 36,000 wounded, with 34 ships sunk and 368 damaged, and the Japanese losses exceeded 100,000 dead.

On April 3, 1945, two days after the first landing on Okinawa, the U.S. command in the Pacific was reorganized: MacArthur was henceforth to be in command of all army units and also in operational control of the U.S. Marines for the invasion of Japan; Nimitz was placed in command of all navy units.

Hiroshima and Nagasaki. Throughout July 1945 the Japanese mainlands, from the latitude of Tokyo on Honshu northward to the coast of Hokkaido, were bombed just as if an invasion was about to be launched. In fact, something far more sinister was in hand, as the Americans were telling Stalin at Potsdam.

In 1939 physicists in the United States had learned of experiments in Germany demonstrating the possibility of nuclear fission and had understood that the potential energy might be released in an explosive weapon of unprecedented power. On Aug. 2, 1939, Albert Einstein had warned Roosevelt of the danger of Nazi Germany's forestalling other states in the development of an atomic bomb. Eventually, the U.S. Office of Scientific Research and Development was created in June 1941 and given joint responsibility with the war department in the Manhattan Project to develop a nuclear bomb. After four years of intensive and ever-mounting research and development efforts, an atomic device was set off on July 16, 1945, in a desert area at Alamogordo, N.M., generating an explosive power equivalent to that of more than 15,000 tons of TNT. Thus the atomic bomb was born. Truman, the new U.S. president, calculated that this monstrous weapon might be used to defeat Japan in a way less costly of U.S. lives than a conventional invasion of the Japanese homeland. Japan's unsatisfactory response to the Allies' Potsdam Declaration decided the matter. On Aug. 6, 1945, an atomic bomb carried from Tinian Island in the Marianas in a specially equipped B-29 was dropped on Hiroshima, at the southern end of Honshu: the combined heat and blast pulverized everything in the explosion's immediate vicinity, generated spontaneous fires that burned almost 4.4 square miles completely out, and killed between 70,000 and 80,000 people, besides injuring more than 70,000 others. A second bomb, dropped on Nagasaki on August 9, killed between 35,000 and 40,000 people, injured a like number, and devastated 1.8 square miles.

The use of the atomic bomb

The Japanese surrender. News of Hiroshima's destruction was only slowly understood in Tokyo. Many members of the Japanese government did not appreciate the power of the new Allied weapon until after the Nagasaki attack. Meanwhile, on August 8, the U.S.S.R. had declared war against Japan. The combination of these developments tipped the scales within the government in favour of a group that had, since the spring, been advocating a negotiated peace. On August 10 the Japanese government issued a statement agreeing to accept the surrender terms of the Potsdam Declaration on the understanding that the emperor's position as a sovereign ruler would not be prejudiced. In their reply the Allies granted Japan's request that the emperor's sovereign status be maintained, subject only to their supreme commander's directives. Japan accepted this proviso on August 14, and the emperor Hirohito urged his people to accept the decision to surrender. It was a bitter pill to swallow, though, and every effort was made to persuade the Japanese to accept the defeat that they had come to regard as unthinkable. Even princes of the Japanese Imperial house were dispatched to deliver the Emperor's message in person to distant Japanese Army forces in China and in Korea, hoping thus to mitigate the shock. A clique of diehards nevertheless attempted to assassinate the new prime minister, Admiral Suzuki Kantarō; but by September 2, when the formal surrender ceremonies took place, the way had been smoothed.

Truman designated MacArthur as the Allied powers' supreme commander to accept Japan's formal surrender, which was solemnized aboard the U.S. flagship *Missouri* in Tokyo Bay: the Japanese foreign minister, Shigemitsu Mamoru, signed the document first, on behalf of the Emperor and his government. He was followed by General Umezu Yoshijiro on behalf of the Imperial General Headquarters. The document was then signed by MacArthur, Nimitz, and representatives of the other Allied powers. Japan concluded a separate surrender ceremony with China in Nanking on Sept. 9, 1945. With this last formal surrender, World War II came to an end.

COSTS OF THE WAR

Killed, wounded, prisoners, or missing. The statistics on World War II casualties are inexact. Only for the United States and the British Commonwealth can official figures showing killed, wounded, prisoners or missing for the armed forces be cited with any degree of assurance. For most other nations, only estimates of varying reliability exist. Statistical accounting broke down in both Allied and Axis nations when whole armies were surrendered or dispersed. Guerrilla warfare, changes in international boundaries, and mass shifts in population vastly complicated postwar efforts to arrive at accurate figures even for the total dead from all causes.

Civilian deaths from land battles, aerial bombardment, political and racial executions, war-induced disease and famine, and the sinking of ships probably exceeded battle casualties. These civilian deaths are even more difficult to determine, yet they must be counted in any comparative evaluation of national losses. There are no reliable figures for the casualties of the Soviet Union and China, the two countries in which casualties were undoubtedly greatest. Mainly for this reason, estimates of total dead in World War II vary anywhere from 35,000,000 to 60,000,000—a statistical difference of no small import. Few have ventured even to try to calculate the total number of persons who were wounded or permanently disabled.

However inexact many of the figures, their main import is clear. The heaviest proportionate human losses occurred in eastern Europe where Poland lost perhaps 20 percent of its prewar population, Yugoslavia and the Soviet Union around 10 percent. German losses, of which the greater proportion occurred on the Eastern Front, were only slightly less severe. The nations of western Europe, however great their suffering from occupation, escaped with manpower losses that were hardly comparable with those of World War I. In East Asia, the victims of famine and pestilence in China are to be numbered in the millions, in addition to other millions of both soldiers and civilians who perished in battle and bombardment.

Table 7: World War II Casualties

country	military			civilian deaths due to war	estimated total deaths
	killed, died of wounds, or in prison*	wounded	prisoners or missing†		
Allied Powers					
Belgium	12,000	—	—	76,000	88,000
Brazil	943	4,222	—	—	1,000
British Commonwealth	373,372	475,047	251,724‡	92,673	466,000
Australia	23,365	39,803	32,393	—	24,000
Canada	37,476	53,174	10,888	—	38,000
India	24,338	64,354	91,243	—	—
New Zealand	10,033	19,314	10,582	—	10,000
South Africa	6,840	14,363	16,430	—	7,000
United Kingdom	264,443	277,077	213,919	92,673§	357,000
Colonies	6,877	6,972	22,323	—	7,000
China ‖	1,310,224	1,752,951	115,248	—	—
Czechoslovakia¶	10,000	—	—	215,000	225,000
Denmark	1,800	—	—	2,000♀	4,000
Franceδ	213,324	400,000	—	350,000	563,000
Greece▫	88,300	—	—	325,000	413,000
Netherlands	7,900	2,860	—	200,000	208,000
Norway	3,000	—	—	7,000	10,000
Poland♦	123,178	236,606	420,760	5,675,000	5,800,000
Philippines	27,000	—	—	91,000	118,000
United States△	292,131	671,801	139,709	6,000	298,000
U.S.S.R.+	11,000,000	—	—	7,000,000	18,000,000
Yugoslavia	305,000	425,000	—	1,200,000	1,505,000
Axis Powers					
Bulgaria⊕	10,000	—	—	10,000	20,000
Finland	82,000	50,000	—	2,000	84,000
Germany**	3,500,000	5,000,000	3,400,000	780,000	4,200,000
Hungary⊕	200,000	—	170,000	290,000	490,000
Italy††	242,232	66,000	350,000	152,941	395,000
Japan	1,300,000‡‡	4,000,000	810,000	672,000	1,972,000
Romania⊕	300,000	—	100,000	200,000	500,000

*Figures for deaths, insofar as possible, exclude those who died of natural causes or were suicides. †As far as possible the figures in this column exclude those who died in captivity. ‡Figures for all Commonwealth nations include those still missing in 1946, some of whom may be presumed dead. §This figure comprises 60,595 killed in aerial bombardment, 30,248 in the merchant marine service, 624 in women's auxiliary services, and 1,206 in the Home Guard. ‖ The figures for China comprise casualties of the Chinese Nationalist forces during 1937–45, as reported in 1946, and do not include figures for local armies and Communists. Estimates of 2,200,000 military dead and 22,000,000 civilian deaths appear in some compilations but are of doubtful accuracy. ¶Czech military figures include only those who fought on the Allied side, not Sudeten Germans and others who served in the German Army. ♀Includes merchant marine personnel who served with Allies. δFrench military casualties include those dead from all causes in the campaign of 1939–40, those of Free French, of rearmed French units that fought with Allies during 1942–45, and of French units that fought with Axis forces in Syria and North Africa during 1941–42 (1,200 dead). ▫These figures released in 1946 are possibly too high. Merchant seamen are included with military dead. ♦Military figures drawn from statement released by Polish government in 1946 and include casualties in the campaign of 1939, those of the underground, of Polish forces serving with British and Soviet armies, and those incurred in the Warsaw uprising. Civilian casualty figures, which include 3,200,000 Jews, are based on this statement as modified by the calculations of population experts. △Military figures include those of Army Ground and Air Forces, and those of the Navy, Marines, and Coast Guard. There were an additional 115,187 deaths of U.S. servicemen from non-battle causes. Civilians listed in 1946 as dead or missing include 5,638 of the merchant marine service. +Available estimates of Soviet casualties vary widely. A Soviet officer who served with the high command in Berlin and left the Soviet service in 1949 placed total military losses at 13,600,000—8,500,000 dead or missing in battle; 2,600,000 dead in prison camps; 2,500,000 died of wounds—and estimated civilian casualties at 7,000,000. These figures have been widely accepted in Germany, but most U.S. compilations, based on Soviet announcements, list 6,000,000 to 7,500,000 battle deaths. Calculations made on the basis of population distribution by age and sex in the 1959 U.S.S.R. census give some credence to the higher figures, for they seem to indicate losses of from 15,000,000 to 20,000,000 males of military age in World War II. The figures used here are a compromise estimate, not intended to obscure the fact that Soviet casualties are, in reality, unknown in the West. ⊕Estimates based on fragmentary data. **Military estimates include men from outside Germany who served with the German armed forces and are based on the assumption that about 1,000,000 of the 1,250,000 men still listed as missing in Soviet territory in 1955 were dead. In addition, perhaps 250,000 military personnel died of natural causes, committed suicide, or were executed. Civilian figures are for Germany and Austria only, and they do not include an estimated 2,384,000 German deaths during 1944–46 resulting from Soviet invasion and forced transfers of population in the eastern provinces given to Poland after the war. ††Figures for dead include those listed as still missing in compilation made by the Italian government in 1952 (131,419 military personnel and 3,651 civilians), but not 49,144 military deaths from natural causes or suicide. Known dead from enemy action amounted to 110,823, making a total of 159,957 military deaths from all causes if the missing are not inlcuded. Of this number, 92,767 occurred before the 1943 Armistice, 67,190 afterward. ‡‡Based on an estimate of 1,600,000 total military deaths on the assumption that about half of those listed as missing in Soviet territory in 1949 were dead. About 300,000 of these probably resulted from causes not related to battle.

Table 7 contains what appear to be the best available statistics on armed forces casualties of all types resulting from battle, of civilian deaths from war-related causes, and estimated total deaths in each of the major nations involved in World War II. Figures rounded to thousands (and this device has been employed in all cases for total deaths) are estimates of varying reliability while omissions in any category indicate that any estimate would be the wildest of conjectures. Estimated casualties of resistance movements have been included in military figures, other victims of Nazi persecution in the civilian ones. In the latter category fall about 5,700,000 Jews, more than half of them from Poland, who died in Nazi concentration and death camps.

Human and material cost. There can be no real statistical measurement of the human and material cost of World War II. The money cost to governments involved has been estimated at more than $1,000,000,000,000 but this figure cannot represent the human misery, deprivation, and suffering, the dislocation of peoples and of economic life, or the sheer physical destruction of property that the war involved.

Europe. The Nazi overlords of occupied Europe drained their conquered territories of resources to feed the German war machine. Industry and agriculture in France, Belgium, The Netherlands, Denmark, and Norway were forced to produce to meet German needs with a resulting deprivation of their own peoples. Italy, though at first a German ally, fared no better. The resources of the occupied territories in eastern Europe were even more ruthlessly exploited. Millions of able-bodied men and women were drained away to perform forced labour in German factories and on German farms. The whole system of German economic exploitation was enforced by cruel and brutal methods, and the guerrilla resistance it aroused was destructive in itself and provoked German reprisals that were even more

destructive, particularly in Poland, Yugoslavia, and the occupied portions of the Soviet Union.

Great Britain, which escaped the ravages of occupation, suffered heavily from the German aerial blitz of 1940–41 and later from V-bombs and rockets. On the other side, German cities were leveled by Allied bombers, and in the final invasion of Germany from both east and west there was much retaliatory devastation, destruction, and pillage.

The destruction of physical plant was immense and far exceeded that of World War I, when it was largely confined to battle areas. France estimated the total cost at an amount equivalent to three times the total French annual national income. Belgium and The Netherlands suffered damage roughly in similar proportions to their resources. In Great Britain about 30 percent of the homes were destroyed or damaged; in France, Belgium, and The Netherlands about 20 percent. Agriculture in all the occupied countries suffered heavily from the destruction of facilities and farm animals, the lack of machinery and fertilizers, and the drain on manpower. Internal transport systems were completely disrupted by the destruction or confiscation of rail cars, locomotives, and barges, and the bombing of bridges and key rail centres. By 1945 the economies of the continental nations of western Europe were in a state of virtually complete paralysis.

In eastern Europe the devastation was even worse. Poland reported 30 percent of its buildings destroyed, as well as 60 percent of its schools, scientific institutions, and public administration facilities, 30–35 percent of its agricultural property, and 32 percent of its mines, electrical power, and industries. Yugoslavia reported 20.7 percent of its dwellings destroyed. In the battlegrounds of the western portion of the Soviet Union, the destruction was even more complete. In Germany itself, the U.S. Strategic Bombing Survey found that in 49 of the largest cities, 39 percent of the dwelling units were destroyed or seriously damaged. Central business districts had generally been reduced to rubble, leaving only suburban rings standing around a destroyed core.

Millions throughout Europe were rendered homeless. There were an estimated 21,000,000 refugees, more than half of them "displaced persons" who had been deported from their homelands to perform forced labour. Other millions who had remained at home were physically exhausted by five years of strain, suffering, and undernourishment. The roads of Europe were swamped by refugees all through 1945 and into 1946 as more than 5,000,000 Soviet prisoners of war and forced labourers returned eastward to their homeland and more than 8,000,000 Germans fled or were evacuated westward out of the Soviet-occupied portions of Germany. Millions of other persons of almost every European nationality also returned to their own countries or emigrated to new homes in other lands.

The Far East. The devastation of World War II in China was inflicted on a country that was already suffering from the economic ills of overpopulation, underdevelopment, and a half-century of war, political disunity, and unrest. The territory occupied by Japanese forces was roughly equivalent to that occupied by the Axis in Europe and the period of occupation was longer. That area of China unoccupied by the Japanese was virtually cut off from the outside world after the Japanese conquest of Burma in early 1942, and its economy continually tottered on the brink of collapse. In both areas, famines, epidemics, and civil unrest were recurrent, much farmland was flooded, and millions of refugees fled their homes, some several times. Cities, towns, and villages were laid waste by aerial bombardment and marching armies. The transportation system, poor to begin with, was thoroughly disrupted. Most of the limited number of hospitals and health institutions in China were destroyed or lost.

In India famine was recurrent, and the Indian economy was severely strained to support the burden the Allied military authorities placed upon it. The Philippines suffered from three years of Japanese occupation and exploitation and from the destruction wrought in the reconquest of the islands by the Americans in 1944–45. The harbour at Manila was wrecked by the retreating Japanese, and many portions of the city were demolished by bombardment.

In Japan the U.S. Strategic Bombing Survey found the damage to urban centres comparable to that in Germany. In the aggregate, 40 percent of the built-up areas of 66 Japanese cities was destroyed, and approximately 30 percent of the entire urban population of Japan lost their homes and many of their possessions. Hiroshima and Nagasaki suffered the peculiar and lasting damage done by atomic explosion and radiation.

BIBLIOGRAPHY

World War I: DWIGHT E. LEE (ed.), *The Outbreak of the First World War: Causes and Responsibilities*, 4th ed. (1975), is a good introduction to the debate on the origins of the war. A more detailed account is provided in LUIGI ALBERTINI, *The Origins of the War of 1914*, 3 vol. (1952–57, reprinted 1980; originally published in Italian, 1942–43). A comprehensive general account of the war is BERNADOTTE E. SCHMITT and HAROLD C. VEDELER, *The World in the Crucible, 1914–1919* (1984). The standard military histories are BASIL HENRY LIDDELL HART, *A History of the World War, 1914–1918*, enl. ed. (1934, reprinted 1970); CYRIL B. FALLS, *The Great War* (1959); and MARC FERRO, *The Great War, 1914–1918* (1973, reprinted 1987; originally published in French, 1969). J.E. EDMONDS (comp.), *A Short History of World War I* (1951, reprinted 1968); and JOHN TERRAINE, *The Great War, 1914–1918: A Pictorial History* (1965, reprinted 1978), are useful introductions. Naval operations during the war are discussed in ARTHUR J. MARDER, *From the Dreadnought to Scapa Flow: The Royal Navy in the Fisher Era, 1904–1919*, 5 vol. (1961–70). International diplomacy during the war is treated in Z.A.B. ZEMAN, *The Gentlemen Negotiators* (1971; U.K. title, *A Diplomatic History of the First World War*).

World War II: Good summaries of the origins of World II include JOACHIM REMAK, *The Origins of the Second World War* (1976); and ANTHONY P. ADAMTHWAITE, *The Making of the Second World War* (1979). The origins of the war in the Pacific are treated in HERBERT FEIS, *The Road to Pearl Harbor: The Coming of the War Between the United States and Japan* (1950, reprinted 1971). The classic comprehensive military histories of the war are BASIL HENRY LIDDELL HART, *History of the Second World War* (1970, reissued 1979); J.F.C. FULLER, *The Second World War, 1939–45: A Strategical and Tactical History* (1948, reprinted 1968). An excellent overall history of Europe during the war is GORDON WRIGHT, *The Ordeal of Total War, 1939–45* (1968). The first phase of the war is treated in JOHN LUKACS, *The Last European War, 1939–1941* (1976). Treatments of the war in the Pacific include JOHN COSTELLO, *The Pacific War* (1981); AKIRA IRIYE, *Power and Culture: The Japanese-American War, 1941–1945* (1981); and WILLIAM CRAIG, *The Fall of Japan* (1967, reissued 1979). The standard naval histories are S.W. ROSKILL, *The War at Sea, 1939–1945*, 3 vol. in 4 (1954–61); and FRIEDRICH RUGE, *Der Seekrieg: The German Navy's Story, 1939–1945* (1957; U.K. title, *Sea Warfare, 1939–1945: A German Viewpoint*; originally published in German, 1954). Aerial operations are treated in R.J. OVERY, *The Air War, 1939–1945* (1980, reprinted 1987); and CHARLES WEBSTER and NOBLE FRANKLAND, *The Strategic Air Offensive Against Germany, 1939–1945*, 4 vol. (1961). Allied cooperation and diplomacy are treated in WILLIAM HARDY MCNEILL, *America, Britain, and Russia: Their Cooperation and Conflict, 1941–1946* (1953, reprinted 1970); HERBERT FEIS, *Churchill, Roosevelt, Stalin: The War They Waged and the Peace They Sought*, 2nd ed. (1967); and CHRISTOPHER THORNE, *Allies of a Kind: The United States, Britain, and the War Against Japan, 1941–1945* (1978). Decisions concerning the atomic bomb are outlined in HERBERT FEIS, *The Atomic Bomb and the End of World War II*, rev. ed. (1966); and MARTIN J. SHERWIN, *A World Destroyed: The Atomic Bomb and the Grand Alliance* (1975, reprinted 1987).

(J.R.-S./Ed.)

Writing

Writing is a form of human communication by means of a set of visible marks that are related, by convention, to some particular structural level of language. This definition highlights the fact that writing is in principle the representation of language rather than a direct representation of thought and the fact that spoken language has a number of levels of structure including sentences, words, syllables, and phonemes (the smallest units of speech used to distinguish one word or morpheme from another), any one of which a writing system can "map onto" or represent. Indeed, the history of writing is in part a matter of the discovery and representation of these structural levels of spoken language in the attempt to construct an efficient, general, and economical writing system capable of serving a range of socially valuable functions. Literacy is a matter of competence with a writing system and with the specialized functions that written language serves in a particular society.

This article treats the nature and origin of writing and its development in the world's major cultural regions. For discussion of the study of writing as a tool of historical research, see the sections *Epigraphy* and *Paleography* in HISTORY, THE STUDY OF.

For coverage of related topics in the *Macropædia* and *Micropædia,* see the *Propædia,* section 514.

The article is divided into the following sections:

The nature and origin of writing 1025
- Writing as a system of signs 1025
- The functions of writing 1027
- Types of writing systems 1027
- History of writing systems 1028
  - Sumerian writing
  - Chinese writing
- Literacy: The uses of writing 1032
  - The rise of literacy
  - Literacy and schooling

Systems of writing 1034
- Cuneiform 1034
  - Origin and character of cuneiform
  - Spread and development of cuneiform
  - Decipherment of cuneiform
  - Influence of cuneiform
- Hieroglyphic writing 1036
  - Development of Egyptian hieroglyphic writing
  - Characteristics of hieroglyphic writing
  - Hieratic script
  - Demotic script
  - Decipherment of hieroglyphic writing
- Alphabetic writing 1040
  - Theories of the origin of the alphabet
  - Development and diffusion of alphabets
  - Major alphabets of the world
- Chinese writing 1049

Adjuncts to writing 1050
- Punctuation 1050
  - Punctuation in Greek and Latin to 1600
  - Punctuation in English since 1600
  - Punctuation in French, Spanish, German, and Russian
  - Punctuation in Oriental and African languages
- Shorthand 1052
  - History and development of shorthand
  - Modern symbol systems
  - Modern abbreviated longhand systems
  - Machine shorthand

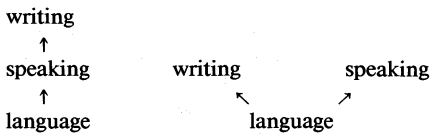
Handwriting and calligraphy 1054
- Early Semitic writing 1054
  - Old Hebrew
  - Spread of Aramaic to the Middle East and Asia
- Arabic calligraphy 1055
- Indic calligraphy 1057
- Greek handwriting 1057
  - Origins to the 8th century AD
  - The 8th to 16th century
- Latin-alphabet handwriting 1063
  - Ancient Roman styles
  - The Anglo-Celtic and other "national" styles (5th to 13th century)
  - Carolingian reforms in the scriptorium (8th and 9th centuries)
  - The black-letter, or Gothic, style (9th to 15th century)
  - The scripts of humanism (14th to 16th century)
  - Writing manuals and copybooks (16th to 18th century)
  - Revival of calligraphy (19th and 20th centuries)
- East Asian calligraphy 1072
  - Chinese calligraphy
  - Korean calligraphy
  - Japanese calligraphy

Bibliography 1075

The nature and origin of writing

WRITING AS A SYSTEM OF SIGNS

Languages are systems of symbols; writing is a system for symbolizing these symbols. A writing system may be defined as any conventional system of marks or signs that represents the utterances of a language. Writing renders language visible; while speech is ephemeral, writing is concrete and, by comparison, permanent. Both speaking and writing depend upon the underlying structures of language. Consequently, writing cannot ordinarily be read by someone not familiar with the linguistic structure underlying the oral form of the language. Yet writing is not merely the transcription of speech; writing frequently involves the use of special forms of language, such as those involved in literary and scientific works, which would not be produced orally. In any linguistic community the written language is a distinct and special dialect; usually there is more than one written dialect. Scholars account for these facts by suggesting that writing is related directly to language but not necessarily directly to speech. Consequently, spoken and written language may evolve somewhat distinctive forms and functions. These alternative relations may be depicted as follows:



It is the fact that writing is an expression of language rather than simply a way of transcribing speech that gives to writing, and hence to written language and to literacy, their special properties. As long as writing was seen merely as transcription, as it was by such pioneering linguists as Ferdinand de Saussure and Leonard Bloomfield earlier in the 20th century, its conceptual significance was seriously underestimated. Once writing was seen as providing a new medium for linguistic expression, its distinctness from speech was more clearly grasped. Scholars such as Milman Parry, Marshall McLuhan, Eric Havelock, Jack Goody, and Walter Ong were among the first to analyze the conceptual and social implications of using written as opposed to oral forms of communication.

Writing is merely one, albeit the most important, means of communicating by visible signs. Gestures—such as a raised hand for greeting or a wink for intimate agreement—are visible signs but they are not writing in that

they do not transcribe a linguistic form. Pictures, similarly, may represent events but do not represent language and hence are not a form of writing.

But the boundary between pictures and writing becomes less clear when pictures are used conventionally to convey particular meanings. In order to distinguish pictures from pictorial signs it is necessary to notice that language has two primary levels of structure that the French linguist André Martinet referred to as the "double articulation" of language: the meaning structures on one hand and the sound patterns on the other. Indeed, linguists define grammar as a system for mapping—establishing a system of relations between—sound and meaning. These levels of structure admit of several subdivisions, any one of which may be captured in a writing system. The basic unit of the meaning system is called a morpheme; one or more morphemes make up a word. Thus the word *boys* is composed of two morphemes, *boy* and plurality. Grammatically related words make up clauses that express larger units of meaning. Still larger units make up such discourse structures as propositions and less well-defined units of meaning such as prayers, stories, and poems.

Units of meaning

The basic linguistic unit of the sound system is called a phoneme; it is a minimal, contrastive sound unit that distinguishes one utterance from another. Phonemes may be further analyzed in terms of a set of underlying distinctive features, features specifying the ways the sound is physically produced by passing breath through the throat and positioning the tongue and lips. Phonemes may be thought of as roughly equivalent to the sound segments known as consonants and vowels, and combinations of these segments make up syllables.

Writing systems can serve to represent any of these levels of sound or any of the levels of meaning, and, indeed, examples of all of these levels of structure have been exploited by some writing system or other. Writing systems, consequently, fall into two large general classes, those that are based on some aspect of meaning structure, such as a word or morpheme, and those that are based on some aspect of the sound system, such as the syllable or phoneme.

The earlier failure to recognize these levels of structure in language led some scholars to believe that some writing systems, so-called ideograms and pictograms, had been invented to express thought directly, bypassing language altogether. The 17th-century German philosopher Gottfried Leibniz set out to invent the perfect writing system, which would reflect systems of thought directly and thereby be readable by all human beings regardless of their mother tongues. We now know that such a scheme is impossible. Thought is too intimately related to language to be represented independently of it.

More recently there have been attempts to invent forms for communicating explicit messages without assuming a knowledge of any particular language. Such messages are communicated by means of pictorial signs. Thus the skirted human figure painted on the door to a toilet, the human figure with an upraised hand on the Pioneer spacecraft, the Amerindian drawing of a horse and rider upside down painted on a rock near a precipitous trail, and the visual patterns branded on range cattle are all attempts to use visual marks to communicate without making any appeal to the structure of any particular language.

However, such signs function only because they represent a high level of linguistic structure and because they function to express one of a highly restricted range of meanings already known to the reader and not because they express ideas or thoughts directly. The sign on the toilet door is an elliptical way of writing "women's washroom," just as the word "women" had been earlier. The plaque on the spacecraft can be read as a greeting only if the reader already knows how to express a human greeting symbolically. The inverted horse and rider expressed the message that horses and riders should avoid the trail. And the brand can be read as the name of the owner's ranch.

Such signs, therefore, express meanings, not thoughts, and they do so by representing meaning structures larger than can be expressed by a single word. They do so by expressing these meanings elliptically. Such signs are readable because the reader has to consider only a restricted

Some of the pictorial signs used at the 1984 Summer Olympic Games in Los Angeles, Calif.
By courtesy of the Comite International Olympique

set of possible meanings. While such pictorial signs could not be turned into a general writing system, they can be extremely efficient in serving a restricted set of functions.

The differences between such pictorial signs and other forms of writing are sufficiently great for some scholars to maintain that they are not legitimate types of writing. These differences are that pictorial signs are "motivated," that is, they visually suggest their meanings, and that they express whole propositions rather than single words. Other scholars would include such signs as a form of writing because they are a conventional means for expressing a particular linguistic meaning. Scholars agree that such a collection of signs could express only an extremely limited set of meanings.

A similar case is the ancient mosaic found at the entrance of a house in Pompeii, depicting a snarling dog on a chain and bearing the inscription *cave canem,* "beware of the dog." Even nonreaders could "read" the message; the picture is therefore a form of writing rather than of picture making. Such pictorial signs, including logotypes, trademarks, and brand names, are so common in modern urban societies that even very young children learn to read them. Such reading ability is described as "environmental" literacy, not associated with books and schooling.

Similarly, number systems have posed a problem for theorists because such symbols as the Arabic numerals *1, 2, 3,* etc., which are conventional across many languages, appear to express thought directly without any intermediary linguistic structure. However, it is more useful to think of these numerals as a particular orthography for representing the meaning structure of these numbers rather than their sound structures. The advantages of this orthography are that the orthography permits the user to carry out mathematical operations, such as carrying, borrowing, and the like, and that the same orthography may be assigned different phonological equivalents in different languages using the same number system. Thus, the numeral *2* is pronounced "two" in English, "deux" in French, "zwei" in German, and so on. Yet it represents not a thought but the word, a piece of language.

It is for these reasons that writing is said to be a system for transcribing language, not for representing thought directly. There are, of course, other systems for representing thought, including such activities as picture making, dance, and mime. These, however, are not representations of ordinary language; rather, they constitute what the American philosopher Nelson Goodman has called the "languages of art." These "languages," or semiotic systems, are systems of signs that are used for expressive and representational purposes. Each of these semiotic systems may, in turn, be represented by a notational system, a system for representing the semiotic system. Thus writing can be defined formally as a notational system for representing some level or levels of linguistic form.

Writing is so pervasive in everyday life that many people take it to be synonymous with language, and this confusion affects their understanding of language. The word *word* denotes ambiguously both the oral form and the written

form, and so people may confuse them. This occurs, for example, when people think that the sounds of language are made up of letters. Even Aristotle used the same word, *gramma,* to refer to the basic units of both speech and writing. Yet it is important to distinguish them. People may have competence in a language and yet know nothing about its written form. Similarly, writing is so fundamental to a modern, literate society that its significance has often been overestimated. Since the 18th century it has been common to identify literacy with civilization, indeed with all civil virtues. When European countries colonized other regions they thought it as important to teach "savages" to read and write as to convert them to Christianity. Modern anthropology has helped to revise what now seems a quaint set of priorities by showing not only that there are no genuinely primitive languages, but that differing languages mask no unbridgeable differences between human beings. All human beings are rational, speak a language of enormous expressive power, and live in, maintain, and transmit to their young a complex social and moral order.

Scholars of literature have in the past half-century amassed compelling evidence to demonstrate that a complex social order and a rich verbal culture can exist in nonliterate societies. The American scholar Milman Parry, writing in the 1920s, showed that the Homeric epic poems, long regarded as models of literary virtuosity, were in fact the product not of a literate but of an oral tradition. These poems were produced by bards who could not write and were delivered in recitals to audiences who could not read. Writing made possible the recording of these poems, not their composition. The hard and fast dividing line that put civilization and literacy on one side and savagery and irrationality on the other has been abandoned. To be unlettered is no longer confused with being ignorant.

Similarly, it was once generally held that all writing systems represent some stage in a progression toward the ideal writing system, the alphabet. The accepted view today is that all writing systems represent relatively optimal solutions to a large and unique set of constraints, including the structure of the language represented, the functions that the system serves, and the balance of advantages to the reader as opposed to the writer. Consequently, while there are important differences between speaking and writing and between various forms of writing, these differences vary in importance and in effect from language to language and from society to society.

THE FUNCTIONS OF WRITING

Given that literacy is not a prerequisite of rationality and civilization, it may be asked why writing systems were invented and why, when they were, they so completely displaced preexisting oral traditions. Many accounts have been given of the dramatic impact on an "oral" culture of the encounter with written text. Isak Dinesen, in her autobiographical *Out of Africa,* reported on the response of Kikuyu tribesmen to their first exposures to written texts: "I learned that the effect of a piece of news was many times magnified when it was imparted by writing. The messages that would have been received with doubt and scorn if they had been given by word of mouth were now taken as gospel truth."

Certainly writing has been observed to displace oral traditions. The American scholar Albert Lord wrote: "When writing is introduced and begins to be used for the same purposes as the oral narrative song, when it is employed for telling stories and is widespread enough to find an audience capable of reading, this audience seeks its entertainment and instruction in books rather than in the living songs of men, and the older art gradually disappears."

The adoption and use of writing systems depend primarily on their ability to preserve language and information through time and across space. But the use of a writing system for this purpose is shaped in part by the nature of the system and by the cultural practices in the society that has adopted it. These uses tend, therefore, to be local and specific and characteristic of a particular literate society.

The Canadian economist Harold Innis classified writing systems into two basic types: those that bind through time, exemplified by Egyptian hieroglyphics carved in stone and Akkadian cuneiform incised in clay; and those that bind across space, exemplified by the portable papyri used by the Romans. Writing used to store information for posterity may be considered to serve an archival function. Such writing may be used not only for constructing, accumulating, and preserving records of political, religious, scientific, and literary interest but also for the more mundane purpose of keeping trade accounts and records. Writing used to transmit information across space, as in letters, encyclicals, newspapers, and the like, may be considered to serve a communicative function. Writing used for purely private ends, to record notes, diaries, or other personal data, may be considered to serve a mnemonic function.

Almost any notational form may be used for mnemonic purposes, for only the person who "wrote" the message needs to be able to "read" it. The carved notches in a wooden counting stick or the pebbles in a counting sack corresponding to the number of cattle under the care of a cowherd are suitable aide-mémoire, since the writer knows what the notches or pebbles represent. But such a system could not be read by others; it would not be clear what the notches represented or even that they represented anything at all. For a writing system to be communicative, the signs must be conventionalized so that the meaning can be grasped by other readers. Such a system may be restricted to a small set of familiar messages that can be read by a limited circle of acquaintances. But for a writing system to serve an archival function it must be sufficiently conventionalized to permit decoding and interpretation by readers who may know nothing about the writer or his message. It is only with the development of explicit writing systems capable of representing the nuances conveyed in speech that writing can be used archivally or communicatively.

TYPES OF WRITING SYSTEMS

A writing system, technically referred to as a script or orthography, consists of a set of visible marks, forms, or structures called characters or graphs that are related to some structure in the linguistic system. Roughly speaking, if a character represents a meaningful unit, such as a morpheme or a word, the orthography is called a logographic writing system; if it represents a syllable, it is called a syllabic writing system; if a segment of a syllable, it is called a consonantal writing system or an unvocalized syllabary; and if a phoneme, it is called an alphabetic system. (A phonetic alphabet, such as the one devised by the International Phonetic Association, is one designed to transcribe any oral language into a common script.) Finally, a writing system based upon the articulatory features that underlie the phoneme, such as voicing and place of articulation, is called a featural writing system. These relations may be depicted as follows:

linguistic structure		orthographic structure
meaning-based	text	—
	topic	—
	speech act	pictorial signs
	word / morpheme	logographic writing
sound-based	syllable	syllabic writing
	segment	consonantal writing
	phoneme	alphabetic writing
	phone	phonetic alphabet
	feature	featural writing system

While relatively pure examples of these different types of script are known, most writing systems that have been used for general purposes combine properties of more than one type.

Pictorial signs, such as the informational signs at a modern international airport (insofar as they can properly be called writing) can bear explicit linguistic messages only because of the extremely limited set of alternatives a reader is required to choose among. Such writing is of little use for conveying new messages since there is no convention for decoding them, and to that extent it cannot be a general writing system. It can, however, serve a limited set of purposes efficiently.

General writing systems all analyze the linguistic form

into constituents of meaning or sound. Chinese script is primarily a logographic script; each word or morpheme is represented by a single graph or character. Two words, even if they sound exactly the same, will be represented by entirely dissimilar characters. But as the number of distinguishable words in a language can run into the tens of thousands (written English has a recorded vocabulary of some 1,500,000 words), the number of logographic characters to be memorized is extremely large.

Syllabaries provide a distinctive symbol for each distinct syllable. A syllable is a unit of speech composed of a vowel sound or a combination of consonant and vowel sounds; the sounds *pa, pe, pi, po, pu* are different syllables and are easily distinguished in a word. The word *paper* has two syllables, *pa-per.* A syllabary such as Linear B, the Mycenaean script dating from about 1400 BC, would have a graph for each of those syllables. Syllables are the most readily distinguishable units of speech, and consequently, the earliest of the sound-based, or phonographic, writing systems are syllabic. The number of syllables in a language, while differing considerably from language to language, is always quite large, hence some hundreds of graphs may be required to make a functioning syllabary. Even then such writing systems are far from explicit, for any string of syllabic graphs may be read in a number of different ways. Reading of such a script would rely upon the reader's prior knowledge and ability to work from the context, along with some guesswork.

Consonantal systems

Consonantal writing systems, as the name implies, represent the consonantal value of a syllable while ignoring the vocalic element. Such a system, therefore, would represent the syllables *pa, pe, pi, po, pu* with a single character. Such scripts have graphs for consonant sounds but not for vowel sounds, with the result that a certain amount of guesswork is involved in determining which syllable is being represented. This ambiguity, however, should not be overemphasized. When a consonantal system is used to represent a language like English, in which vowels differentiate root morphemes (in English, *pat, pet, pit, pot, put* are all different morphemes), discarding the vowel results in a highly ambiguous written expression that can be understood only by a reader who already had a good idea of the content of the written message. But in Semitic languages, such as Hebrew and Arabic, the absence of characters representing vowels is much less serious, because in these languages vowel differences generally do not distinguish morphemes. Vowel differences mark inflections, such as tense and aspect, that, while of some importance to the representation of meaning, are both more readily recovered from context and less likely to change the overall meaning. The failure to notice the intimate relation between the morphophonemic structure of the language and the type of orthography has led some scholars to underestimate the efficiency of consonantal writing systems and, perhaps, to overestimate the centrality of the invention of the alphabet to the evolution of Western culture.

Alphabetic writing systems represent the phonological structure of the language. The smallest pronounceable segment of speech is a syllable, but a syllable may be analyzed into the distinctive underlying constituents called phonemes. The syllable *pa* is produced by passing a column of air through the vocal chords, an action that constitutes the vocalic element, bounded at the outset by sudden release of air through the lips, an action that constitutes the consonantal element. The achievement of the alphabet is to analyze the syllable into its underlying consonant and vowel constituents. The economy of representation comes from the fact that a large number of syllables can be generated from a small set of these constituents. An alphabet consisting of 21 consonants and five vowels can generate 105 simple consonant-and-vowel syllables and more than 2,000 consonant-vowel-consonant syllables. In short, an alphabet can represent a full range of phonological differences. It is a script particularly suited to representing a language in which morphological differences are marked in phonological differences; it is less useful for a language, such as Chinese, in which one syllable represents a large number of morphemes. For the Chinese language a logographic system is more efficient.

Featural systems

Featural writing systems exploit the fact that even phonemes are not the most fundamental units of analysis of speech. Rather, phonemes may be analyzed into sets of distinctive features. The phonemes represented by the letters *n* and *d* share the feature of the tongue touching the alveolar ridge above the upper teeth. Featural writing systems analyze the sounds described as consonants and vowels into their shared and distinguishing features. Examples of writing systems that employ at least in part a featural approach are the Korean Han'gŭl script created, according to tradition, by King Sejong in the 15th century and Pitman shorthand, a system for rapid writing invented in Britain in the 19th century. In Han'gŭl, vowels are represented by long horizontal or vertical lines distinguished by small marks, while consonants are represented by two-dimensional signs that suggest the articulations involved: pairs of lines representing lips together, tongue touching the roof of the mouth, an open throat, and the like. As the phonological system is organized around some dozen such features, an efficient script can be constructed out of 24 basic graphs. In addition, such a script makes syllables visually discriminable by organizing them into blocks to facilitate rapid reading. Such properties led the British linguist Geoffrey Sampson to say: "Whether or not it is ultimately the best of all conceivable scripts for Korean, Han'gul must unquestionably rank as one of the great intellectual achievements of humankind."

No orthography is a pure system. The clearest example of logographic writing, Chinese, consists not only of characters representing meanings but also of secondary characters based on sound similarity for representing meanings that were difficult to picture. It therefore relies upon both word-based and sound-based principles. On the other hand, alphabets, which are primarily sound-based, also use fixed letter strings to represent the same meaningful unit even if the pronunciation of that unit varies in different contexts. So, for example, the common spelling for the root *photo* is preserved in the words *photograph* and *photography* even though they are pronounced somewhat differently. Conversely, alphabets often provide different graphic representations for homophones (words that sound identical but have different meanings) the more clearly to distinguish their meanings, as in *meat, meet, mete; pain, pane; be, bee.* The morphemic unit is so fundamental to the reading process that some linguists have concluded that for an orthography to be practical and efficient, it is more important to provide an invariant visual form for each meaningful unit than for each sound unit.

The shaping of a writing system to make it suitable for a wide range of cultural purposes required other developments besides the invention of a system of characters for representing linguistic form. To facilitate fast and accurate recognition, the form of writing was improved by introducing spaces between the words, developing conventions for punctuation and paragraphing, and simplifying graphic forms. This evolution continued through the invention of printing and the invention of type fonts. And to exploit the aesthetic properties of the writing system, artistic forms of writing were developed (see below *Calligraphy*).

HISTORY OF WRITING SYSTEMS

While speaking is a universal human competence that has been characteristic of the species from the beginning and that is acquired by all normal human beings without systematic instruction, writing is a technology of relatively recent history that must be taught to each generation of children. Historical accounts of the evolution of writing systems have until recently concentrated on a single aspect, increased efficiency, with the Greek invention of the alphabet being regarded as the culmination of a long historical evolution. This efficiency is a product of a limited and manageable set of graphs that can express the full range of meanings in a language. As Eric Havelock wrote, "At a stroke the Greeks provided a table of elements of linguistic sound not only manageable because of economy, but for the first time in the history of *homo sapiens,* also accurate." Ignace Gelb distinguished four stages in this evolution, beginning with picture writing, which expressed ideas directly; followed by word-based writing systems;

then by sound-based syllabic writing systems, including unvocalized syllabaries or consonantal systems; and concluding with the Greek invention of the alphabet.

The alphabet

The invention of the alphabet is a major achievement of Western culture. It is also unique; the alphabet was invented only once, though it has been borrowed by many cultures. It is a model of analytic thinking, breaking down perceptible qualities like syllables into more basic constituents. And because it is capable of conveying subtle differences in meaning, it has come to be used for the expression of a great many of the functions served by speech. The alphabet requires little of the reader beyond familiarity with its orthography. It allows the reader to decipher words newly encountered and permits the invention of spellings for new patterns of sound, including proper names (a problem that is formidable for nonalphabetic systems). Finally, its explicitness permits readers to make a relatively sharp distinction between the tasks of deciphering and interpreting. Less explicit orthographies require the reader first to grasp the meaning of a passage as a whole in order to decide which of several possible word meanings a particular graphic string represents.

It must be remembered, however, that efficiency depends not only on the nature of the writing system but also on the functions required of it by its users, for orthographies are invented to serve particular cultural purposes. Furthermore, an orthography invented to satisfy one purpose may acquire new applications. For instance, writing systems invented to serve mnemonic purposes were subsequently elaborated and used for communicative and archival purposes. Orthographies were not invented as art forms but once invented could serve aesthetic functions.

Notions of explicitness of representation depend on the morphophonemic structure of the language. An alphabet was a notable advance for representing the Greek language but not necessarily for representing a Semitic language. Moreover, for languages such as Chinese and Japanese, which have simple syllabic structure and a great number of homophones, a writing system that depended on phonological structure, such as a syllabary or an alphabet, would be extremely inefficient. It is with such factors in mind that more recent accounts of writing systems have stressed how many different orthographies may function efficiently, given the particular language they are used to represent. Just as linguists have abandoned the notion of progressive evolution of languages, with some languages ranking as more primitive than others, so historians of writing have come to treat existing orthographies as appropriate to the languages they represent.

Nonetheless, all contemporary orthographies have a history of development, and there are many common features in these histories. It is unlikely that writing was invented only once and then borrowed by different cultural groups. While all Western writing systems may be traced back to the beginnings of symbol-making in Sumer, there is no reason to believe that Oriental writing systems were borrowed from the Sumerian form. Consequently, there are two quite separate histories of writing, that of the writing system developed by the Sumerians and that of the one developed by the Chinese.

Sumerian writing. The outline of the development of the Sumerian writing system has been worked out by paleographers. It has long been known that the earliest writing system in the world was Sumerian script, which in its later stages was known as cuneiform. The earliest stages of development are still a matter of much speculation based on fragmentary evidence. The French-American archaeologist Denise Schmandt-Besserat, building on a hypothesis advanced by Pierre Amiet of the Louvre, has demonstrated a series of small steps leading from the use of tokens for simple bookkeeping purposes to the development of written tablets on which graphs of the script stand for morphemes of spoken Sumerian. Archaeologists have discovered in lower Mesopotamia (now southern Iraq) large numbers of small, distinctively shaped clay objects. These are thought to date back to as early as 8000 BC, about the time that hunter-gatherer societies were giving way to an agricultural way of life. A greatly elaborated set of these clay shapes, some shaped like jars and some like various animals, and occasionally inserted in clay envelopes, date from 3500 BC, about the time of the rise of cities. Some of the envelopes have markings corresponding to the clay shapes inside. Moreover, these markings are more or less similar to the shapes drawn on clay tablets that date back to about 3100 BC and that are unambiguously related to the Sumerian language. These markings are thought to constitute a logographic form of writing consisting of some 1,200 different characters representing numerals, names, and such material objects as cloth and cow.

The first writing

The theory advanced to explain this transformation by Schmandt-Besserat is that the clay shapes are tokens representing agricultural goods such as grain, sheep, and cattle and that they were used as a form of bookkeeping. The multiplication of types of tokens could correspond to the increase in the number of kinds of goods that were exchanged with the rise of urbanization in the 4th millennium BC. Tokens placed in an envelope might have constituted a sort of "bill of lading" or a record of indebtedness. To serve as a reminder of the contents of the envelope, so that every reader would not need to break open the envelope to read the contents, corresponding shapes were impressed upon the envelope. But if the content was marked on the envelope, there was no need to put the tokens in an envelope at all; the envelope could be flattened into a convenient surface and the shapes impressed on it. Now that there was no need for the tokens at all, their message was simply inscribed into the clay. These shapes, drawn in the wet clay with a reed stylus or pointed stick, constituted the first writing.

The historical record is much more explicit after 3200 BC and reveals clearly the stages involved in the evolution from a limited system of notation suitable for recording particular events into a full, general-purpose orthography. Archaic Sumerian used mostly graphs representing numerals, names for objects, and names of persons. Graphs for numerals were geometric shapes, while those for objects were often stylized pictures of the things they represented. Yet the system was a genuine logographic writing system generally adequate to economic and administrative purposes (see below *Systems of writing: Cuneiform*). With the substitution of a blunt writing stylus for a pointed one, the symbols become less picturelike and more conventionalized. The writing system takes the name cuneiform from the shape of the strokes that form the symbols (from Latin *cuneus,* "wedge").

The next major stage in the evolution of Sumerian writing was the adoption of the phonographic principle, the use of a sign to represent a common sound rather than a common meaning. The graph representing "water" appears to have been used also to represent the locative suffix "in" because the latter sounded the same as, or similar to, the word "water." It is as if in English a person used the word for ball to stand for the person named Bill on the grounds that it is easy to represent the ball with a circular graph while there is no obvious way to represent Bill, and the two words sound similar. The Sumerian script, however, remained primarily logographic and resorted to phonographic signs only when forced to, for representing unpicturable words and for distinguishing ambiguous graphs. Sumerian script was adopted in the 3rd millennium BC by the Akkadians, who greatly expanded the phonographic properties of the script. The Assyrians and the Babylonians, both speaking dialects of the Akkadian language, were responsible for most of the cuneiform writing in a form known today as Akkadian cuneiform.

While cuneiform had many graphs that represented syllables, many syllables were not represented. The methods used for representing syllables that did not have distinctive graphs were quite unsystematic. The first writing system consistently based on the sound structure of a language was Linear B, a Mycenaean Greek orthography developed around 1400 BC and deciphered in modern times by an English architect, Michael Ventris, in 1952. The script is strictly syllabic; each consonant-vowel pair is given a distinctive graph. As an example, a set of syllables that an alphabetic system would represent with the consonant *p* plus a vowel are all represented in Linear B by different graphs. Although the script is highly systematic, it provides

Linear B

a limited representation of the phonology of Mycenaean Greek. Greek contains many syllables that are not simple consonant-vowel combinations, and not all consonantal sounds are followed by vowels. Linear B is thus an incomplete script for representing the phonological structures of the spoken language. Hence, there are usually several ways of reading a series of Linear B graphs, and a correct reading depends upon the reader's knowing what the text is about.

The final stage in the evolution of writing systems was the discovery of the alphabetic principle, the procedure of breaking the syllable into its constituent consonantal and vowel sounds. According to Geoffrey Sampson, the British linguist, "Most, and probably all, 'alphabetic' scripts derive from a single ancestor: the Semitic alphabet, created sometime in the 2nd millennium [BC]." The Semitic script was invented by speakers of some Semitic language, possibly Phoenician, who lived in the northern part of the Fertile Crescent. Modern versions of Semitic script include the Hebrew script and the Arabic script. Their most prominent characteristic is that they have graphs for consonants but not for vowels.

The inventors of the Semitic orthography apparently took the acrophonic principle, that of representing sounds by pictures of things whose names begin with that sound, from Egyptian hieroglyphic, a form of writing not different in principle from Akkadian cuneiform. The hieroglyphic sign , depicting waves of water, represented the sound /n/, the first sound of the spoken word for water. By means of this principle a 22-graph system was constructed with a memorized order, beginning *alef, bet, gimel,* that was suitable for representing a full range of meanings. These graphs represented the consonants of the language, vowels remaining unrepresented. This fact has led some scholars, notably Gelb and Havelock, to claim that Semitic scripts are not true alphabets but rather unvocalized syllabaries. Other scholars, noting that the graphs represent consonants rather than syllables—for example, *pa, pe, pi, po,* and *pu* would all be represented by the same character—insist that the script is an alphabet. The controversy is circumvented by referring to Semitic scripts, following Sampson, as consonantal writing systems. While such a script would be greatly limited in explicitness or completeness for a language with complex syllable structure such as English, it is relatively complete for Semitic languages in which vowel differences are rarely contrastive.

To illustrate, the following oral forms have in common the three consonantal phonemes /k/, /t/, and /b/ with different vowel sounds interdigitated. The meanings all contain the root meaning "write," and the vowel differences mark subject, tense, and aspect: *katab* "he wrote," *katabi* "I wrote," *katebu* "they wrote," *ketob* "write," *koteb* "writing," *katub* "being written." All are written simply *ktb.*

Because vowel sounds generally distinguish grammatical rather than lexical meaning, some Semitic writing systems never developed any device for representing them. This is not necessarily a flaw in the orthography. Indicating the vowels could cause some confusion for the reader because instead of a single root there would now be a multiplicity of written words, each reflecting a particular grammatical context. Nonetheless, ignoring the vowels does result in an orthography that is far from explicit or complete; many ambiguities in decoding remain. Consequently, some scripts, such as Hebrew, added *matres lectionis,* literally, "mothers of reading," a pointing system to distinguish the vowel sounds. These were used especially for preserving the precise reading of sacred texts. To this day they are used in books written to be read by beginning readers and in poetry and other writings of which the prior knowledge of the reader may not be sufficient to reduce the residual ambiguity.

The transition from consonantal writing to alphabetic writing, writing with full representation of both consonants and vowels, occurred when the Semitic script was adapted to the Greek language. This occurred about 1000–900 BC. Scholars have traditionally considered the Greek invention as a stroke of genius. While not minimizing the significance of the Greek invention, it is now recognized that the invention of the alphabet was, in fact, the rather straightforward consequence of applying a script invented for representing one kind of language to a quite different kind.

Greek alphabet

The letters used by the Greeks to represent consonantal sounds were borrowed rather directly from the Semitic script. What was distinctive was that the Greeks used six of the Semitic letters, those that represented sounds that did not occur in Greek, to represent vowel sounds. Greek, like English, is an Indo-European language that uses vowel distinctions to make lexical contrasts. Moreover, words may consist simply of vowels, words may begin with vowels, and words with adjacent vowels are not uncommon. Such forms are rare in Semitic languages in which simple consonant-vowel syllable structures predominate and in which vowel differences usually mark only grammatical inflections. Sampson has suggested that in the Semitic language some of the consonants that preceded a vowel sound may have been nonphonemic to the Greeks, who thus in hearing the syllable would have heard only a vowel corresponding to a vowel already prominent in the Greek language.

The Romans borrowed the Greek alphabet (along with many Greek words and much of Greek culture) to form the Roman, or Latin, alphabet. Written "learned" Latin was the language of state and of scholarship in Europe until the end of the Middle Ages. Further developments of the alphabet resulted from changes in the phonology of Latin and of the Romance languages that evolved from it. For English, the differentiation of all the 26 letters was completed only in the 19th century.

While the invention of logographic writing, the later invention of the principle of phonetization, the analysis of syllables into a consonantal writing system, and the addition of vowels to make a full alphabet do constitute progress toward an efficient, economical, explicit, and complete writing system, this progress was not simply a matter of increasing insight. Advances resulted from attempts to apply a writing system invented for one language to another language for which it was not completely appropriate. Yet the accumulated discoveries yielded an analysis of deeper and deeper levels of linguistic structure of the type associated with discoveries in the natural sciences. For this reason, writing has almost always been the means not only for transcribing speech but also for uncovering its underlying structure. That is, to a large extent writing is what has made people conscious of the properties of speech.

Observation of children learning to read and write an alphabetic orthography suggests that children pass through some of the same stages in interpreting the code that the writing system itself passed through in the course of its development. The youngest child's hypothesis about writing is that words must be similar in some way to the objects they represent. Thus, at the earliest stage children think that the word "train" must be represented by a long word because it is a long thing. Similarly, they think that "two little pigs" must be represented by two words, one for each pig, and so on. Later, they invent the hypotheses that writing represents words rather than things and that these words are series of sounds. At this point children may write the word with a series of consonants; cat becomes *kt.* Only later do they recognize the alphabetic principle that words must be written with both consonants and vowels.

Yet the evolution of the alphabet, an invention of enormous importance for Greek and for all Indo-European languages, was of little use for Semitic languages, in which the vowels played a smaller role than in Greek. And it was of no use at all for Chinese, which is a monosyllabic language with a great many homophones.

Chinese writing. Chinese is a language with clearly distinguished syllables, each of which corresponds to a meaningful unit, a morpheme. As it is an "isolating" language, rather than an inflected language like Latin or, to a lesser degree, English, each morpheme is represented separately by a separate syllable. Whereas in English one word, for example, *make,* yields, when inflected, a family of related words (*make, makes, making, made,* etc.), in Chinese one character would represent one morpheme (*e.g., make*). Because each morpheme is represented by a different

character, and because the number of morphemes in a language is far larger than the number of syllables, such a writing system needs an extremely large number of characters or graphs.

It is not known when Chinese writing originated. The earliest known samples are from the time of the Shang dynasty (18th–12th century BC), but by then it was already a highly developed system, essentially similar to its present form. By 1400 BC the script included some 2,500 to 3,000 characters, most of which can be read to this day. The script was fixed in its present form during the Ch'in period (221–206 BC). The earliest graphs were schematic pictures of what they represented; the graph for "man" resembled a standing man, that for "woman" depicted a kneeling figure, perhaps doing housework.

Because basic characters or graphs were "motivated," that is, the graph was made to resemble the object it represented, it has sometimes been concluded that Chinese writing is ideographic, representing ideas rather than the structures of a language. It is now recognized that the system represents the Chinese language by means of a logographic script. Each graph or character corresponds to one meaningful unit of the language, not directly to a unit of thought.

Although it was possible to make up simple signs representing common objects, many words were not readily picturable. To represent such words, the phonographic principle was adopted. A graph that pictured some object was borrowed to write a different word that happened to sound similar. With this invention the Chinese approached the form of writing invented by the Sumerians. However, because of the enormous number of Chinese words that sound the same, to have carried through the phonographic principle would have resulted in a writing system in which many of the words could be read in more than one way. That is, a written character would be enormously ambiguous.

The solution to the problem of ambiguity of characters, a solution adopted about 213 BC, was to distinguish two words having the same sound and represented by the same graph by adding another graph to give a clue to the meaning of the particular word intended. Such complex graphs or characters consist of two parts, one part suggesting the sound, the other part the meaning. The system was then standardized so as to approach the ideal of one distinctive graph representing each morpheme in the language. The limitation is that a language that has thousands of morphemes would require thousands of characters. As the characters are formed from simple lines in various orientations and arrangements, they came to possess great complexity.

Not only did the principle of the script change with time, so too did the form of the graphs. The earliest writing consisted of carved inscriptions. Before the beginning of the Christian Era, the script came to be written with brush and ink on paper. The result was that the shapes of the graphs lost their pictorial, "motivated" quality. The brushwork allowed a great deal of scope for aesthetic considerations.

The relation between the written Chinese language and its oral form is very different from the analogous relation between written and spoken English. In Chinese many different words are expressed by the identical sound pattern—188 different words are expressed by the syllable /yi/—while each of those words is expressed by a distinctive visual pattern. A piece of written text read orally to another person is often quite incomprehensible because of the large number of homophones. In conversation literate Chinese speakers frequently draw characters in the air to distinguish between homophones. Written text, on the other hand, is completely unambiguous. In English, by contrast, writing is often thought of as a reflection, albeit imperfect, of speech.

To make the script easier to read, a system of transcribing Chinese into the Roman alphabet was adopted in 1958. The system was not intended to replace the logographic script but to indicate the sounds of graphs in dictionaries and to supplement graphs on such things as road signs and posters. A second reform simplified the characters by reducing the number of strokes used in writing them. Simplification, however, tends to make the characters more similar in appearance; thus they are more easily confused, and the value of the reform is limited.

Most scholars now believe that neither the logographic Chinese writing system nor the alphabetic writing system of Indo-European possesses any overall advantage. The Chinese system requires more memorization, while the alphabet requires more analysis and synthesis; both appear to be relatively optimal devices for the transcription of their respective, very different, languages.

A second factor makes the use of a logographic system particularly significant for Chinese. A single logographic system is capable of representing very different spoken forms. As was mentioned earlier, the numerals *1, 2,* and *3* are logograms that represent different words in a number of different languages.

Chinese logographs form a common medium of communication for a vast nation because they can be read by people speaking mutually incomprehensible dialects of Chinese. Since the Communist Revolution the grammar and vocabulary of modern Mandarin Chinese has served as the standard written language.

Japanese writing. The Japanese came into contact with Chinese culture during the Chinese Han dynasty (206 BC–AD 220), and they began to write their own language in the 5th century AD, basing their writing system on the Chinese model. But the two languages are fundamentally different in structure; whereas Chinese words are monosyllables, Japanese words often consist of several syllables, and whereas Chinese is an isolating language, Japanese is an inflected language. To write such a language, the Japanese developed a mixed system, partly logographic, based on the Chinese system, and partly syllabic, using the same characters in a second way for their sound values. In *kun* writing Chinese characters were used to represent Japanese words that have a similar meaning, while other characters were adopted to represent sounds.

In the 8th century the phonographic principle was applied more systematically in a writing system called *man'yōgana,* a syllabary very similar in form to the Semitic alphabet. However, given the large number of homophones and the fact that *man'yogana* was combined with *kun* writing, it was almost impossible to establish a single correct reading of a text. Indeed, scribes took pride in being able to read the same text in various ways.

In the 9th or 10th century two sets of syllabic signs evolved, one called *hiragana,* or "plain" *kana,* which consists of simplified outlines, written cursively, of Chinese characters, the other called *katakana,* or "partial" *kana,* which consists of carefully written parts of the original Chinese characters. Writing with the full Chinese characters is called *kanji.* The two sets of *kana* characters are limited as are other syllabaries in that they are not unambiguous; *kanji* are unambiguous but are very complex visually. Consequently, modern Japanese writing uses a combination of characters from all three of these systems. In 1946 a standardizing reform established a limited list of 1,850 *kanji* (enlarged to 1,945 in 1981) and encouraged the use of *kana* for all other words. Modern written Japanese uses many more *hiragana* graphs than *kanji* in a piece of text.

Even with modern reforms, written Japanese is difficult to read unambiguously because of the great degree of homophony in the vocabulary. The word *kan,* for example, means "sweet," "be affected," "print," "be accustomed to," "view," "investigate," "slow," "tube," "enjoy," "a volume," "Chinese," and "Korean," among others. As a result a reader must know rather precisely what is being discussed in order to read a text accurately. Poetry, in particular, takes quite a different form in Japanese than in Indo-European languages.

Korean writing. Korea, too, adopted its institutions and culture from the Chinese. Until the 20th century the normal medium of written communication was in Chinese, using the Chinese writing system. But beginning in about the 6th century the Chinese script was adapted to write Korean. The application of Chinese script to the Korean language created problems almost identical to those that arose in using Chinese to write the Japanese language. Yet

the borrowed *kanji* script continues to be used for some purposes to this day. The most remarkable development in Korean writing was the invention of Han'gŭl by King Sejong in 1446. It is a featural script consisting of some 28 letters that have a systematic visual structure directly related to the phonetic features of the phonemes. This writing system owes nothing to the Chinese orthography (see above *Types of writing systems*).

Because the principles employed by various writing systems vary greatly and because the languages they represent are organized so differently, it is difficult to state any general principles of the evolution of writing systems. Yet it appears that they all began with motivated pictorial signs representing objects. To turn such signs into a general orthography required the recognition that the signs must represent sound patterns and the consequent invention of the phonographic principle. Depending on the language, such sound-based systems developed in two directions. Western scripts went furthest in the phonographic direction, representing words by means of syllables and syllables by means of consonantal writing systems and eventually developing a full vocalic alphabet. Eastern scripts preserved the logographic principle even though some of the logographs were sound-based; each word was represented by a distinctive visual character. Only one practical orthography, Korean, adopted a featural system, and that invention bore little or no relation to neighbouring orthographies.

LITERACY: THE USES OF WRITING

The rise of literacy. The invention of devices for representing language is inextricably related to issues of literacy; that is, to issues of who can use the script and what it can be used for. Competence with written language, both in reading and writing, is known as literacy. High levels of literacy are required for using scripts for a wide range of somewhat specialized functions. When a large number of individuals in a society are competent in using written language to serve these functions, the whole society may be referred to as a literate society.

Just as scripts have a history, so, too, does literacy have a history. This history closely reflects the increasing number of ways in which written materials have been used and the increasing number of readers who have been able to use them. Scripts were elaborated to serve new purposes; more importantly, new kinds of writing systems permitted them to serve a wider range of purposes by a larger number of individuals.

Learnability and expressive power

Although the uses of writing reflect a host of religious, political, and social factors and hence are not determined simply by orthography, two dimensions of the script are important in understanding the growth of literacy: learnability and expressive power. Learnability refers to the ease with which the script can be acquired, and expressive power refers to the resources of the script for unambiguously expressing the full range of meanings available in the oral language. These two dimensions are inversely related to each other. Simple, restricted scripts are readily learned. Pictographic signs such as those used in "environmental writing" and logographic scripts with a limited set of characters are easiest to learn and, indeed, are acquired more or less automatically by children. Syllabaries such as the Cree syllabary are reported to be learnable in a day, while the indigenous Liberian Vai syllabary is learned in a few days. Consonantal scripts and alphabets are difficult to learn and usually require a few years of schooling. Full logographic systems, such as Chinese, or mixed systems, such as Japanese, are difficult to acquire because they require the memorization of thousands of distinctive characters. Once learned, however, they appear to function as well as alphabets.

But pictographic signs and logographic scripts with a limited, readily learnable set of graphs are restricted to expressing a limited range of meanings. Syllabaries are highly ambiguous and hence dependent on knowledge not only of the script but also on the likely content of the message. Syllabaries therefore serve a restricted set of functions, primarily personal correspondence. They are of limited use in expressing novel meanings that could be read in the same way by all readers of the script. Consonantal and alphabetic writing systems can express essentially all the lexical and grammatical meanings in the language (but not the intonation) and are thus highly suitable for the expression of original meanings. They constitute an ideal medium for technical, legal, literary, and scientific texts that must be read in the same way by readers dispersed in both time and space. Some scholars have held that the high degree of literacy in the West is a consequence of the optimality of the alphabet in balancing the two dimensions of learnability and expressive power. Such generalizations, however, ignore the fact that the "optimal" balance may differ from language to language. A consonantal writing system is almost as complete for Hebrew as the alphabet is for Greek, but a consonantal writing system would be hopelessly ambiguous for Greek. Similarly, a syllabary or an alphabet would be quite useless for Chinese, a language with a staggering degree of homophony. Logographic systems achieve a comparable level of explicitness by the addition of new characters, but the ease of addition is traded off against the ease of acquisition. Instead of attempting to determine whether one system is better than another, it is perhaps more reasonable to assume that each script is optimal for the language it represents and for the functions it has evolved to serve.

The ease of acquisition of a script is an important factor in determining whether a script remains the possession of an elite or whether it can be democratized, that is, turned into a possession of ordinary people. Syllabaries are readily learned, but their residual ambiguity tends to restrict their uses. Alphabets have been viewed by many historians as decisive in the democratization of writing; alphabetic writing could become a possession of ordinary people and yet serve a full range of functions. However, democratization of a script appears to have more to do with the availability of reading materials and of instruction in reading and the perceived relevance of literacy skills to the readers. Even in a literate society, most readers learn to read only a narrow range of written materials; specialized materials, such as those pertaining to science or government, remain the domain of elites who have acquired additional education.

Functions of a script

The second factor determining the social breadth of the use of writing is the range of functions that a script serves. The functions served are directly related to the orthography. Early forms of writing served an extremely narrow range of functions and were wholly unsuitable for others. While tokens served for simple record keeping, and early Sumerian writing was useful for a range of administrative purposes, a relatively complete script is required for writing histories, edicts, treaties, and scientific and literary works that, to be useful, must be read in the same way by all readers. Considerable scholarly controversy surrounds the question of the role of the invention of more complete or explicit scripts, such as the alphabet, in the evolution of these more specialized uses of language. If the alphabet were decisive, one could look for the basis of many of the particular features of Western culture in the invention of an alphabetic orthography.

This question is far from resolved. Historically, the rise of cities coincided with the development of a script suitable for serving bureaucratic purposes. Later, the scientific and philosophical tradition that originated in classical Greece and that prevails in the West to this day developed along with the alphabet. Many writers, including Eric Havelock, have maintained that the alphabet was a decisive factor in the cultural development of the West. Marshall McLuhan and Walter Ong have claimed that the rise of literacy and the decline of "orality" in the later Middle Ages were fundamental to the cultural flowering known as the Renaissance.

It is perhaps characteristic of alphabet-based conceptions of literacy to draw a strict distinction between reading and interpreting. As interpretation came to be seen as interpolation into or distortion of the text, the attempt was made to write texts in such a manner as to reduce the possibility of variant interpretations. This resulted in the attempt to write texts with univocal meanings, texts that mean neither more nor less than what they say. To achieve this required the formalization of grammatical structures, the conventionalization of meanings of terms, and the inven-

tion of standard punctuation. Such textual developments were especially important for the specialized functions of science and philosophy. The distinction between meaning and interpretation fostered the idea that texts have a literal meaning, that knowledge can be completely expressed by means of such literal meanings, and that texts can be autonomous and objective. In the Western tradition, knowledge is treated as if it were an ideal text, as something that is regarded by most learners as given rather than created. These assumptions about meaning were important to both the literary and the scientific traditions that took form in western Europe in the 17th century and that continue to this day.

The particular form of writing, whether logographic, syllabic, or alphabetic, is less important than the existence of some form that is general enough to serve a full range of purposes. Literate societies, whether Chinese or Sumerian, have always been esteemed by nonliterate societies, which have borrowed heavily from them. Thus the Romans borrowed Greek literacy, and the Japanese and Koreans borrowed Chinese literacy. Once adopted and used for administrative, scientific, legal, and literary purposes, literacy altered the society that it was part of in a variety of ways.

Writing allows exactly repeatable statements to be circulated widely and preserved. It allows readers to scan a text back and forth and to study, compare, and interpret at their leisure. It allows writers to deliberate over word choice and to construct lists, tables, recipes, and indexes. It fosters an objectified sense of time, a linear conception of space. It separates the message from the author and from the context in which it was written, thereby "decontextualizing," or universalizing the meaning of, language. It allows the creation of new forms of verbal structure, like the syllogism, and of numerical structures, like the multiplication table. When writing becomes a predominant institutional and archival form it has contributed to the replacement of myth by history and the replacement of magic by skepticism and science. Writing has permitted the development of extensive bureaucracy, accounting, and legal systems organized on the basis of explicit rules and procedures. Writing has replaced face-to-face governance with written law and depersonalized administrative procedures. And, on the other hand, it has turned writers from scribes into authors and thereby contributed to the recognition of the importance of the thoughts of individuals and consequently to the development of individualism.

Literacy and schooling. Whereas oral language is learned quite independently of whether it is taught or not, literacy is largely dependent upon teaching. While some local or indigenous scripts are taught relatively informally by parents or someone who knows the script well, widespread or universal literacy is dependent upon schooling. Indeed, in many societies schooling and literacy have been almost synonymous. Schools in such diverse places as Sumer and China developed concurrently with the development of a full writing system and were concerned primarily with teaching first adults and later children to read and write. And it is inconceivable that modern, technological societies could survive without schools to develop high levels of literacy.

Although schooling is critical to the development of literacy, it is not, by itself, sufficient. Historians have shown that the level of literacy produced by the schools of any society is directly tied to the functions and levels of literacy in the society as a whole. Consequently, it is unrealistic to expect that a modern, literate society could be created simply through establishing schools and teaching children to read. Schools tend to reflect the society rather than to change it dramatically. Schooling in Western societies is successful in achieving relatively high levels of literacy in part because of the literacy practices in the larger society. When compulsory schooling was introduced in Britain, Europe, and America in the 19th century, it was nurtured by an environment of "lay" literacy in which as much as 75 percent of the population could use written materials for such informal purposes as keeping diaries, reading and writing notes and letters, and personal record keeping. Such a climate of widespread practical literacy is important to the effectiveness of schooling. The relation between literate practices in the home and the level of literacy achieved by children in the school has been amply documented.

It is common to think of literacy as the simple ability to read and write. In part such thinking is a consequence of the naive assumption that alphabetic literacy is a matter simply of decoding graphs into sounds and vice versa. In fact, literacy involves competence in reading, writing, and interpreting texts of various sorts. It involves both skill in decoding and higher levels of comprehension and interpretation. These higher levels depend upon knowledge both of specialized uses of language and of specialized bodies of knowledge. The intimate relations among language, literacy, and specialized bodies of knowledge have contributed to the identification of literacy with schooling.

As different scripts serve different functions and make different demands upon readers, it is a complex matter to define literacy in universal terms and so to judge the literacy levels of a society at different periods or to compare one society with another. Scripts that, because of incompleteness or inexplicitness, rely heavily upon the prior knowledge of reader and writer remain the domain of a specialized elite, as did cuneiform, or they are used for rather restricted purposes, as is Cree syllabic. Scripts that are relatively explicit and complete permit a reader who is unfamiliar with a text to read it in a reliable way and hence can be used for a much broader range of functions.

The form of the script may be less crucial than the range of functions a script serves and the breadth of its readership, that is, the degree of literacy of the society. With the growth of readership come increased production of materials to be read, increased number of social functions the script is used for, and the invention of new, more specialized genres of writing. The novel form, for example, was invented in Europe only in the 17th century, when there was a broadly based reading public. Other specialized uses of writing developed much earlier. As European societies became more literate during the Middle Ages, writing came to be used for functions that earlier had been performed by oral language and by ritual. Indenture of servants, deeding of property, evidence at trials, and accounts of the lives of saints all came to be functions of written texts. As literacy began to be required for these vital social purposes, oral language came to be seen as loose and unruly and lacking in social authority. And people who could not read and write came to be regarded as rude and ignorant—in short, unlettered.

Rise of literacy in Europe

Rising levels of literacy in Europe were closely related to great social transformations, notably the Protestant Reformation and the rise of modern science. The right to read the Bible for oneself and to discover its meaning was the fundamental tenet of Protestantism, and the private study and verification of written texts was important to science. Both of these functions were enormously facilitated by the development of printing from movable type and by the translation of important books from scholarly Latin into vernacular languages. With the increase in the uses of writing and the spread of printing there were more texts to read. Concurrently, European society as a whole became more literate in two ways: more individuals learned to read and write their native tongues, and even those who could not themselves read and write came to rely upon written documents as loci of authority and significance. In the 18th and 19th centuries in western Europe and in America, even before the establishment of compulsory schooling, more than half of the population had some competence in reading and writing. Compulsory schooling had, by the end of the 19th century, made some level of literacy more or less universal.

Partly because of the close tie between schooling and literacy, literacy levels are often defined exclusively in terms of the number of years that a person has attended school. Educational institutions usually differentiate a basic, or functional, level of literacy, roughly equivalent to six years of schooling, from a high level of literacy, a level of competence roughly equivalent to 10 to 12 years of schooling. Such categorical distinctions have been criticized because they are insensitive to the diversity of particular uses of literacy in even a literate society and the irrelevance of

the school to many of them. Many people incapable of or uninterested in reading continuous texts pertaining to science and literature nonetheless read menus, catalogs, letters, labels, warnings, invoices, and a range of other materials of relevance and interest to them.

Moreover, literacy levels are judged against a sliding standard. The more literate the society becomes, the higher a standard of literacy is judged as functional. In Sweden in the 17th century a person was judged as literate, and allowed to marry, if he could read bits of the catechism and sign the church registry. In the United States at the time of World War II, when soldiers were screened for military service the army defined a minimal level of literacy as that normally achieved in the fifth grade (about 10 years of age). By 1966 the criterion of functional literacy in the United States had been raised to completion of secondary school by the Adult Education Act passed by Congress in that year.

Using this criterion, some writers have claimed that 25 percent of U.S. adults are functionally illiterate. Some commentators see in such figures a social problem of great importance and promote various programs of educational reform intended to produce higher levels of literacy. However, most scholars criticize such statistics as meaningless on two grounds. First, they are based on a questionable identification of competence with success in a single institution, the school, rather than in the relevant contexts of application. Second, they do not adequately reflect the extent to which even those individuals who are classified as functionally illiterate depend upon and participate in literate activities in modern societies. Such persons know how to participate in a great many literacy-based institutions—how to read signs, labels, and letters, how to deal with ballots, how to sign checks and write notes—if not the special literate skills usually acquired in school. More important, even if they are not highly skilled in literate activities they know what it is to be literate, what texts are, how they are written and interpreted, how they accumulate to form a tradition, and how they are consulted and used in multiple ways in a literate society.

As an alternative to simply identifying levels of literacy with years of schooling, some scholars have distinguished levels of literacy in another way. Environmental or lay literacy is the term used to designate that form of unspecialized competence involved in generally dealing with a literate environment. Such literacy need never be taught. It is a type of literacy that is acquired through participating in a literate environment in which written signs, labels, trademarks, headlines, sports scores, and the like are ubiquitous. Such a general, if low, level of literacy, which stands somewhat apart from the particular skills of reading and writing, first arose in Europe in the later Middle Ages with the development of what the Canadian historian Brian Stock refers to as "textual communities." A textual community consisted of a band of believers formed around an interpreter who read and interpreted religious texts. Because the authority of the teacher rested in the text rather than in the church, members of the community came to know certain general truths about texts and about writing: that they could be read, understood, studied, consulted; that they were more reliable than hearsay; that they were permanent; and that they possessed authority. Everyone in a literate society is literate in this sense; all know the nature, uses, and functions of writing even if they do not personally practice it.

Lay literacy

A literate society is also dependent upon the development of elite literacy, a high level of literate competence, possessed by a relatively small percentage of the population, in such specialized fields of endeavour as science, law, or literature. High levels of literate competence involve learning a somewhat specialized vocabulary as well as the nuances of meaning that are relevant to lexical choice. It is estimated that literate people have a reading vocabulary, consisting of words that are encountered only in reading and writing, that may be more than double the size of their ordinary speaking vocabulary. In addition to specialized vocabularies, high levels of literate competence involve knowledge of specialized grammatical constructions that serve to set out explicitly the logical form of an argument and of specialized genres or literary forms such as description, explanation, argument, and instructions that can be used for building complex linguistic structures or genres, such as narrative and expository texts. These specialized skills require for mastery many years of formal schooling. Once such forms are acquired in literate contexts they can also be used in speech. For this reason literacy is not tied exclusively to writing; just as one can write in an essentially oral style, so one can speak in a manner characteristic of written language. Literacy makes it possible to speak a written language. (D.R.O.)

Elite literacy

Systems of writing

CUNEIFORM

Cuneiform (a coinage from Latin and Middle French roots meaning "wedge-shaped") has been the modern designation, from the early 18th century onward, for the most widespread and historically significant writing system in the ancient Middle East. Its active history comprises the last three millennia BC; its long development and geographic expansion involved numerous successive cultures and languages; and its overall significance as an international graphic medium of civilization is second only to that of the Phoenician-Greek-Latin alphabet.

Origin and character of cuneiform. The origins of cuneiform may be traced back approximately to the end of the 4th millennium BC. At that time the Sumerians, a people of unknown ethnic and linguistic affinities, inhabited southern Mesopotamia and the region west of the mouth of the Euphrates known as Chaldea. While it does not follow that they were the earliest inhabitants of the region or the true originators of their system of writing, it is to them that the first attested traces of cuneiform writing are conclusively assigned. The earliest written records in the Sumerian language are pictographic tablets from Uruk (Erech), evidently lists or ledgers of commodities identified by drawings of the objects and accompanied by numerals and personal names. Such word writing was able to express only the basic ideas of concrete objects. Numerical notions were easily rendered by the repetitional use of strokes or circles. However, the representation of proper names, for example, necessitated an early recourse to the rebus principle—*i.e.*, the use of pictographic shapes to evoke in the reader's mind an underlying sound form rather than the basic notion of the drawn object. This brought about a transition from pure word writing to a partial phonetic script. Thus, for example, the picture of a hand came to stand not only for Sumerian *šu* ("hand") but also for the phonetic syllable *šu* in any required context. Sumerian words were largely monosyllabic, so the signs generally denoted syllables, and the resulting mixture is termed a word-syllabic script. The inventory of phonetic symbols henceforth enabled the Sumerians to denote grammatical elements by phonetic complements added to the word signs (logograms or ideograms). Because Sumerian had many identical sounding (homophonous) words, several logograms frequently yielded identical phonetic values and are distinguished in modern transliteration—(as, for example, *ba, bá, bà, ba*$_4$). Because a logogram often represented several related notions with different names (*e.g.*, "sun," "day," "bright"), it was capable of assuming more than one phonetic value (this feature is called polyphony).

Use of the rebus

In the course of the 3rd millennium the writing became successively more cursive, and the pictographs developed into conventionalized linear drawings. Due to the prevalent use of clay tablets as writing material (stone, metal, or wood also were employed occasionally), the linear strokes acquired a wedge-shaped appearance by being pressed into the soft clay with the slanted edge of a stylus. Curving lines disappeared from writing, and the normal order of signs was fixed as running from left to right, without any word-divider. This change from earlier columns running downward entailed turning the signs on one side (see above *History of writing systems*).

Spread and development of cuneiform. Before these developments had been completed, the Sumerian writing system was adopted by the Akkadians, Semitic invaders who established themselves in Mesopotamia about the

middle of the 3rd millennium. In adapting the script to their wholly different language, the Akkadians retained the Sumerian logograms and combinations of logograms for more complex notions but pronounced them as the corresponding Akkadian words. They also kept the phonetic values but extended them far beyond the original Sumerian inventory of simple types (open or closed syllables like *ba* or *ab*). Many more complex syllabic values of Sumerian logograms (of the type *kan, mul, bat*) were transferred to the phonetic level, and polyphony became an increasingly serious complication in Akkadian cuneiform (*e.g.*, the original pictograph for "sun" may be read phonetically as *ud, tam, tú, par, laḫ, ḫiš*). The Akkadian readings of the logograms added new complicated values. Thus the sign for "land" or "mountain range" (originally a picture of three mountain tops) has the phonetic value *kur* on the basis of Sumerian but also *mat* and *šad* from Akkadian *mātu* ("land") and *šadû* ("mountain"). No effort was made until very late to alleviate the resulting confusion, and equivalent "graphies" like *ta-am* and *tam* continued to exist side by side throughout the long history of Akkadian cuneiform.

The earliest type of Semitic cuneiform in Mesopotamia is called the Old Akkadian, seen for example in the inscriptions of the ruler Sargon of Akkad (died *c.* 2279 BC). Sumer, the southernmost part of the country, continued to be a loose agglomeration of independent city-states until it was united by Gudea of Lagash (died *c.* 2124 BC) in a last brief manifestation of specifically Sumerian culture. The political hegemony then passed decisively to the Akkadians, and King Hammurabi of Babylon (died 1750 BC) unified all of southern Mesopotamia. Babylonia thus became the great and influential centre of Mesopotamian culture.

By courtesy of the Oriental Institute, the University of Chicago

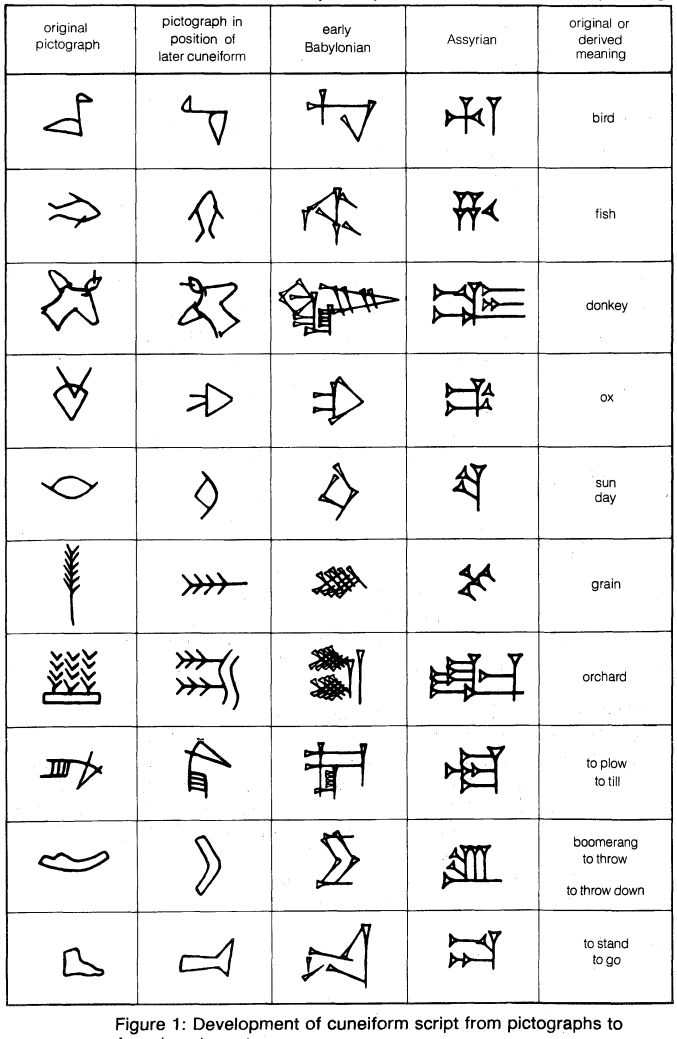


Figure 1: Development of cuneiform script from pictographs to Assyrian characters.

The Code of Hammurabi is written in Old Babylonian cuneiform, which developed throughout the shifting and less brilliant later eras of Babylonian history into Middle and New Babylonian types. Farther north in Mesopotamia the beginnings of Assur were humbler. Specifically Old Assyrian cuneiform is attested mostly in the records of Assyrian trading colonists in central Asia Minor (*c.* 1950 BC; the so-called Cappadocian tablets) and Middle Assyrian in an extensive Law Code and other documents. The Neo-Assyrian period was the great era of Assyrian power, and the writing culminated in the extensive records from the library of Ashurbanipal at Nineveh (*c.* 650 BC).

Assyrian cuneiform

The expansion of cuneiform writing outside Mesopotamia began in the 3rd millennium, when the country of Elam in southwestern Iran was in contact with Mesopotamian culture and adopted the system of writing. The Elamite sideline of cuneiform continued far into the 1st millennium BC, when it presumably provided the Indo-European Persians with the external model for creating a new simplified quasi-alphabetic cuneiform writing for the Old Persian language. The Hurrians in northern Mesopotamia and around the upper stretches of the Euphrates adopted Old Akkadian cuneiform around 2000 BC and passed it on to the Indo-European Hittites, who had invaded central Asia Minor at about that time.

In the 2nd millennium the Akkadian of Babylonia, frequently in somewhat distorted and barbarous varieties, became a lingua franca of international intercourse in the entire Middle East, and cuneiform writing thus became a universal medium of written communication. The political correspondence of the era was conducted almost exclusively in that language and writing. Cuneiform was sometimes adapted, as in the consonantal script of the Canaanite city of Ugarit on the Syrian coast (*c.* 1400 BC), or simply taken over, as in the inscriptions of the kingdom of Urartu or Haldi in the Armenian mountains from the 9th to 6th centuries BC; the language is remotely related to Hurrian, and the script is a borrowed variety of Neo-Assyrian cuneiform. Even after the fall of the Assyrian and Babylonian kingdoms in the 7th and 6th centuries BC, when Aramaic had become the general popular language, rather decadent varieties of Late Babylonian and Assyrian survived as written languages in cuneiform almost down to the time of Christ.

Decipherment of cuneiform. Many of the cultures employing cuneiform (Hurrian, Hittite, Urartian) disappeared one by one, and their written records fell into oblivion. The same fate overtook cuneiform generally with astonishing swiftness and completeness. One of the reasons was the victorious progress of the Phoenician script in the western sections of the Middle East and the classical lands in Mediterranean Europe. To this writing system of superior efficiency and economy, cuneiform could not offer serious competition. Its international prestige of the 2nd millennium had been exhausted by 500 BC, and Mesopotamia had become a Persian dependency. Late Babylonian and Assyrian were little but moribund artificial literary idioms. So complete was the disappearance of cuneiform that the classical Greeks were practically unaware of its existence, except for the widely traveled Herodotus, who in passing mentions *Assyria Grammata* ("Assyrian characters").

Old Persian and Elamite. The rediscovery of the materials and the reconquest of the recondite scripts and languages have been the achievements of modern times. Paradoxically the process began with the last secondary offshoot of cuneiform proper, the inscriptions of the Achaemenid kings (6th to 4th centuries BC) of Persia. This is understandable, because almost only among the Persians was cuneiform used primarily for monumental writing, and the remains (such as rock carvings) were in many cases readily accessible. Scattered examples of Old Persian inscriptions were reported back to Europe by western travelers in Persia since the 17th century, and the name cuneiform was first applied to the script by Engelbert Kämpfer (*c.* 1700). During the 18th century many new inscriptions were reported; especially important were those copied by Carsten Niebuhr at the old capital Persepolis. It was recognized that the typical royal inscriptions contained three different scripts, a simple type with about

Persian inscriptions

40 different signs and two others with considerably greater variations. The first was likely to reflect an alphabet, while the others seemed to be syllabaries or word writings. Assuming identical contents in three different languages, scholars argued on historical grounds that those trilingual inscriptions belonged to the Achaemenid kings and that the first writing represented the Old Persian language, which would be closely related to Avestan and Sanskrit. The recognition of a diagonal wedge as word-divider simplified the segmentation of the written sequences. The German scholar Georg Friedrich Grotefend in 1802 reasoned that the introductory lines of the text were likely to contain the name, titles, and genealogy of the ruler, the pattern for which was known from later Middle Iranian inscriptions in an adapted Aramaic (*i.e.*, ultimately Phoenician) alphabet. From such beginnings, he was eventually able to read several long proper names and to determine a number of sound values. The initial results of Grotefend were expanded and refined by other scholars.

Next the second script of the trilinguals was attacked. It contained more than 100 different signs and was thus likely to be a syllabary. Mainly by applying the sound values of the Old Persian proper names to appropriate correspondences, a number of signs were gradually determined and some insight gained into the language itself, which is New Elamite; the study of it has been rather stagnant, and considerable obscurity persists. The same holds true for the Old Elamite of the late 2nd millennium.

Akkadian and Sumerian. The third script of the Achaemenian trilinguals had in the meantime been identified with that of the texts found in very large numbers in Mesopotamia, which obviously contained the central language of cuneiform culture, namely Akkadian. Here also the proper names provided the first concrete clues for a decipherment, but the extreme variety of signs and the peculiar complications of the system raised difficulties which for a time seemed insurmountable. The serious external divergencies between older and newer types of Akkadian cuneiform, the distribution of ideographic and syllabic uses of the signs, the simple (*ba, ab*) and complex (*bat*) values of the syllables, and especially the bewildering polyphony of many notations were only gradually surmised by scholars. Once the Semitic character of the language had been established, the philological science of Assyriology developed rapidly from the closing decades of the 19th century onward, especially because of scholars like Friedrich Delitzsch and, later, Benno Landsberger and Wolfram von Soden.

The Akkadian key

Once Akkadian had been deciphered, the very core of the system was intelligible, and the prototype was provided for the interpretation of other languages in cuneiform. Until the 20th century Sumerian was not definitely recognized as a separate language at all but rather as a special way of noting Akkadian. Even when its independent character was established, the difficulties of interpretation were appalling because of its strange and unrelated structure. After Sumerian finally died out as a living language toward the middle of the 2nd millennium, it lingered on as a cult idiom of Babylonian religion. To facilitate its artificial acquisition by the priesthood, grammatical lists and vocabularies were compiled and numerous religious texts were provided with literal translations into Babylonian. These have facilitated the penetration of unilingual Sumerian texts, and Sumerian studies advanced greatly through the efforts of such scholars as Delitzsch, François Thureau-Dangin, Arno Poebel, Anton Deimel, and Adam Falkenstein.

Hittite and other languages. An important new dimension was added to cuneiform studies in the early years of the 20th century, through the discovery in 1906 of the royal archives of the Hittites at the ancient capital site of Hattusas, near the Turkish village of Boğazköy, east of Ankara. Some years earlier the existence of an Indo-European idiom in some cuneiform letters found in the Egyptian diplomatic archives of the 18th dynasty at Tell el-Amarna had been suspected by Johan Knudtzon. This unlikely surmise was confirmed by Friedrich Hrozný during World War I, when his initial interpretation of the Boğazköy materials proved that the predominant language in the thousands of tablets was that of the Indo-European Hittites, whose rule in central Asia Minor filled most of the 2nd millennium. The tablets offered no serious cryptological problems, being edited in a type of borrowed Akkadian cuneiform. The interpretation of the unknown language was aided by the partial ideographic nature of the script, which revealed elements of meaning independent of linguistic factors. Even more important was a series of bilingual parallel texts, in which the Akkadian versions served as a clue to the analysis of linguistic structure.

In the absence of close affinity to known languages, which vouches adequate safeguards against the notoriously misleading comparative method of interpretation, inner analysis of the unknown language is the only trustworthy procedure. Hurrian and Urartian are definitely related languages, but neither may yet be safely used to explain the other. Urartian has been solved to some extent with the help of its rather free use of ideograms and the Assyrian versions of two bilingual inscriptions.

Excavations at Ras Shamra in 1929 unearthed the remains of Ugarit. Inscriptions in an unknown simple system of cuneiform were found; the low number of 30 different signs pointed to an alphabetic type. The use of a vertical stroke as word-divider facilitated the decipherment, which was based on the correct assumption that an early North Semitic Canaanite dialect was involved. Thus the script was solved with astonishing speed by Hans Bauer, Edouard Dhorme, and Charles Virolleaud, yielding a Semitic dialect named Ugaritic, closely related to Old Phoenician. Hurrian inscriptions in the same script were also found, as were texts in conventional Middle Babylonian cuneiform.

Deciphering of Ugaritic

Influence of cuneiform. The main type of cuneiform, with its inventory of ideograms (including "determinatives" or "classifiers") and phonetic signs, is a word-syllabic system like the Egyptian, hieroglyphic Hittite, Minoan-Mycenaean, proto-Elamite, and proto-Indic. The Sumerian system seems to be the oldest. To what extent it stimulated the origin or influenced the development of the others is a difficult problem connected with the monogenesis or polygenesis (common or multiple origin) of writing. The Phoenician consonantal script provided the new typological pattern on which the Ugaritic and Old Persian systems were constructed, keeping only the outer likeness of the wedge form. (J.Pl./Ed.)

HIEROGLYPHIC WRITING

Hieroglyphic writing is a system that employs characters in the form of pictures. These individual signs, called hieroglyphs, may be read either as pictures, as symbols for pictures, or as symbols for sounds. The name hieroglyphic (from the Greek word for "sacred carving") is first encountered in the writings of Diodorus Siculus (1st century BC). Earlier, other Greeks had spoken of sacred signs when referring to Egyptian writing. Among the Egyptian scripts, the Greeks labeled as hieroglyphic the script that they found on temple walls and public monuments, in which the characters were pictures sculpted in stone. The Greeks distinguished this script from two other forms of Egyptian writing that were written with ink on papyrus or on other smooth surfaces. These were known as the hieratic, which was still employed during the time of the ancient Greeks for religious texts, and the demotic, the cursive script used for ordinary documents.

Hieroglyphic, in the strict meaning of the word, designates only the writing on Egyptian monuments. The word has, however, been applied for about 100 years to the writing of other peoples, insofar as it consists of picture signs used as writing characters. The name hieroglyphics is, for example, always used to designate the scripts of the Indus civilization and of the Hittites, who also possessed other scripts, in addition to the Mayan, the Incan, and Easter Island writing forms, and also the signs on the Phaistos Disk on Crete. Colloquially, the word hieroglyphics has been extended to mean any sort of illegible or barely legible writing.

Non-Egyptian hieroglyphic scripts

Because of their pictorial form, hieroglyphs were difficult to write and were used only for monument inscriptions. They were usually supplemented in the writing of a people

by other, more convenient scripts. Among living writing systems, hieroglyphic scripts are no longer used.

The rest of this section is concerned only with Egyptian hieroglyphic writing.

Development of Egyptian hieroglyphic writing. The most ancient hieroglyphs date from the end of the 4th millennium BC and comprise annotations to the scenes cut in relief—found on slabs of slate in chapels or tombs—that had been donated as votive offerings. Although by no means all of these earliest signs can be read today, it is nonetheless probable that these forms are based on the same system as the later classical hieroglyphs. In individual cases, it can be said with certainty that it is not the copied object that is designated but rather another word phonetically similar to it. This circumstance means that hieroglyphs were from the very beginning phonetic symbols. An earlier stage consisting exclusively of picture writing using actual illustrations of the intended words cannot be shown to have existed in Egypt; indeed, such a stage can with great probability be ruled out. No development from pictures to letters took place; hieroglyphic writing was never solely a system of picture writing. It can also be said with certainty that the jar marks (signs on the bottom of clay vessels) that occur at roughly the same period do not represent a primitive form of the script. Rather, these designs developed in parallel fashion to hieroglyphic writing and were influenced by it.

Improbability of relationship to cuneiform

It is not possible to prove the connection of hieroglyphs to the slightly older cuneiform characters used by the Sumerians in southern Mesopotamia. Such a relationship is improbable because the two scripts are based on entirely different systems. What is conceivable is a general tendency toward words being fixed by the use of signs, without transmission of particular systems.

Invention and uses of hieroglyphic writing. The need to identify a pictorial representation with a specific, unique event, such as a hunt or a particular battle, led to the invention of hieroglyphic writing. Hieroglyphs added to a scene signified that this illustration represented a particular war rather than an unspecified one or war in general. This new attitude toward time and toward history as unique events in time led to the invention of hieroglyphic writing. The system first appeared only in connection with relief depictions, which they explained by means of place-names. Beginning in the 1st dynasty (*c.* 2925–*c.* 2775 BC), images of persons were also annotated with their names or titles, a further step toward expressing individuality and uniqueness. The so-called annalistic tablets of the first two dynasties were pictorial representations of the events of a year with specifically designated personal names, places, and incidents. For example, accompanying a scene of the pharaoh's triumph over his enemies is the annotation "the first occasion of the defeat of the Libyans." Simultaneously, the writing of the Egyptians began to appear unaccompanied by pictorial representations, especially on cylindrical seals. These roller-shaped incised stones were rolled over the moist clay of jar stoppers. Their inscription prevented the sealed jar from being covertly opened and at the same time described its contents and designated the official responsible for it. In the case of wine, its origin from a specific vineyard and often also the destination of the shipment were designated, and, as a rule, so was the name of the reigning king.

From the stone inscriptions of the 1st dynasty, only individual names are known, these being mainly the names of kings. In the 2nd dynasty, titles and names of offerings appear, and, at the end of this dynasty, sentences occur for the first time. The discovery of a blank papyrus scroll in the grave of a high official, however, shows that longer texts could have been written much earlier; *i.e.,* since the early part of the 1st dynasty.

Relationship of writing and art. The form of these hieroglyphs of the archaic period (the 1st to 2nd dynasty) corresponds exactly to the art style of this age. Although definite traditions or conventions were quickly formed with respect to the choice of perspective—*e.g.,* a hand was depicted only as a palm, an eye or a mouth inscribed only in front view—the proportions remained flexible. The prerequisite of every writing system is a basic standardization, but such a standardization is not equivalent to a canon (an established body of rules and principles) in the degree of stylistic conformity that it requires. A recognized canon of Egyptian hieroglyphic writing arose in the 3rd dynasty and was maintained until the end of the use of the script.

Formation of the standard and canon of Egyptian writing

In that hieroglyphic signs represented pictures of living beings or inanimate objects, they retained a close connection to the fine arts. The same models formed the basis of both writing and art, and the style of the writing symbols usually changed with the art style. This correspondence occurred above all because the same craftsmen painted or incised both the writing symbols and the pictures. Deviations from the fine arts occurred when the writing, which was more closely bound to convention, retained patterns that the fine arts had eliminated. The face in front view is an example of this. This representation, apart from very special instances, was eventually rejected as an artistic form, the human face being shown only in profile. The front view of the face was, however, retained as a hieroglyph from the archaic period to the end of the use of hieroglyphic writing. Similar cases involve the depiction of various tools and implements. Although the objects themselves fell out of use in the course of history—*e.g.,* clubs used as weapons—their representations, mainly misunderstood, were preserved in the hieroglyphic script. The hieroglyphs corresponding to objects that had disappeared from daily life were therefore no longer well known and were often distorted beyond recognition. But the style of representation in the hieroglyphs still remained closely bound to the art of the respective epoch. Thus there appeared taut, slender forms or sensuous, fleshy ones, or even completely bloated characters, according to the art style of the period.

Media for hieroglyphic writing. In historical times (2800 BC–AD 300), hieroglyphic writing was used for inscribing stone monuments and appears in Egyptian relief techniques, both high relief and bas-relief; in painted form; on metal, sometimes in cast form and sometimes incised; and on wood. In addition, hieroglyphs appear in the most varied kinds of metal and wood inlay work. All of these applications correspond exactly with the techniques used in fine art, and the same craftsmen who produced the works of art painted or incised the hieroglyphic inscriptions.

Subject matter of inscriptions

Hieroglyphic texts are found primarily on the walls of temples and tombs, but they also appear on memorials and gravestones, on statues, on coffins, and on all sorts of vessels and implements. Hieroglyphic writing was used as much for secular texts—historical inscriptions, songs, legal documents, scientific documents—as for religious subject matter—cult rituals, myths, hymns, grave inscriptions of all kinds, and prayers. These inscriptions were, of course, only a decorative monumental writing, unsuitable for everyday purposes. For popular use, hieratic script was developed, an abbreviated form of the picture symbols such as would naturally develop in writing with brush and ink on smooth surfaces like papyrus, wood, and limestone.

Writing and religion. The influence of religious concepts upon hieroglyphic writing was confined to two cases. In the 3rd millennium, certain signs were avoided or used in garbled form in grave inscriptions for fear that the living beings represented by these signs could harm the deceased who lay helpless in the grave. Among these taboo symbols were human figures and dangerous animals, such as scorpions and snakes. Second, in all periods and for all uses of the writing, symbols to which a positive religious significance was attached were regularly placed in front of other signs, even if they were to be read after them. Among these were hieroglyphs for God or individual gods, as well as those for the king or the palace. Thus, for example, the two signs, denoting the word combination "servant of God" (priest), are written so that the symbol for God, , stands in front of that for servant, , although the former is to be read last. Moreover, theology traced the invention of hieroglyphic writing back to the god Thoth, although this myth of its divine origin did not have an effect on the development of the script. In the late

period, Egyptian texts referred to hieroglyphic inscriptions as "writing of God's words"; earlier, in contrast, they were simply called pictures.

Literacy and knowledge of hieroglyphic writing. At all periods only a limited circle understood the script. Only those who needed the knowledge in their professions acquired the arts of writing and reading. These people were, for example, officials, doctors, and priests (insofar as they had to be able to read rituals and other sacred texts), as well as craftsmen whose work included the making of inscriptions. Under Greek and especially under Roman rule, the knowledge declined and was entirely confined to temples where priests instructed their pupils in the study of hieroglyphic writing. From the time of the rule of the Ptolemies (305 to 30 BC), national consciousness became more and more narrowly bound up with religion, and for both the national consciousness and religion alike the tradition-filled hieroglyphic writing was an outward sign—in the fullest sense, a symbol. There was no lack of attempts to replace the hieroglyphic writing, cumbersome and ever more divergent from the spoken language, with the simpler and more convenient Greek script. Such experiments, however, remained ineffective precisely because of the emotional value that the old writing system had when the country was under the foreign domination of the Macedonian Greeks and the Romans.

Decline of hieroglyphic writing

Christianity and the Greek alphabet. The situation was altered with the conversion of the country to Christianity in the 2nd and 3rd centuries AD. The new religion fought against the Egyptian polytheism and traditions, and with its victory, the Greek script triumphed. From the beginning, Egyptian Christians used the Greek alphabet for writing their spoken Egyptian language. This practice involved enlarging the Greek alphabet with seven supplementary letters for Egyptian sounds not present in Greek. As a consequence, the knowledge of hieroglyphic writing quickly declined. The last evidence of the writing system is a rock inscription from the island of Philae, dating from Aug. 24, 394, from the reign of the emperor Theodosius I. The language as well as the writing system of the Egyptian Christians is called Coptic.

Characteristics of hieroglyphic writing. The system of hieroglyphic writing has two basic features: first, representable objects are portrayed as pictures (ideograms) and, second, the picture signs are given the phonetic value of the words for these represented objects (phonograms). At the same time, these signs are also written to designate homonyms, similar-sounding words. The writing disregards vowels and also, in earlier times, the semivowels *i, y,* and *w,* thus offering more possibilities for the transference of signs to words with identical consonant combinations. For example, the sign for "wood" is written as a branch, , which is pronounced with the consonants *ḫ* and *t,* which occur in the Egyptian word for wood. Other words with the same series of consonants can also be written with the same basic sign—*e.g., ḫt* "after," *ḫtiʾ* "to retreat," or *ḫtiʾ* "to carve." Words that consisted of only one consonant, plus one or more vowels, supplied single consonant signs. The Egyptians, however, never reduced their writing to an alphabet by discarding the multiconsonant signs; rather, they retained clearly the form of the original words. When doubts occurred, as in the case of the three signs for the frequent consonant series *m* + *r* (the hoe, , the chisel, , and the pyramid,), the plurality was used to make clear distinctions between words: all derivations from the stem *mr* "to love" were written with the hoe; those from the stem *mr* "to be ill," with the chisel; and those words related to pyramids with the sign for pyramid. Thus, two or more existing signs for the same sound or combination of sounds were retained and used in conscious distinction to promote easier readability. Although each sign originally had only one reading, occasional ambiguities did develop through the convergence of two symbols of similar form, such as those for the thighbone and the shankbone of an animal. A few signs, therefore, had two or, less commonly, three readings in classical Egyptian writing.

Ideograms and phonograms

Reading aids: spelling, phonetic complements, determinatives. By means of this rebus system in which letters and pictures were combined, the Egyptians could write a large number of the words of their language. But there remained a residue for which no drawable word with the same consonant framework presented itself—*e.g., nḫt* "strong." Here the Egyptians spelled out the word: for *n,* they had a sign, the water symbol (from the word *nwy* "flood"), and for *ḫt* they had the above-mentioned sign for wood, , so that they could now write *nḫt* as .

Two additional reading aids that were quickly added to this system promoted distinctness and readability. For multiconsonant signs, one or more consonants, or in some cases all of them, were also written to serve as a phonetic complement. Thus there is for *mr* "to be ill," in which the owl (top, left) possesses the phonetic value *m* and the mouth ( ) that of an *r.* In cases like this, the consonants, according to the conception of modern scholars, were written twice but were read only once. For the Egyptians, the single consonant signs were there simply as reading aids for the clarification of the word sign, the logogram. Accordingly, they wrote , in which is complemented by the two signs , *ḫ* , and , *t,* which appear after it.

Phonetic complements

In addition, determinatives—signs that do not represent a phonetic value but serve only to inform the reader as to the family of meanings to which the designated word belongs—were quickly formed out of these. The consonant combination *ḫtiʾ* "to engrave" has a knife written after it; on the other hand, *ḫtiʾ* "to retreat" has legs striding backward. Thus, these two words, otherwise written identically, are differentiated graphically as and . In this manner, each Egyptian word possessed its own writing picture with which it was strictly associated. Grammatical endings were attached to this word picture and stood after the determinative. From the outset, therefore, Egyptian writing was a complete script; that is, it could unequivocally fix any word, including all derivations and all grammatical forms.

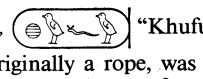
Summary of the types of signs. In summation, hieroglyphs can be separated into three groups, of which the first two render a phonetic value and the third represents mute reading aids: (1) ideograms, or signs that should be read as the word they represent—*e.g.,* , "branch"; (2) phonograms, or signs that do not refer back to the objects they represent but stand simply for one or more consonants—*e.g.,* as *n* and as *ḫt* in , *nḫt* "strong"; and (3) determinatives, which possess no phonetic value but which aid the reader by leading him to the correct interpretation of the meaning—*e.g.,* in , *ḫtiʾ* "to retreat."

Egyptian writing lacked punctuation in our meaning of the term. Line and stanza signs appeared only in certain literary texts.

Number of symbols. In the classical period of Egyptian writing, the number of hieroglyphs totaled approximately 700. Their number multiplied considerably in the late period (about 600 BC); this proliferation occurred because scholars constantly invented new forms or signs. The additional hieroglyphs were, however, always in accordance with the principles that had governed Egyptian writing from its beginnings. The hieroglyphic system remained flexible throughout all periods, always open to innovation, even though, as with every writing system, convention played a preponderant role.

Invention of new signs

Direction of the writing. The lines of hieroglyphs were written from right to left or, less frequently, from left to right. Vertical rows of signs could be placed next to horizontal rows, according to the particular demands of the architectural setting. The direction of the writing is immediately ascertainable because the signs almost always face the beginning of the row. Occasionally, some signs are turned around in the row, presumably so that two human figures can face one another and thus avoid standing with their backs toward each other. These rotations of signs are infrequent, however, and are found almost exclusively in the names of kings. Royal names were enclosed in a ring, the

so-called cartouche; *e.g.*,  "Khufu," in Egyptian *Ḫwfw*. This ring, originally a rope, was supposed to protect the bearer of the enclosed name from injury and, in particular, from harmful magic.

Egyptian pedagogical traditions. To understand hieroglyphic writing, one must know about its tradition within Egypt. The Egyptian student of writing, who brought with him a knowledge of the spoken language as his mother tongue, began by learning the script picture corresponding to each word without having isolated its elements; *i.e.*, its individual signs. Through centuries this pedagogical tradition in the schools helped Egyptian words retain the original established spelling, with only minor—usually stylistic—changes, even when the phonetic form had radically changed. Hieroglyphic writing thus conceals historical sound changes.

Sound changes shown by misspellings

The mistakes in hearing made by pupils in the writing schools have helped scholars to understand the phonetic changes that occurred in the development of the Egyptian language. When the pupil who was learning to write the hieroglyphic script did not recognize a word dictated to him, he wrote it badly—that is, just as he heard it. Because he had not yet learned to spell in the orthodox manner, what appeared on his papyrus was usually a word that sounded similar to the dictated but misunderstood term and whose word picture was familiar. Thus, although Egyptian writing was originally composed of symbols that represented a phonetic value, the system was transmitted in the form of word pictures—that is, closed or indivisible groups, generally of several signs per word.

Cryptographic hieroglyphic writing. That knowledge of the hieroglyphic system and the principles upon which it was devised had not become lost is attested by two phenomena: cryptography and the development of the hieroglyphic writing during the last millennium of its existence. From the middle of the 3rd millennium but more frequently in the New Kingdom (from *c.* 1539 BC), hieroglyphic texts are encountered that have a very strange appearance. The absence of familiar word groups and the presence of many signs not found in the canon characterize these texts at first glance as cryptographic, or secret, writing. This kind of hieroglyphic writing was probably intended as an eyecatcher, to entice people to seek the pleasure of deciphering it. Composed according to the original principles of the script, these inscriptions differed only in that certain features excluded when the original canon was formulated were now exploited. The new possibilities involved not only the forms of the signs but also their selection. For example, the mouth was not drawn in front view ( ), as in the classical script, but in profile ( ), although it had the same phonetic value. An example of a change in the choice of signs is the case in which a man carrying a basket on his head ( ), a determinative without phonetic value in the classical script, was later to be read as *f* and was used in lieu of the familiar sign having this phonetic value, that of the horned viper. In the new selection of the sign, the phonetic value is obtained from the word *f3i* "to carry" (neglecting its two weak consonants), in accordance with a principle that the inventors of the writing had applied in 3000 BC. These cryptographic inscriptions prove that alongside the method of instruction in the schools, which was based on memorization or recognition, not upon analytical understanding, there was another tradition that transmitted knowledge of the basic principles of the hieroglyphic script. A command of the principles of hieroglyphics similar to that which the composers of the cryptic inscriptions had was presupposed for the puzzle-happy decipherers. Because the encoded texts often consisted of a petition by the inventor of the text to say a prayer on his behalf, the number of these decipherers must surely not have been small.

Transmission of the hieroglyphic principles

Growth of hieroglyphic writing during the 1st millennium BC. At about the middle of the 1st millennium BC, Egyptian writing experienced new developments and a revival of interest. Again the inscriptions abounded with new signs and sign groups unknown in the classical period, all generated according to the same principles as the classical Egyptian script and the cryptographic texts. The writing of this late period was distinguished from the cryptograms in that this script, like every normal system of writing, developed a fixed tradition, being intended not to conceal but to be read easily, whereas the cryptography strove for originality.

Stages of hieroglyphic writing. The development of hieroglyphic writing thus proceeded approximately as follows: at first only the absolutely necessary symbols were invented, without a canonization of their artistic form. In a second stage, easier readability (*i.e.*, increased rapidity of reading) was achieved by increasing the number of signs (thereby eliminating some doubts) and by employing determinatives. Finally, after the second stage had endured, essentially unaltered, for about 2,000 years, the number of symbols increased to several thousand in about 500 BC. This rampant growth process occurred through the application of hitherto unused possibilities of the system. With the triumph of Christianity, the knowledge of hieroglyphic writing was extinguished along with the ancient Egyptian religion.

Tools. The tools used by the craftsmen for writing hieroglyphic symbols consisted of chisels and hammers for stone inscriptions and brushes and colours for wood and other smooth surfaces. Only for the cursive scripts, hieratic and demotic, were special materials developed. Leather and papyrus became writing surfaces, and the stems of rushes in lengths of six to 13 inches (15 to 33 centimetres), cut obliquely at the writing end and chewed to separate the fibres into a brushlike tip, functioned as writing implements. The split calamus reed was introduced into Egypt by the Greeks in the 3rd century BC.

Development of special writing materials

Hieratic script. The Egyptian cursive script, called hieratic writing, received its name from the Greek *hieratikos* ("priestly") at a time when the script was used only for sacred texts. Everyday secular documents were written in another style, the demotic (Greek *dēmotikos*, "for the people" or "in common use") script.

Relation of hieratic to hieroglyphic script. The structure of the hieratic script corresponds with that of hieroglyphic writing. Changes occurred in the characters of hieratic simply because they could be written rapidly with brush or rush and ink on papyrus. In general, the picture form is not, or not easily, recognizable. Because their models were well known and in current use throughout Egyptian history, the hieratic symbols never strayed too far from them. Nevertheless, the system differs from the hieroglyphic script in some important respects:

1. Hieratic was written in one direction only, from right to left. In earlier times the lines had run vertically and later, about 2000 BC, horizontally. Subsequently the papyrus scrolls were written in columns of changing widths.
2. There were ligatures in hieratic so that two, but no more than two, signs could be written in one stroke.
3. As a consequence of its decreased legibility, the spelling of the hieratic script was more rigid than that of hieroglyphic writing. Variations from uniformity at a given time were minor; but, during the course of the various periods, the spelling developed and changed. As a result, hieratic texts do not correspond exactly to contemporary hieroglyphic texts, either in the placing of signs or in the spelling of words.
4. Hieratic used diacritical additions to distinguish between two signs that had grown similar to one another because of cursive writing. For example, the cow's leg received a supplementary distinguishing cross, because in hieratic it had come to resemble the sign for the leg of a man. Certain hieratic signs were taken into the hieroglyphic script.

Role of hieratic in Egypt

All commonplace documents—*e.g.*, letters, catalogs, and official writs—were written in hieratic script, as were literary and religious texts. In the life of the Egyptians, hieratic script played a larger role than hieroglyphic writing and was also taught earlier in the schools. In offices, hieratic was replaced by demotic in the 7th century BC, but it remained in fashion until much later for religious texts of all sorts. The latest hieratic texts stem from the end of the 1st century or the beginning of the 2nd century AD.

Demotic script. Demotic script is first encountered at

the beginning of the 26th dynasty, in about 660 BC. The writing signs plainly demonstrate its connection with the hieratic script, although the exact relationship is not yet clear. The demotic characters are more cursive (flowing and joined) and thus more similar to one another, with the result that they are more difficult to read than are the hieratic forms. Countering this difficulty, there is less freedom for the writer's individual variations. It appears that demotic was originally developed expressly for government office use—that is, for documents in which the language was extensively formalized and thus well suited for the use of a standardized cursive script. Only some time after its introduction was it used for literary texts in addition to documents and letters; much later it was employed for religious texts also. The latest dated demotic text, from Dec. 2, 425, consists of a rock inscription at Philae. In contrast to hieratic, which is almost without exception written in ink on papyrus or other flat surfaces, demotic inscriptions are not infrequently found engraved in stone or carved in wood.

Alternative demotic spelling. The demotic system corresponds to the hieratic and hence also to the hieroglyphic system. Alongside the traditional spelling, however, there was another spelling that took account of the markedly altered phonetic form of the words by appropriate respelling. This characteristic applied especially to a large number of words that did not occur in the older language and for which no written form had consequently been passed down. The nontraditional spelling could also be used for old, familiar words.

Decipherment of hieroglyphic writing. With the possible exception of Pythagoras, no Greek understood the nature of hieroglyphic writing. The Greeks did not obtain guidance from their Egyptian contemporaries, some of whom even lived on Italian soil and wrote proper hieroglyphic inscriptions on Roman obelisks. Rather, the Greek tradition taught that hieroglyphs were symbolic signs or allegories. The Egyptian-born Greek philosopher Plotinus interpreted hieroglyphic writing entirely from the viewpoint of his esoteric philosophy. Only one of the numerous works on the hieroglyphic script written in late antiquity has been preserved: the *Hieroglyphica* of Horapollon, a Greek Egyptian who probably lived in the 5th century AD. Horapollon made use of a good source, but he himself certainly could not read hieroglyphic writing and began with the false hypothesis of the Greek tradition, namely, that hieroglyphs were symbols and allegories, not phonetic signs.

Greek beliefs concerning hieroglyphic writing

From G. Moller, *Zeitschrift des Deutschen Vereins fur Buchwesen und Schrifttum*, ii (1919), 78

Figure 2: *Egyptian hieroglyphs and their cursive equivalents.* The hieroglyphs depict (top to bottom): three fox skins tied together; a whip; a single-barbed harpoon; an adz at work on a block of wood; a stone jug with handle; a scribe's outfit; a roll of papyrus tied with a cord.

The Middle Ages neither possessed any knowledge of hieroglyphic writing nor took any interest in it. But a manuscript of Horapollon brought to Florence in 1422 stirred great interest among the humanists. Without giving a thought to the possibility that ancient Egyptian originals might be available in Rome, Renaissance artists designed hieroglyphs after Horapollon's descriptions, as well as from their own imaginations. They used hieroglyphs as wisdom-laden symbols in architecture and also in drawings and paintings.

Kircher's attempts at decipherment. The great German scholar Athanasius Kircher (1602–80) began his attempts at decipherment with the Coptic language and with the correct hypothesis that the hieroglyphs recorded an earlier stage of this language. He also believed, again correctly, that the signs recorded phonetic values. In spite of this, he did not arrive at correct results—with the exception of a single character. This failure can be attributed not only to Kircher's erroneous assumption that the hieroglyphs must correspond phonetically to an alphabet but primarily to the fact that he was most interested in the Renaissance conception of a supposed symbolic meaning constituting the deeper significance of hieroglyphs. In his view, the phonetic value of the hieroglyphs was merely the commonplace, superficial part of the sign.

Discovery of the Rosetta Stone. Both the intellectual and the physical prerequisites for the deciphering of the hieroglyphic script first presented themselves at the end of the 18th century. By accident, a stone that exhibited three different scripts—hieroglyphic, demotic, and Greek—was discovered by members of Napoleon's expedition to Egypt in 1799 near Rashīd (French Rosette; English Rosetta) on the Mediterranean coast. The Greek text stated clearly that the document set forth the same text in the sacred script, the folk or popular script, and Greek. The stone was promptly made known to all interested scholars. Important partial successes in the effort of deciphering the scripts were achieved by the Swede Johan David Åkerblad and by the English physicist Thomas Young, who mainly studied the demotic text, again beginning with the false hypothesis that the hieroglyphs were symbols. Young succeeded in proving that they were not symbols—at least that the proper names were not—and that the demotic and hieratic signs had come from the hieroglyphs. (He first published this result in the supplement to the 4th, 5th, and 6th editions of the *Encyclopædia Britannica*.) He was the first to isolate correctly some single-consonant hieroglyphic signs. But a wrong turn in the course of his investigations then prevented him from fully deciphering the writing.

Thomas Young's work

Champollion's decipherment. This task of complete decipherment was first accomplished by the Frenchman Jean-François Champollion (1790–1832) in 1822, after long years of intensive work and many setbacks. His success was due to the recognition that hieroglyphic writing, exactly like the hieratic and demotic scripts derived from it, did not constitute a writing system of symbols but rather a phonetic script. He arrived at this breakthrough by an exact comparison of the three Egyptian forms of writing, as well as by reference to Coptic, the late phase of the Egyptian language that was written with the Greek alphabet and was thus directly readable. The Coptic language was also understood at that time. Starting, as had his predecessors, from Ptolemy and Cleopatra, both ring-enclosed royal names, and adding the hieroglyphic spelling of Ramses' name, Champollion determined, essentially correctly, the phonetic values of the signs. Soon after, he also learned to read and translate a large number of Egyptian words. Since then, precise research has confirmed and refined Champollion's approach and most of his results.

(He.B.)

ALPHABETIC WRITING

An alphabet is a set of graphs, or characters, used to represent the phonemic structure of a language; in most

alphabets the characters are arranged in a definite order, or sequence (*e.g.*, A, B, C, etc.). In the usual case, each character represents either a consonant or a vowel, rather than a syllable or group of consonants and vowels. As a result, the number of characters required can be held to a relative few. A language that has 30 consonant sounds and five vowels, for example, needs at most only 35 separate letters. In a syllabary, on the other hand, the same language would require 30×5 symbols to represent each possible consonant-vowel syllable (*e.g.*, separate forms for *ba, be, bi, bo, bu; da, de, di,* and so on) and an additional five symbols for the vowels, thereby making a total of 155 individual characters. Both syllabaries and alphabets are phonographic symbolizations; that is, they represent the sounds of words rather than units of meaning.

The word alphabet, from the first two letters of the Greek alphabet—*alpha* and *beta*—was first used, in its Latin form, *alphabetum,* by Tertullian (2nd–3rd century AD), a Latin ecclesiastical writer and church father, and by St. Jerome. The classical Greeks customarily used the plural of *to gramma* ("the letter"); the later form *alphabētos* was probably adopted under Latin influence.

Theories of the origin of the alphabet. The evolution of the alphabet involved two important achievements. The first was the step taken by a group of Semitic-speaking people, perhaps the Phoenicians, on the eastern shore of the Mediterranean between 1700 and 1500 BC. This was the invention of a consonantal writing system known as North Semitic. The second was the invention, by the Greeks, of characters for representing vowels. This step occurred between 800 and 700 BC. While some scholars consider the Semitic writing system an unvocalized syllabary and the Greek system the true alphabet, both are treated here as forms of the alphabet.

Over the centuries, various theories have been advanced to explain the origin of alphabetic writing, and, since classical times, the problem has been a matter of serious study. The Greeks and Romans considered five different peoples as the possible inventors of the alphabet—the Phoenicians, Egyptians, Assyrians, Cretans, and Hebrews. Among modern theories are some that are not very different from those of ancient days. Every country situated in or more or less near the eastern Mediterranean has been singled out for the honour. Egyptian writing, cuneiform, Cretan, hieroglyphic Hittite, the Cypriot syllabary, and other scripts have all been called prototypes of the alphabet. The Egyptian theory actually subdivides into three separate theories, according to whether the Egyptian hieroglyphic, the hieratic, or the demotic script is regarded as the true parent of alphabetic writing. Similarly, the idea that cuneiform was the precursor of the alphabet may also be subdivided into those singling out Sumerian, Babylonian, or Assyrian cuneiform.

Egyptian theory of the origin of the alphabet

Among the various other theories concerning the alphabet are the hypotheses that the alphabet was brought by the Philistines from Crete to Palestine, that the various ancient scripts of the Mediterranean countries developed from prehistoric geometric symbols employed throughout the Mediterranean area from the earliest times, and that the proto-Sinaitic inscriptions (discovered since 1905 in the Sinai Peninsula) represent a stage of writing intermediate between the Egyptian hieroglyphics and the North Semitic alphabet. Another hypothesis, the Ugaritic theory, evolved after an epoch-making discovery in 1929 (and the years following) at the site of the ancient Ugarit, on the Syrian coast opposite the most easterly cape of Cyprus. Thousands of clay tablets were found there, documents of inestimable value in many fields of research (including epigraphy, philology, and the history of religion). Dating from the 15th and 14th centuries BC, they were written in a cuneiform alphabet of 30 letters.

The Early Canaanite theory is based on several undeciphered inscriptions also discovered since 1929 at various Palestinian sites; the writings belong in part to *c.* 1700 BC and are thus the earliest preserved documents in an alphabetic writing.

Despite the conflict in theories, scholars are generally agreed that, for about 200 years before the middle of the 2nd millennium BC, alphabet making was in the air in the Syro-Palestinian region. It is idle to speculate on the meaning of the various discoveries referred to. That they manifest closely related efforts is certain; what the exact relationship among these efforts was, and what their relationship with the North Semitic alphabet was, cannot be said with certainty.

Conditions favourable for alphabetic writing

It can, however, be ascertained that the period from 1730 to 1580 BC in Syria, Palestine, and Egypt, during which there was an uprooting of established cultural and ethnic patterns in the Fertile Crescent, provided conditions favourable to the conception of an alphabetic script, a kind of writing that would be more accessible to larger groups of people, in contrast to the scripts of the old states of Mesopotamia and Egypt, which were confined largely to the priestly class. In default of other direct evidence, it is reasonable to suppose that the actual prototype of the alphabet was not very different from the writing of the earliest North Semitic inscriptions now extant, which belong to the last two or three centuries of the 2nd millennium BC. The North Semitic alphabet was so constant for many centuries that it is impossible to think that there had been any material changes in the preceding two to three centuries. Moreover, the North Semitic languages, based as they are on a consonantal root (*i.e.*, a system in which the vowels serve mainly to indicate grammatical or similar changes), were clearly suitable for the creation of a consonantal alphabet.

The inventor or inventors of the alphabet were, no doubt, influenced by Egyptian writing—perhaps also by other scripts. Indeed, it is probable that those who invented the alphabet were acquainted with most of the scripts current in the eastern Mediterranean lands at the time. Though the nationality of the inventor or inventors of the alphabet is unknown, it is now generally agreed that he or they belonged to the Northwest Semitic linguistic group, which includes the ancient Canaanites, Phoenicians, and Hebrews.

Originally, graphs were perhaps "motivated" pictorial signs that were subsequently used to represent the initial sound of the name of the pictured object. The North Semitic alphabet remained almost unaltered for many centuries. If the signs' external form (which, it must be emphasized, had no particular significance) is ignored and only their phonetic value, number, and order are considered, the modern Hebrew alphabet may be regarded as a continuation of the original alphabet created more than 3,500 years ago. The Hebrew order of the letters seems to be the oldest. The earliest evidence that the Hebrew alphabet was learned systematically was left in the form of a schoolboy's scribbling on the vertical face of the upper step of a staircase leading up to the palace at Tel Lakhish, in southern Israel. It includes the scratching of the first five letters of the early Hebrew alphabet in their conventional order, and it belongs to the 8th or 7th century BC.

Development and diffusion of alphabets. At the end of the 2nd millennium BC, with the political decay of the great nations of the Bronze Age—the Egyptians, Babylonians, Assyrians, Hittites, and Cretans—a new historical world began. In Syria and Palestine, the geographical centre of the Fertile Crescent, three nations—Israel, Phoenicia, and Aram—played an increasingly important political role. To the south of the Fertile Crescent, the Sabaeans, a South Arabian people (also Semites, though South Semites), attained a position of wealth and importance as commercial intermediaries between the East and the Mediterranean. To the west, seeds were sown among the peoples who later constituted the nation of Hellas—the Greeks. As a result, an alphabet developed with four main branches: (1) the so-called Canaanite, or main branch, subdivided into Early Hebrew and Phoenician varieties; (2) the Aramaic branch; (3) the South Semitic, or Sabaean, branch; and (4) the Greek alphabet, which became the progenitor of the Western alphabets, including the Etruscan and the Latin. The Canaanite and Aramaic branches constitute the North Semitic main branch.

Development of four main branches of the alphabet

The Canaanite alphabet. The two Canaanite branches may be subdivided into several secondary branches. First, Early Hebrew had three secondary branches—Moabite, Edomite, and Ammonite—and two offshoots—the script

of Jewish coins and the Samaritan script, still in use today for liturgical purposes only. Second, Phoenician can be divided into Phoenician proper and "colonial" Phoenician. Out of the latter developed the Punic and neo-Punic scripts and probably also the Libyan and Iberian scripts.

The term Early Hebrew is used to distinguish this branch from the later so-called Square Hebrew. The Early Hebrew alphabet had already begun to acquire its distinctive character by the 11th century BC. It was used officially until the 6th century BC and lingered on for several centuries more. In a stylized form it was used on Jewish coins from 135 BC to AD 132–135. The most ancient example of Early Hebrew writing is that of the Gezer Calendar of the period of Saul or David (*i.e., c.* 1000 BC). The oldest extant example of the Early Hebrew ABCs is the 8th–7th-century-BC schoolboy graffito mentioned above. A cursive style reached its climax in the inscriptions at Tel Lakhish, dating from the beginning of the 6th century BC. The Leviticus and other small Early Hebrew fragments found in the Dead Sea caves, which are probably from the 3rd century BC, are the only remains of what is considered to be the Early Hebrew book, or literary, hand.

Paramount importance of the Phoenician alphabet

It is difficult to overestimate the importance of the Phoenician alphabet in the history of writing. The earliest definitely readable inscription in the North Semitic alphabet is the so-called Ahiram inscription found at Byblos in Phoenicia (now Lebanon), which probably dates from the 11th century BC. There is, however, no doubt that the Phoenician use of the North Semitic alphabet went further back. By being adopted and then adapted by the Greeks, the North Semitic, or Phoenician, alphabet became the direct ancestor of all Western alphabets. Only very few inscriptions have been found in Phoenicia proper. This rarity of indigenous documents is in contrast to the numbers of Phoenician inscriptions found elsewhere—on Cyprus, Malta, Sicily, and Sardinia, and in Greece, North Africa, Marseille, Spain, and other places.

The Aramaic alphabet. The adaptation of the North Semitic alphabet to the Aramaic language took place at some time in the 10th century BC, when Aramaic was spoken in several petty kingdoms in northern Mesopotamia and Syria, the most important of them being Dammeshek (Damascus). The process of the reestablishment of the Assyrian Empire and its hegemony over a good part of the Middle East began in the 9th century. One after another, the Aramaean states gave way under Assyrian onslaught. Dammeshek, the last survivor, fell in 732 BC. The end of Aramaean political independence marked the beginning of Aramaean cultural and economic supremacy in western Asia. The transplantation of masses of Aramaeans by the Assyrians, a political measure designed to break up military alliances, bore remarkable fruit. By the end of the 8th century BC the use of the Aramaic language and alphabet had become very widespread in Assyria itself; by the end of the following century all of Syria and a large part of Mesopotamia had become thoroughly Aramaized.

Aramaic inscriptions

On the whole, the few early Aramaic inscriptions that have been found belong to the 9th, 8th, and 7th centuries BC. Inscriptions from the 6th and later centuries are more numerous; the increase reflects the rapid spread of the Aramaic alphabet throughout the Middle East. Numerous Aramaic papyri and ostraca (inscribed pottery fragments) have been found in Egypt; the earliest of these can be dated to *c.* 515 BC, while the most famous are the Elephantine papyri, containing information of a religious and economic nature about a 5th-century Hebrew military colony in Egypt. Aramaic inscriptions have been found in northern Arabia, Palestine, Lycia, Cappadocia, Lydia, Cilicia, Assyria, and as far afield as Greece, Afghanistan, and India.

Almost as if by prearrangement, all of the alphabetic scripts west of Syria seem to have been derived, directly or indirectly, from the Canaanite alphabet, whereas the hundreds of alphabetic writings of the East apparently have sprung from the offshoots of the Aramaic alphabet. On the whole, the direct and indirect descendants of the Aramaic alphabet can be divided into two main groups: the scripts employed for Semitic languages and those adapted to non-Semitic tongues. With regard to the Semitic offshoots, six separate alphabets may be discerned: the Hebrew, the Nabataean-Sinaitic-Arabic, the Palmyrene, the Syriac-Nestorian, the Mandaean, and the Manichaean. Some of these alphabets became links between the Aramaic alphabet and the numerous scripts used for the non-Semitic languages of Central, South, and Southeast Asia.

Adaptations of the Aramaic alphabet

Among these scripts, which were directly or mainly indirectly adapted to non-Semitic languages from the Aramaic alphabet, are: (1) the Persian (Iranian) scripts known as Pahlavi, which were used for such writings as sacred (pre-Islāmic) Persian literature; (2) Sogdian, a script and language that constituted the lingua franca of Central Asia in the second half of the 1st millennium AD; (3) Kök Turki, a script used from the 6th to the 8th century AD by Turkish tribes living in the southern part of central Siberia, in northwestern Mongolia, and in northeastern Turkistan (this alphabet was the prototype of the early Hungarian alphabet); (4) the alphabet of the Uighur, a Turkic-speaking people who lived in Mongolia and eastern Turkistan in the early 13th century; this script was adapted, with Tibetan influence, and adopted as the writing of the Mongol Empire (the so-called Kalika script); (5) the early scripts of the Mongols, including Kalmuck, Buryat (Buriat), Mongolian proper, and the allied Manchu alphabet.

The Aramaic alphabet was probably also the prototype of the Brāhmī script of India, a script that became the parent of nearly all Indian writings. Derived from the Aramaic alphabet, it came into being in northwest India. The Armenian and Georgian alphabets, created by St. Mesrob (Mashtots) in the early 5th century AD, were also based on the Aramaic alphabet.

The South Semitic alphabet. The South Semitic, or Sabaean, branch remained within the confines of the Arabian Peninsula for most of its history. It was in use at the beginning of the 1st millennium BC. The most that can be said about its origins is that it neither developed from nor directly depended upon the North Semitic alphabet. It may have been derived, ultimately, from the proto-Sinaitic script, with some influence from the North Semitic. Offshoots from the South Semitic branch include the Minaean, Himyaritic, Qatabanic, and Hadhramautic alphabets in southern Arabia, and Thamudene, Dedanite, and Safaitic alphabets in the northern part of the peninsula. Numerous inscriptions in these alphabets are the principal source for the study of those once-flourishing kingdoms, including Saba' (the biblical Sheba), relegated by the rise of Islām to the backwaters of history.

The Sabaean offshoot, a graceful and elegant script consisting of 29 letters, spread into Africa, where it became the progenitor of the Ethiopic alphabet; this in turn gave birth to the modern Amharic, Tigré, Tigrinya, and other alphabets of modern Ethiopia. These are the only South Semitic scripts still in use today.

The Greek alphabet. As in so many other things, the importance of the ancient Greeks in the history of the alphabet is paramount. All of the alphabets in use in European languages today are directly or indirectly related to the Greek. The Greek achievement was to provide representations for vowel sounds. Consonants plus vowels made a writing system that was both economical and unambiguous (see above *History of writing systems*). The true alphabetic system has remained for 3,000 years, with only slight modifications, an unparalleled vehicle of expression and communication in and among the most diverse nationalities and languages. The Greek alphabet, created early in the 1st millennium BC, spread in various directions in Asia Minor, Egypt, Italy, and other places, but far and away its most important descendants were the Etruscan Latin and the Cyrillic alphabets.

Theories explaining diffusion. There is no complete agreement among scholars as to how or why certain alphabets have come to dominate much of the world. Some believe that diffusion is explained by the efficiency of the orthography; the Greek alphabet, capable of representing unambiguously a full range of meanings, was adopted throughout western Europe. Others hold that the alphabet follows the flag; that is, that the diffusion of an alphabet results from political and military conquests by the people who use it. Still others hold that the alphabet follows trade

The theory that the alphabet follows religion

or religion. A few examples may illustrate the point: (1) The Latin language and script were carried by Roman legionaries and imperial officers to all parts of the vast Roman Empire, particularly to the regions that were not Hellenized. In later centuries, however, churchmen and missionaries carried the Latin language and script still farther afield. The ascendancy of Latin led to the adoption of the Latin (Roman) alphabet by a large majority of nations; it became used for tongues of the most diverse linguistic groups, not only in Europe but in all other parts of the world as well. (2) Two alphabets, the Cyrillic and the Latin, are used for writing Slavic languages. Cyrillic is used by those Slavic peoples who accepted their religion from Byzantium, whereas Roman Christianity brought the use of the Latin alphabet to the Poles, Lusatians, Wends, Czechs, Slovaks, Slovenes, and Croats. Particularly interesting is the case of Yugoslavia, in which a single language is written differently by the Roman Catholic Croats and the Greek Orthodox Serbs. (3) The Arabic alphabet is, after Latin, the most generally used in Asia and Africa. The rise of Islām in the 7th century AD and the tremendous Islāmic expansion and conquest carried the Islāmic holy book, the Qur'ān, written in the Arabic alphabet, over a vast area: the Middle East, North and Central Africa, South and Southeast Asia, and even southern Europe. The Arabic alphabet was, therefore, adapted to Semitic and Indo-European forms of speech, to Tatar-Turkish, Iranian, and Austronesian (Malayo-Polynesian) tongues, and to several African languages. (4) The movement eastward from India of the Indian Brāhmī-Buddhist alphabets was much more peaceful than that of the Arabic alphabet. These offshoots, which took root in Ceylon, Burma, Thailand, Cambodia, Laos, Vietnam, Indonesia, and the Philippines, were again the result of the spreading of a religion—Buddhism—but by missionaries and not by armies.

From D. Diringer, *The Alphabet: A Key to the History of Mankind*

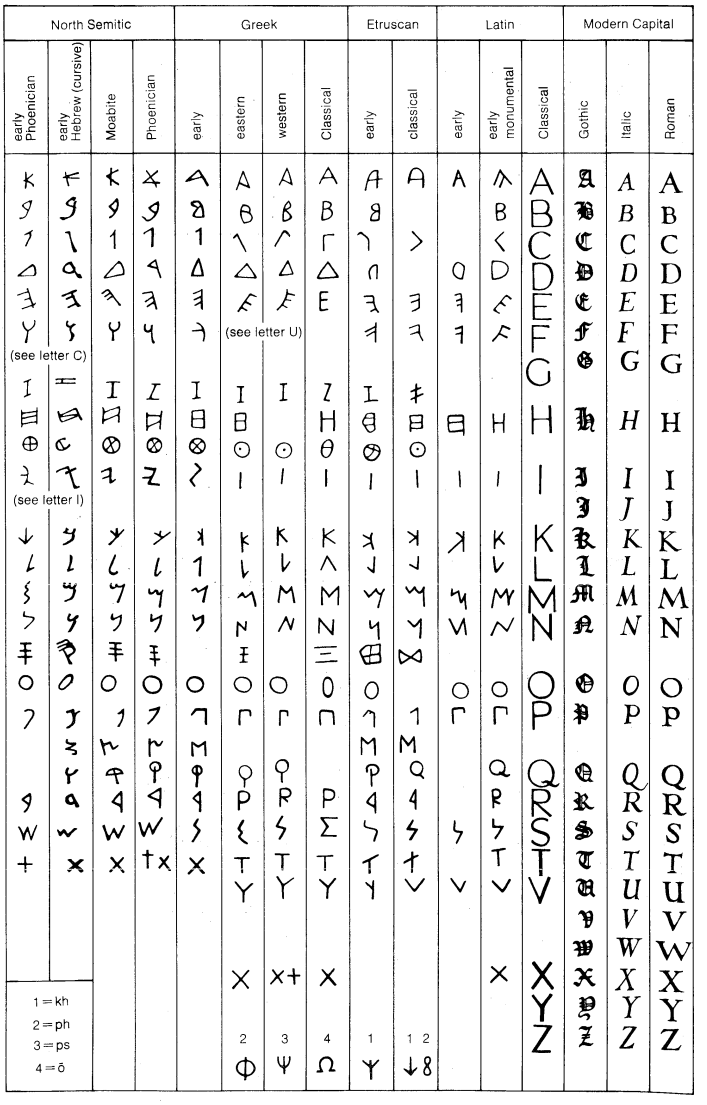


Figure 3: Alphabet characters from North Semitic to modern Roman capital letters. The North Semitic, Greek, and Etruscan symbols do not always designate the same sound as the modern letters.

Major alphabets of the world. *Hebrew alphabet.* It is generally believed, in accordance with Jewish tradition, that the Early Hebrew alphabet was superseded in the Holy Land by the Aramaic alphabet during the period of the Babylonian Exile (586–516 BC) and that the Aramaic script therefore became the parent of the Square Hebrew (in Hebrew *ketav meruba'* ["square script"] or *ketav ashuri* ["Assyrian writing"]). The theory may be only partly right, for in the Holy Land the Early Hebrew alphabet was an object of such local attachment that for several centuries it was used side by side with the Aramaic script.

At any rate, there is little doubt that the Square Hebrew did derive from the Aramaic alphabet. A distinctive Jewish variety of the Aramaic alphabet that can be regarded as the Square Hebrew script can be traced from the 3rd century BC. It became standardized just before the Christian Era, and it was from this script that the modern Hebrew alphabet, in all its styles, eventually developed. The development was gradual and purely external (*i.e.,* in the shapes of the single letters); from the internal standpoint (*i.e.,* considering the phonetic values of the letters), there has been no development, though it must be borne in mind that for several letters (*e.g., waw, ḥet, tzade, qof, shin, sin,* etc.) the exact original phonetic value is still uncertain.

Standardization of Square Hebrew

When the Square Hebrew alphabet became standardized, it took (at least, in its formal style and, much later, in its printed form) the form that, with insignificant changes, it now has. Minute rules laid down by the Talmud made further development of the Square Hebrew impossible.

In the Square Hebrew alphabet there are five letters—*kaf, mem, nun, pe,* and *tzade*—that have dual forms. That is, there is one character for initial or medial position and another for final position.

The Hebrew alphabet consists of 22 letters, all consonants, though four of them—*alef, he, waw,* and *yod*—are also employed to represent long vowels. The absence of vowel letters was not at first a problem, because Hebrew, like other Semitic languages, has consonantal roots, with vowels serving principally to denote inflections in nouns, moods of verbs, and other grammatical variations. As Hebrew speech passed out of daily use (being superseded by Aramaic, which became the vernacular of the Jews) and the knowledge of biblical Hebrew steadily declined, it became necessary to introduce some form of vocalic distinction so that the Bible could be read and explained correctly. The three main vowel systems now extant are the Babylonian, the Palestinian, and the Tiberiadic; of these the latter is the most important and, indeed, the only one still in use. The Tiberiadic system consists of dots, combinations of dots, and small dashes.

Before the discovery of the celebrated Dead Sea Scrolls, several Square Hebrew inscriptions belonging mainly to the 1st century BC and the succeeding centuries were known; they were found in Palestine, Syria, North Africa, and Italy, on rocks, tombs, or ossuaries (depositories for the bones of the dead), in synagogues and catacombs. The biblical manuscripts, except for some fragments written on papyrus, belong to a much later date. The earliest fragment is the Nash papyrus of approximately the 1st century BC, now in the University of Cambridge Library. Many thousands of fragments of Hebrew biblical and other manuscripts, partly of the 7th and 8th centuries AD, were discovered in the Geniza, an archive in the old synagogue in Cairo.

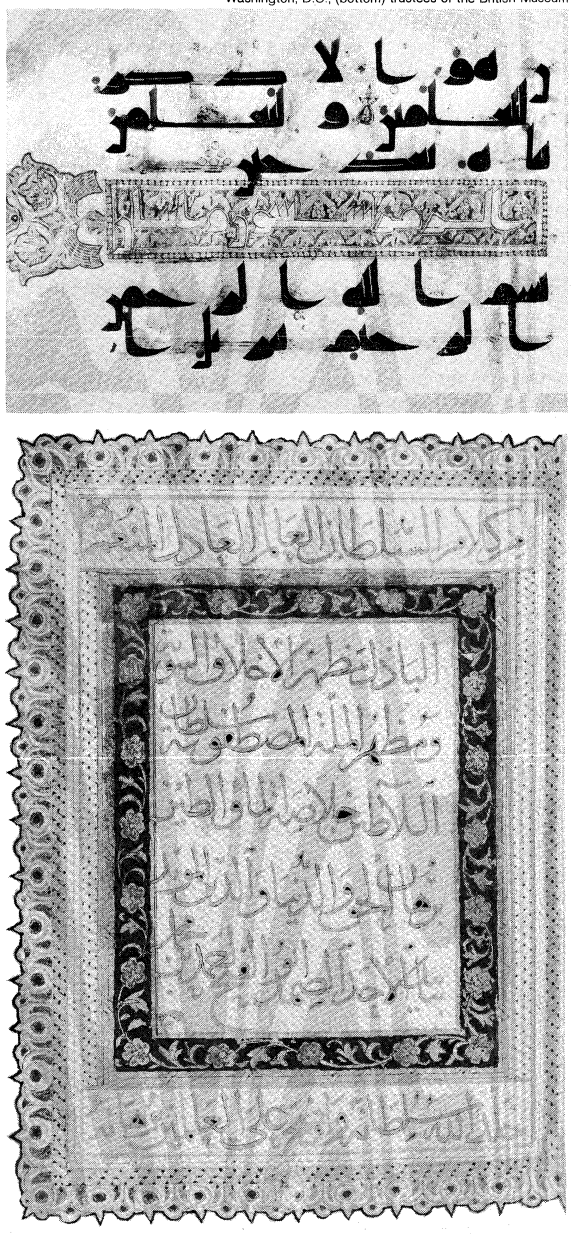
Dead Sea Scrolls

The focus of scholarly interest during the late 1940s and the successive years was the sensational discovery of Hebrew biblical and nonbiblical scrolls in caves near the Dead Sea. The tens of thousands of fragmentary manuscripts, composing what are popularly called the Dead Sea Scrolls, may be divided into several groups, the oldest being a collection of biblical and other Hebrew manuscripts dating approximately from the 3rd century BC.

In the more than bimillenary development of the Square Hebrew alphabet, four fundamental types can be noticed: (1) the square script, which evolved into the well-proportioned printing type of modern Hebrew (the majority of Dead Sea Scrolls are in this Square Hebrew script); (2) the medieval formal styles; (3) the rabbinic, also known as Rashi writing, which was the medieval book or literary hand; and (4) a cursive script or daily handwriting, which gave rise to many local varieties (Oriental, Spanish, Italian, Franco-German, and so on), of which the Polish-German became the current Hebrew handwriting of today. The Hebrew script has been adapted to some other languages, such as Arabic, Turkish (for the Karaite people of Crimea), and so forth, but particularly to German—hence, Yiddish—and Spanish—hence, Judeo-Spanish.

Arabic alphabet. The Arabic script descended from the Aramaic through the Nabataean and the neo-Sinaitic alphabets. After the Latin script, it is the most widely used form of alphabetic writing in the modern world. The Arab conquests of the 7th and 8th centuries AD brought the language and the script to the vast expanse of territory extending from India to the Atlantic Ocean. The Arabic alphabet was adapted, with some necessary modifications, to such diverse languages as the Slavic tongues, Spanish, Persian, Urdu, Turkish, Hebrew, Berber, Swahili, Malay, Sudanese, and others.

By courtesy of the (top) Freer Gallery of Art, Smithsonian Institution, Washington, D.C., (bottom) trustees of the British Museum



Early Arabic writing.
(Top) Early Kūfic book style, leaf from a Qur'ān, 8th–9th century. In the Freer Gallery of Art, Smithsonian Institution, Washington, D.C. (Bottom) Old Ottoman *naskhī*, opening of the Qur'ān, 1394. In the British Museum (MS. OR 4126).

Kūfic and *naskhī* styles

The Arabic alphabet probably originated at some time in the 4th century AD, but the earliest extant Arabic writing is a trilingual inscription—Greek-Syriac-Arabic—of AD 512. The two principal types of Arabic writing, which developed quite early in the Muslim period, were the Kūfic, from the town of Kūfah in Mesopotamia, seat of a famous Muslim academy, and the *naskhī,* or Mecca-Medina script. Kūfic, a heavy, bold, and lapidary style, appeared toward the end of the 7th century AD. It was particularly suitable for writing on stone or metal, for painting or carving inscriptions on the walls of mosques, and for lettering on coins. Its letters are generally thick, squat, and unslanted. With the high development of Arabic calligraphy, Kūfic writing became an exceptionally beautiful script. From it there were derived a number of other styles, chiefly medieval, in North and Central Africa, Spain, and northern Arabia. Thereafter, it was virtually discontinued except for formal and monumental writing. Nevertheless, it was also used for writing precious manuscripts of the Qur'ān, many of which are extant today.

The *naskhī* style was from the very outset a more cursive form. It was always employed chiefly for writing on papyrus. In time, it evolved into innumerable styles and varieties, including the *ta'liq,* the *riqa'*, the *divani,* the *thuluth,* and the *syakat,* and became the parent of the modern Arabic writing.

Like other Semitic scripts, Arabic is written from right to left. Its alphabet contains 28 consonantal letters, 22 being directly derived from the Aramaic-Nabataean branch of the North Semitic alphabet and six being new additions; three of the letters—*alif, wāw,* and *yā'*—are also used as long vowels.

The written letters undergo a slight external change according to their position within a word. When they stand alone or occur at the end of a word, they ordinarily terminate in a bold stroke; when they appear in the middle of a word, they are ordinarily joined to the letter following by a small, upward curved stroke. With the exception of six letters, which can be joined only to the preceding ones, the initial and medial letters are much abbreviated, while the final form consists of the initial form with a triumphant flourish. The essential part of the characters, however, remains unchanged. On the whole, the evolution of the forms of the Arabic letters was the most rapid of all the branches of alphabetic writing.

Vowel marks in Arabic

Although the absence of vowel letters was not strongly felt in Arabic (as in Hebrew and other Semitic languages), for teaching purposes and for correct reading of the Qur'ān, the use of diacritical marks (including signs for short vowels, which are sometimes used in conjunction with the letters *alif, wāw,* and *yā'*) was introduced in Basra in the early 8th century. The practice was probably borrowed from the Syriac script. It not only provides vowel sounds but also distinguishes different consonants; diacritical points are also used as endings in the inflection of nouns and the moods of verbs. These marks—there are three of them—are written above or below the consonants (preceding the vowel), while a sign called *sukūn* indicates the absence of a vowel. Thus, there are, on the whole, a great number of diacritical points; these form a peculiar characteristic of this writing form.

Indian alphabets. The Aramaic alphabet was probably the prototype of the Brāhmī script of India, the ancestor of all Indian scripts. The transmission probably took place in the 7th century BC. Adapting the Aramaic script to the Indo-Aryan tongue of India was by no means simple or straightforward. The shapes of many Brāhmī letters show clear Semitic influence; moreover, the Brāhmī script was originally written from right to left. It is obvious, however, that on the whole it was the idea of alphabetic writing that was transmitted and that the fully developed Brāhmī writing was the outcome of the brilliant philological and phonological elaboration of the scientific Indian school.

During the 5th century BC the second of the prototypal

Indian alphabets—the Kharoṣṭī script—came into being in northwest India (which was then under Persian rule). Although the origin of Brāhmī is still uncertain and hotly discussed, it is commonly accepted that the Kharoṣṭī alphabet is a direct descendant from the Aramaic alphabet. Moreover, the direction of writing in Kharoṣṭī script is as in Aramaic, from right to left, and there is also a likeness of many signs having similar phonetic value.

In the later centuries of its existence, Brāhmī gave birth to eight varieties of script. Three of them—the early and late Maurya and the Śuṅga—became the prototypes of the North Indian subdivision of the Brāhmī script in the 1st centuries BC and AD. Out of this North Indian subdivision there arose the Gupta, a monumental form of writing, which was employed from the 4th to the 6th century AD and became the ancestor of the great majority of Indian scripts.

Gupta writing in India

The western variety of the Gupta spread into eastern (or Chinese) Turkistan, where it was adopted for a number of languages, including the recently discovered Turfanian and Kuchean (Tocharian A and B), and where it strongly influenced the invention or revision of the Tibetan script (AD 639). There were two main offshoots of the Tibetan writing: the 'Phags-pa, adapted to the Chinese and Mongolian languages in 1272; and the Lepchā, which arose in the beginning of the 18th century.

From D. Diringer, *The Alphabet: A Key to the History of Mankind*

phonetic value	Brāhmī: Aśoka	Brāhmī: Bhattiprōlu	Brāhmī: Śuṅga	Brāhmī: Kuṣāna	Brāhmī: Kṣatrapa	Gupta	modern north Indian: Devanāgarī	modern north Indian: Gurmukhi	modern north Indian: Gujarātī	modern north Indian: Marāthī	modern north Indian: Sindhī	modern north Indian: Multanī	main n.e. Indian: Bengali	main n.e. Indian: Assamese	main n.e. Indian: Oriya	south Indian: Vijayanagar	south Indian: Telugu	south Indian: Kannada	south Indian: Tamil	south Indian: Malayalam
a	𑀅	𑀅	𑀅	𑀅	𑀅	𑀅	अ	ਅ	અ	अ	ॲ	ਵ	অ	অ	ଅ	ಅ	అ	ಅ	அ	അ
ā	𑀆	𑀆	𑀆	𑀆	𑀆	𑀆	आ	ਆ	આ	आ	آ	ਾ	আ	আ	ଆ	ಆ	ఆ	ಆ	ஆ	ആ
i	𑀇	𑀇	𑀇	𑀇	𑀇	𑀇	इ	ਇ	ઇ	इ	ا	ਿ	ই	ই	ଇ	ಇ	ఇ	ಇ	இ	ഇ
ī	𑀈	𑀈	𑀈	𑀈	𑀈	𑀈	ई	ਈ	ઈ	ई	ا	ੀ	ঈ	ঈ	ଈ	ಈ	ఈ	ಈ	ஈ	ഈ
u	𑀉	𑀉	𑀉	𑀉	𑀉	𑀉	उ	ਉ	ઉ	उ	ا	ੁ	উ	উ	ଉ	ಉ	ఉ	ಉ	உ	ഉ
ū	𑀊	𑀊	𑀊	𑀊	𑀊	𑀊	ऊ	ਊ	ઊ	ऊ	ا	ੂ	ঊ	ঊ	ଊ	ಊ	ఊ	ಊ	ஊ	ഊ
ṛ	𑀋	𑀋	𑀋	𑀋	𑀋	𑀋	ऋ	਋	ઋ	ऋ	ا	੃	ঋ	ঋ	ଋ	ಋ	ఋ	ಋ	஋	ഋ
ṝ	𑀌	𑀌	𑀌	𑀌	𑀌	𑀌	ॠ	਌	ૠ	ॠ	ا	੄	ৠ	ৠ	ୠ	ೠ	ౠ	ೠ	௠	ൠ
ḷ	𑀍	𑀍	𑀍	𑀍	𑀍	𑀍	ऌ	਍	ઌ	ऌ	ا	੅	ঌ	ঌ	ଌ	ಌ	ఌ	ಌ	஌	ഌ
ḹ	𑀎	𑀎	𑀎	𑀎	𑀎	𑀎	ॡ	਎	ૡ	ॡ	ا	੆	ৡ	ৡ	ୡ	ೡ	ౡ	ೡ	௡	ൡ
e	𑀏	𑀏	𑀏	𑀏	𑀏	𑀏	ए	ਏ	એ	ए	ا	ੇ	এ	এ	ଏ	ಎ	ఎ	ಎ	எ	എ
ai	𑀐	𑀐	𑀐	𑀐	𑀐	𑀐	ऐ	ਐ	ઐ	ऐ	ا	ੈ	ঐ	ঐ	ଐ	ಐ	ఐ	ಐ	ஐ	ഐ
o	𑀑	𑀑	𑀑	𑀑	𑀑	𑀑	ओ	ਓ	ઓ	ओ	ا	ੋ	ও	ও	ଓ	ಒ	ఒ	ಒ	ஒ	ഒ
ō	𑀒	𑀒	𑀒	𑀒	𑀒	𑀒	ऑ	ਔ	ઑ	ऑ	ا	ੌ	ঔ	ঔ	ଔ	ಓ	ఓ	ಓ	ஓ	ഓ
au	𑀓	𑀓	𑀓	𑀓	𑀓	𑀓	औ	ਕ	ક	औ	ا	◌੍	ক	ক	କ	ಕ	క	ಕ	க	ക
ka	𑀔	𑀔	𑀔	𑀔	𑀔	𑀔	क	ਖ	ખ	क	ا	ਖ	খ	খ	ଖ	ಖ	ఖ	ಖ	க	ഖ
kha	𑀕	𑀕	𑀕	𑀕	𑀕	𑀕	ख	ਗ	ગ	ख	ا	ਗ	গ	গ	ଗ	ಗ	గ	ಗ	ங	ഗ
ga	𑀖	𑀖	𑀖	𑀖	𑀖	𑀖	ग	ਘ	ઘ	ग	ا	ਘ	ঘ	ঘ	ଘ	ಘ	ఘ	ಘ	஘	ഘ
gha	𑀗	𑀗	𑀗	𑀗	𑀗	𑀗	घ	ਙ	ઙ	घ	ا	ਙ	ঙ	ঙ	ଙ	ಙ	ఙ	ಙ	ங	ങ
ña	𑀘	𑀘	𑀘	𑀘	𑀘	𑀘	ङ	च	ચ	ङ	ا	ਚ	চ	চ	ଚ	ಚ	చ	ಚ	ச	ച
ca	𑀙	𑀙	𑀙	𑀙	𑀙	𑀙	च	छ	છ	च	ا	ਛ	ছ	ছ	ଛ	ಛ	ఛ	ಛ	஛	ഛ
cha	𑀚	𑀚	𑀚	𑀚	𑀚	𑀚	छ	ज	જ	छ	ا	ਜ	জ	জ	ଜ	ಜ	జ	ಜ	ஜ	ജ
ja	𑀛	𑀛	𑀛	𑀛	𑀛	𑀛	ज	झ	ઝ	ज	ا	ਝ	ঝ	ঝ	ଝ	ಝ	ఝ	ಝ	஝	ഝ
jha	𑀜	𑀜	𑀜	𑀜	𑀜	𑀜	झ	ञ	ઞ	झ	ا	ਞ	ঞ	ঞ	ଞ	ಞ	ఞ	ಞ	ஞ	ഞ
ña	𑀝	𑀝	𑀝	𑀝	𑀝	𑀝	ञ	ट	ટ	ञ	ا	ਟ	ট	ট	ଟ	ಟ	ట	ಟ	ட	ട
ṭa	𑀞	𑀞	𑀞	𑀞	𑀞	𑀞	ट	ठ	ઠ	ट	ا	ਠ	ঠ	ঠ	ଠ	ಠ	ఠ	ಠ	ட	ഠ
ṭha	𑀟	𑀟	𑀟	𑀟	𑀟	𑀟	ठ	ड	ડ	ठ	ا	ਡ	ড	ড	ଡ	ಡ	డ	ಡ	ட	ഡ
ḍa	𑀠	𑀠	𑀠	𑀠	𑀠	𑀠	ड	ढ	ઢ	ड	ا	ਢ	ঢ	ঢ	ଢ	ಢ	ఢ	ಢ	ட	ഢ
ḍha	𑀡	𑀡	𑀡	𑀡	𑀡	𑀡	ढ	ण	ણ	ढ	ا	ਣ	ণ	ণ	ଣ	ಣ	ణ	ಣ	ண	ണ
ṛa							ड़	ੜ		ड़			ড়	ড়	ଡ଼					
ṛha							ढ़			ढ़			ঢ়	ঢ়	ଢ଼					
ṇa	𑀢	𑀢	𑀢	𑀢	𑀢	𑀢	ण	त	ત	ण	ا	ਤ	ত	ত	ତ	ತ	త	ತ	த	ത
ta	𑀣	𑀣	𑀣	𑀣	𑀣	𑀣	त	थ	થ	त	ا	ਥ	থ	থ	ଥ	ಥ	థ	ಥ	ந	ഥ
tha	𑀤	𑀤	𑀤	𑀤	𑀤	𑀤	थ	द	દ	थ	ا	ਦ	দ	দ	ଦ	ದ	ద	ದ	த	ദ
da	𑀥	𑀥	𑀥	𑀥	𑀥	𑀥	द	ध	ધ	द	ا	ਧ	ধ	ধ	ଧ	ಧ	ధ	ಧ	த	ധ
dha	𑀦	𑀦	𑀦	𑀦	𑀦	𑀦	ध	न	ન	ध	ا	ਨ	ন	ন	ନ	ನ	న	ನ	ந	ന
na	𑀧	𑀧	𑀧	𑀧	𑀧	𑀧	न	ऩ	ન઼	न	ا	ਨ਼	ন়	ন়	ନ଼	ನ಼	న಼	ನ಼	ன	ഩ
pa	𑀨	𑀨	𑀨	𑀨	𑀨	𑀨	प	प	પ	प	ا	ਪ	প	প	ପ	ಪ	ప	ಪ	ப	പ
pha	𑀩	𑀩	𑀩	𑀩	𑀩	𑀩	फ	फ	ફ	फ	ا	ਫ	ফ	ফ	ଫ	ಫ	ఫ	ಫ	ஃப	ഫ
ba	𑀪	𑀪	𑀪	𑀪	𑀪	𑀪	ब	ब	બ	ब	ا	ਬ	ব	ব	ବ	ಬ	బ	ಬ	ப	ബ
bha	𑀫	𑀫	𑀫	𑀫	𑀫	𑀫	भ	भ	ભ	भ	ا	ਭ	ভ	ভ	ଭ	ಭ	భ	ಭ	ம	ഭ
ma	𑀬	𑀬	𑀬	𑀬	𑀬	𑀬	म	म	મ	म	ا	ਮ	ম	ম	ମ	ಮ	మ	ಮ	ம	മ
ya	𑀭	𑀭	𑀭	𑀭	𑀭	𑀭	य	य	ય	य	ا	ਯ	য	য	ଯ	ಯ	య	ಯ	ய	യ
ẏa							य़	ਯ਼		य़			য়	য়	ୟ					
ra	𑀮	𑀮	𑀮	𑀮	𑀮	𑀮	र	र	ર	र	ا	ਰ	র	র	ର	ರ	ర	ರ	ர	ര
ṛa							ऱ			ऱ										
la	𑀯	𑀯	𑀯	𑀯	𑀯	𑀯	ल	ਲ	લ	ल	ا	ਲ	ল	ল	ଲ	ಲ	ల	ಲ	ல	ല
ḷa							ळ	ਲ਼		ळ			ল়	ল়	ଳ					
va	𑀰	𑀰	𑀰	𑀰	𑀰	𑀰	व	व	વ	व	ا	ਵ	ব	ব	ବ	ವ	వ	ವ	வ	വ
śa	𑀱	𑀱		𑀱	𑀱	𑀱	श	श	શ	श	ا	ਸ਼	শ	শ	ଶ	ಶ	శ	ಶ	ஶ	ശ
ṣa	𑀲	𑀲	𑀲	𑀲	𑀲	𑀲	ष	ष	ષ	ष	ا	ਸ਼	ষ	ষ	ଷ	ಷ	ష	ಷ	ஷ	ഷ
sa	𑀳	𑀳	𑀳	𑀳	𑀳	𑀳	स	स	સ	स	ا	ਸ	স	স	ସ	ಸ	స	ಸ	ஸ	സ
ha	𑀴	𑀴	𑀴	𑀴	𑀴	𑀴	ह	ਹ	હ	ह	ا	ਹ	হ	হ	ହ	ಹ	హ	ಹ	ஹ	ഹ

Diacritical marks are (ṃ) anusvāra, ◌̐ (m̐) anunāsika, (ḥ) visarga,)((ḫ) jihvāmūlīya, ≍ (ḥ) upadhmānīya, and S (') avagraha.
These are often used with the script and affect pronunciation.

Figure 4: Comparison of several Indian scripts.

From the Devanāgarī writing as used in eastern India in the 11th century there developed the proto-Bengali and the early Nepali, or Newari, scripts, from which the many scripts employed nowadays in northern India and Bangladesh descended (*e.g.*, the Bengali, Oriya, Manipuri, Assamese, Gujarati, and Bihari scripts, and the various Eastern Hindi local scripts).

In northwestern India several other scripts are employed. The Sarada script, a descendant of the western type of the Gupta character, originated in the 8th century and is still employed for Kashmiri. In addition, there are the several varieties of the Takri, used over the lower ranges of the Himalayas north of Punjab; the Dogri, used for a dialect of Punjabi; the Landa, the national alphabet of Punjabi, which has many varieties and is used mainly by shopkeepers all over Punjab and Sindh; and the Gurmukhi script, the characters of the Sikh scriptures.

In South India, which is inhabited by peoples speaking Dravidian languages, several other scripts are used, of which the Kannada, or Kanarese, the Telugu, the Grantha, the Tulu-Malayalam, the Tamil, and the Vaṭṭeḻuttu are the most important.

Long before the existence of the Gupta script, the Brāhmī script had already begun its eastward movement. The Indo-Aryan migration in the 5th century BC to the island now known as Sri Lanka had set the stage there, and the earliest Brāhmī inscriptions in Sri Lanka can be dated to the 3rd century BC. Most dramatic of all, however, was the expansion of Buddhism from India into what are now Sri Lanka, Burma, Thailand, Kampuchea, Laos, Vietnam, Malaysia, and Indonesia. As already mentioned, unlike the conquests of Islām, this was a peaceful movement; its "soldiers" were Buddhist monks, political independents who built an empire founded on the cultural and spiritual community of peoples. Among their many achievements, these monks brought into being offshoots from the Brāhmī script, principally from its South Indian varieties, throughout the vast extent of territory from India itself to the Philippines. Thus arose the many scripts of Southeast Asia, from the Cham writing of Kampuchea to the Kavi character of Java and its Sumatran offshoots, and the Tagalog writing of the Philippines.

Spread of Brāhmī to Southeast Asia

All these Indian and Southeast Asian scripts involve types of semi-syllabaries rather than alphabets. They consist of vowels and diphthongs and basic consonants (*i.e.*, consonants followed by a short *a*); there are no pure consonants (*i.e.*, consonants written by themselves).

Greek alphabet. The Greek alphabet derived from the North Semitic script in the 8th century BC. The direction of writing in the oldest Greek inscriptions—as in the Semitic scripts—is from right to left, a style that was superseded by the boustrophedon (meaning, in Greek, "as the ox draws the plow"), in which lines run alternately from right to left and left to right. This change occurred approximately in the 6th century BC. There are, however, some early Greek inscriptions written from left to right, and after 500 BC Greek writing invariably proceeded from left to right.

The letters for *b, g, d, z, k, l, m, n, p, r,* and *t,* which are sounds common to the Semitic and Greek languages, were taken over without change. The principal Greek change arose in applying a script developed to represent

a Semitic language, in which vowel sounds were of minor importance to the identity of a word, to a language in which such vowel differences are crucial to the identity of a word. In Greek, /kat/, /kit/, and /kot/ are entirely different words, while in Semitic languages they would be the same word in different grammatically inflected forms. The Greek addition of vowels to the alphabet to make it an analogue of the sound pattern produced a writing system that was both manageable and accurate (see above *History of writing systems*). The different ways in which these adaptations were carried out allow the two main branches of the early Greek alphabet—the eastern and the western—to be distinguished. These again subdivided, each into secondary branches. Within this general grouping there were many local peculiarities, but the differences between all of these local alphabets involved variations in detail rather than essential structure.

Ionic and other eastern Greek alphabets

The eastern and western subdivisions were the two principal branches of the early Greek alphabet. The Ionic alphabet was the most important of the eastern variety, which also included the Greek alphabets of Asia Minor and the adjacent islands, of the Cyclades and Attica, of Sicyon and Argos, of Megara, Corinth, and the Ionian colonies of Magna Graecia. A secondary branch of the eastern subdivision was made up of the alphabets used on the Dorian islands of Thera, Melos, and Crete. The alphabets of Euboea (Chalcidian), Boeotia, Phocis, Locris, Thessaly, the Peloponnesus (except its northeastern part), and of the non-Ionian colonies of Magna Graecia belonged to the western subdivision. It is a controversial point whether the eastern or the western branch was the earlier in time, whether there was any derivative link between one and the other, or whether they represent two quite independent adaptations of the Semitic alphabet. The latter alternative seems rather improbable.

Gradually, the Greek local alphabets became more and more similar. In 403 BC the Ionic alphabet of Miletus was officially adopted in Athens and later also in the other states. By the middle of the 4th century BC almost all the local alphabets had been replaced by the Ionic, which became the common, classical Greek alphabet of 24 letters.

After this time the development of the Greek alphabet was almost wholly external, in the direction of greater utility, convenience, and, above all, beauty. The classical style was retained as a monumental script at the same time that more cursive forms grew up for writing on such surfaces as parchment, papyrus, and wax. The classical letters were also retained as the capital letters in the modern print (though some of the capitals in modern Greek handwriting are borrowed from the Latin alphabet). On the other hand, the classical Greek alphabet also evolved into the Greek uncials, the cursive, and the minuscule script. (Uncial letters were somewhat rounded and separated versions of capital letters or cursive forms; minuscule letters developed from cursive writing and have simplified, small forms.) Until about AD 800 the uncials were used as a book hand; later, the minuscule script was employed for the same purpose. The cursive scripts evolved into the modern Greek minuscule.

Accent marks in Greek

In the middle of the 3rd century BC the Greek scholar Aristophanes of Byzantium introduced the three accents—acute, grave, and circumflex—that were thereafter used to assist students, particularly foreigners, in the correct pronunciation of Greek words; these continue to be used in most Greek texts printed today. Originally, these marks indicated tone or pitch, not stress.

Countless inscriptions have been discovered all over the Hellenic and Hellenistic world and beyond. They include official decrees, annals, codes of law, lists of citizens, civic rolls, temple accounts, votive offerings, ostraca (fragments of pottery), sepulchral inscriptions, coins, lettering on vases, and so forth. These, along with many thousands of Greek manuscripts, both ancient and medieval, serve as sources for the studies known as Greek epigraphy and Greek paleography and are of untold importance for all branches of ancient history, philology, philosophy, and other disciplines.

The most direct offshoots from the Greek alphabet were those adapted to the languages of the non-Hellenic peoples of western Asia Minor in the 1st millennium BC: the scripts of the Lycians, Phrygians, Pamphylians, Lydians, and Carians. The first three of these were derived directly from the Greek; the Lydian and Carian were strongly influenced by it. The Coptic alphabet was the other non-European offshoot from the Greek and the only one used in Africa. Twenty-four of its 31 letters were borrowed from the Greek uncial writing, and seven were taken over from a particularly cursive variety of the Egyptian demotic writing; the demotic letters were used to express Coptic sounds not existing in the Greek language.

More significant, however, were the European offshoots. In Italy, two alphabets derived directly from the Greek: the Etruscan and the Messapian (Messapic). The Messapii were an ancient tribe who inhabited the present Apulia (in southern Italy) in pre-Roman times; their language is presumed to belong to the Illyrian group. Over 200 Messapian inscriptions have been discovered. In southeastern Europe there were three offshoots from the Greek alphabet—the Gothic, the Cyrillic, and the Glagolitic alphabets.

The Gothic alphabet

The Gothic alphabet, not to be confused with the so-called Gothic script (a variety of the Latin alphabet), was a script created by the Gothic bishop Ulfilas (or Wulfila), who died *c.* AD 382. The script consisted of 27 letters, of which some 19 or 20 were taken over from the Greek uncial script. Ulfilas translated the Bible into Gothic; of this translation, some fragments are extant in manuscripts of the 5th and 6th centuries. The most important manuscript is the Codex Argenteus, preserved in Uppsala, Swed.

Cyrillic and Glagolitic alphabets. The two early Slavic alphabets, the Cyrillic and the Glagolitic, were invented by St. Cyril, or Constantine (*c.* 827–869), and St. Methodius (*c.* 825–884). These men were Greeks from Thessalonica who became apostles to the southern Slavs, whom they converted to Christianity. An early tradition, in attributing the invention of an early Slavic writing to Cyril, does not indicate whether his contribution was the Cyrillic or the Glagolitic. It is just possible that both alphabets were invented by him. The earliest dated Old Slavic documents belong to the late 10th and the 11th centuries. The Cyrillic and the Glagolitic alphabets differed widely in the form of their letters, in the history of their development, and partly also in the number of the letters, but they were alike in representing adequately the many sounds of Slavic.

The Cyrillic alphabet was based on the Greek uncial writing of the 9th century. It originally had a total of 43 letters; the two Hebrew letters *tzade* and *shin* were transformed into the Cyrillic letters for the sounds *ch, sh,* and *shch.* The modern forms of this alphabet have fewer letters. Glagolitic writing consisted of 40 letters, externally very unlike either the Greek or Cyrillic scripts.

Cyrillic became, with slight modification in each case, the national script of the Bulgarians, Serbs, Russians, Belorussians, and Ukrainians. (The other Slavic peoples—the Slovenes, Croats, Czechs, Slovaks, Wends, Lusatians, and Poles—use the Latin alphabet.) As mentioned above, in Yugoslavia a single language is written in Cyrillic by the Greek Orthodox Serbs and in the Latin alphabet by the Roman Catholic Croats. For a time, Cyrillic was also adapted to the Romanian language, and in recent times, through the medium of Russian script, it became the writing of a number of Finno-Ugric languages (Komi, Votyak, Mordvinian, Vogul, Ostyak, etc.), Turco-Tatar languages (Chuvash, Turkmenian, Azerbaijanian, etc.), Iranian languages (Ossetic, Kurdish, Tajiki), and Caucasian languages (Abkhaz, Circassian, Avar, etc.).

The history of the Glagolitic alphabet is particularly connected with the religious history of the Slavic peoples of southwest central Europe and the western Balkan Peninsula. In the second half of the 9th century it was introduced, together with the Slavonic liturgy, into the Moravian kingdom, but with the banning of this liturgy by the pope it disappeared from Moravia. It was, however, accepted (also with the Slavonic liturgy) in Bulgaria and Croatia and spread along the Dalmatian coast southward into Montenegro and westward into Istria. Although the Glagolitic script soon disappeared among the Greek Orthodox Slavic peoples because of the victory of the Cyrillic, it continued, notwithstanding the opposition of the higher

Roman Catholic authorities, to be employed among the Roman Catholics of the western Balkan Peninsula together with the Slavonic liturgy and finally succeeded in obtaining the special license of the pope. It is still employed in the Slavonic liturgy in some Dalmatian and Montenegrin communities; the inhabitants of these places are the only Roman Catholics to use the Slavonic liturgy. The earliest preserved Glagolitic secular document dates from 1309. Glagolitic had a short flourishing period in the 16th and 17th centuries.

Etruscan alphabet. The Etruscans, a highly civilized people who were the ancestors of the modern Tuscans and the predecessors of the Romans, inhabited what is now Tuscany in central Italy; their language, still mainly undeciphered, has come down in more than 11,000 inscriptions, the earliest being the 8th-century-BC Marsiliana Tablet, preserved in the Museo Archeologico in Florence. This is also the earliest preserved record of a Western alphabet. The early Etruscan alphabet, unlike any early Greek alphabet found in the Greek inscriptions, contains the original—the prototype—Greek alphabet, consisting of the 22 North Semitic letters, with the phonetic values given to them by the Greeks, and the four additional Greek letters at the end of the alphabet. The Etruscans introduced various changes in their script, and several features in the modern alphabets can be attributed to the influence of the ancient Etruscans. An example is the phonetic value of /k/ for the letters *c, k,* and *q.* Like the Semitic and the early Greek alphabets, Etruscan writing nearly always reads from right to left, though a few inscriptions are in boustrophedon style. The probable date of the origin of the Etruscan alphabet is the late 9th or early 8th century BC.

About 400 BC the "classical" Etruscan alphabet took its final form of 20 letters—four vowels and 16 consonants. Because the voiced and voiceless sounds *b* and *p, d* and *t, g* and *k* were not differentiated in the Etruscan language, the letters *b* and *d* never appear in pure Etruscan inscriptions, and after the disappearance of *k* and *q,* the letter *C* was employed for *g* and *k.*

The Etruscan alphabet had many varieties and several offshoots. Among the offshoots, apart from the Latin, were many alphabets used by Italic populations of pre-Roman Italy and by non-Italic tribes (*e.g.,* the Piceni).

Latin alphabet. The adaptation of the Etruscan alphabet to the Latin language probably took place some time in the 7th century BC. From this century there is a gold brooch known as the Praeneste Fibula (preserved in the Museo Preistorico Etnografico Luigi Pigorini in Rome). The inscription, written in an early form of Latin, runs from right to left and reads clearly: *manios: med:fhefhaked-:numasioi,* which in classical Latin is *Manius me fecit Numerio* ("Manius made me for Numerius").

Oldest known record of Latin alphabet

Dating from the end of the 7th or the beginning of the 6th century BC is a famous cippus (small pillar) from the Roman Forum; it is inscribed vertically on its four faces, in boustrophedon style. Another inscription, probably of the 6th century BC, is known as that of the Duenos Vase and was found in Rome, near the Quirinal Hill. It is also written from right to left. Some Sabine inscriptions belong to the 5th or the 4th century BC. There are also a few inscriptions belonging to the 3rd and 2nd centuries BC.

The Roman capital letters, a form of writing that was used under the empire with unparalleled effectiveness for monumental purposes, became a byword for precision and grandeur, despite a very unprepossessing beginning. Indeed, for the first six centuries of its existence, Roman writing was relatively unimpressive. Only with the advent of the 1st century BC were there signs of magnificence to come.

An opinion that used to be commonly held, and still is held by many, is that the Latin alphabet was derived directly from the Greek in a form used by Greek colonists in Italy. The theory rested on an assertion that the Latin alphabet corresponds to the Chalcidian variety of the western group of Greek scripts employed at Cumae in Campania, southern Italy. This theory is unlikely; indeed, as already mentioned, the Etruscan alphabet was the link between the Greek and the Latin. For instance, the most interesting feature in the inscription of the Praeneste Fibula is the device of combining the letters *f* and *h* to represent the Latin sound of *f.* This was one of the Etruscan ways of representing the same sound. Also, most of the Latin letter names, such as *a, be, ce, de* for the Greek *alpha, beta, gamma, delta,* and so on, were taken over from the Etruscans.

Etruscan as link between Greek and Latin

Runic and ogham alphabets. Runes, in all their varieties, may be regarded as the "national" script of the ancient North Germanic tribes. The origin of the name rune (or runic) is probably related to the fact that the ancient Germanic tribes, like all primitive peoples, attributed magic powers to the mysterious symbols scratched on armour, jewels, tombstones, and so forth. This is given credence by two related Germanic forms that mean "mystery, secret, secrecy": the Old Germanic root *ru-* and the Gothic *runa.* The most interesting runic inscriptions are those that were cut for magical purposes and those that appeal to deities.

The origin of the runes offers many difficult problems and has been hotly argued by scholars and others. The theory of the *Urrunen* (forerunners of the runes), a supposed prehistoric north Germanic alphabetic script, holds that it is the parent not only of the runes but also of all the Mediterranean alphabets, including the Phoenician. This belief, based on racial and political grounds, need not be seriously considered. Some scholars propounded the 6th century BC Greek alphabet as the prototype of the runes; others have suggested the Greek cursive alphabet of the last centuries BC. Several eminent scholars have proposed the Latin alphabet as the source of the runes. The most probable theory, supported recently by many scholars, is that the runic script derived from a North Etruscan, Alpine alphabet. In that case, it is very probable that it originated about the 2nd century BC or a little later.

Origin of the runes

It is still unknown whether the runes were originally employed mainly for magical purposes, as suggested by the name *runa,* or as a usual means of communication. The earliest extant runic inscriptions, numbering over 50, come from Denmark and Schleswig and date from the 3rd to the 6th century AD. About 60 inscriptions from Norway date from the 5th to the 8th century, slightly later than the continental ones. There are also about 50 Anglo-Saxon runic inscriptions extant, including the Franks Casket (about AD 650–700); the right side of the casket is in the Museo Nazionale del Bargello, in Florence, and the rest is in the British Museum. The largest number of inscriptions, about 2,500, come from Sweden; most of these date from the 11th and 12th centuries AD.

There is no certain evidence of wide literary use of runes in early times, but some scholars hold that the runic writing was widely employed for all kinds of secular documents, such as legal provisions, contracts, genealogies, and poems. The known manuscripts are, however, rare and relatively late. The gradual displacement of the runes coincided with the increasing influence of the Roman Catholic Church. The runic scripts lingered on for a long time after the introduction of Christianity, however; indeed, the use of runes for charms and memorial inscriptions lasted into the 16th or even the 17th century.

The ogham alphabet was restricted to the Celtic population of the British Isles. There are over 375 known inscriptions: 316 of them have been discovered in Ireland, chiefly in the southern counties, with only 55 from the northern counties; 40 inscriptions have been discovered in Wales; two come from Devon; and one from Cornwall. One inscription was discovered at Silchester in southern England. About 10 come from the Isle of Man, and a few from Scotland. The Welsh inscriptions are usually bilingual, Latin-Celtic. With one exception, the Irish records are in ogham alone. Most peculiar is the runic-oghamic inscription from the Isle of Man (the runes being a kind of "secret" writing and the oghams being a cryptic script). The distribution of the ogham inscriptions, combined with their language and grammatical forms, point to South Wales or southern Ireland as their place of origin and to the 4th century AD as the date of their origin.

The ogham character was used for writing messages and letters (generally on wooden staves), but sometimes it was

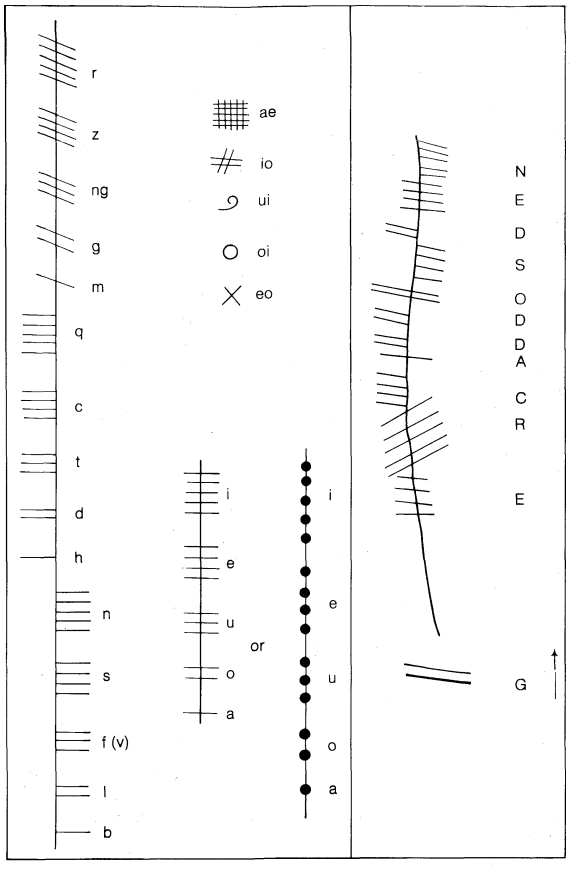


Figure 5: (Left) Ogham alphabet. (Right) Pictish ogham.
From D. Diringer, *The Alphabet: A Key to the History of Mankind*

also written on shields or other hard material and was employed for carving on tombstones. The oghams formed a cryptic script, and there were several varieties, such as wheel oghams, bird oghams, tree oghams, hill oghams, church oghams, colour oghams, and others. The main ogham alphabet consisted of 20 letters represented by straight or diagonal strokes, varying in number from one to five and drawn or cut below, above, or right through horizontal lines, or else drawn or cut to the left, right, or directly through vertical lines. The ogham alphabet was divided into four groups (*aicme*), each containing five letters. Oghams were employed during the Middle Ages; the 14th-century *Book of Ballymote* reproduces the earliest keys for translation. In many cases the ogham inscriptions run upward.

Pictish oghams

Several ogham inscriptions known as the Pictish oghams were found in western Scotland, on the small island of Gigha off the western coast, in Argyll, in northeastern Scotland, and on the northern isles, such as the Shetland Islands. They either belong to the same type as the Irish and Welsh oghams or are written in another ogham variety.

Later development of the Latin alphabet. As already mentioned, the original Etruscan alphabet consisted of 26 letters, of which the Romans adopted only 21. They did not retain the three Greek aspirate letters (*theta, phi,* and *chi*) in the alphabet because there were no corresponding Latin sounds but did employ them to represent the numbers 100, 1,000, and 50. Of the three Etruscan *s* sounds, the Romans kept what had been the Greek *sigma.* The symbol that represented the aspirate later received the shape *H* as it did in Etruscan. *I* was the sign both of the vowel *i* and the consonant *j. X* was added later to represent the sound *x* and was placed at the end of the alphabet. At a later stage, after 250 BC, the seventh letter, the Greek *zeta,* was dropped because Latin did not require it, and a new letter, *G,* made by adding a bar to the lower end of *C,* was placed in its position.

After the conquest of Greece in the 1st century BC, a large number of Greek words were borrowed by the Latin language. At that time the symbols *Y* and *Z* were adopted from the contemporary Greek alphabet, but only to transliterate Greek words; hence, they do not appear in normal Latin inscriptions. They were placed at the end of the alphabet, and the Latin script thus became one of 23 symbols.

A few permanent additions or, rather, differentiations from existing letters occurred during the Middle Ages, when the signs for *u* and *v,* and *i* and *j,* previously written interchangeably for either the vowel or the consonant sound, became conventionalized as *u* and *i* for vowels and *v* and *j* for consonants. *W* was introduced by Norman scribes to represent the English sound *w* (a semivowel) and to differentiate it from the *v* sound.

Derivation of majuscules and minuscules from Latin alphabet

The connection of the capital letters of modern writing with the ancient Semitic-Greek-Etruscan-Latin letters is evident even to a layman. The connection of the minuscules (*i.e.,* the small letters) with the ancient Latin letters is not as evident, but in fact both the majuscules and the minuscules descended from the same ancient Latin alphabet. The different shapes of the small letters are the result of a transformation of the ancient letters by the elimination of a part of the letter—as, for instance, *h* from *H* or *b* from *B*—or by lengthening a part of it—for instance, *d* from *D.* Moreover, the change of the Latin writing into the modern script was induced by the nature of the tool, primarily the pen, and the material of writing, mainly papyrus and parchment, and, from the 14th century onward, also paper. It was the pen, with its preference for curves, that eliminated the angular forms; it was the papyrus, and still more the parchment or vellum, and, in modern times, paper, that made these curves possible.

In ancient times the minuscule did not exist, but there were several varieties of the capital and the cursive scripts. There were three varieties of the capitals: the lapidary capitals (used mainly on stone monuments); the elegant book capitals, somewhat rounded in shape; and the rustic capitals, which were less carefully elaborated than the lapidary script and not as round as the book capitals but

modern European	Latin	runes	Greek	Cyrillic	Glagolitic
A	A		A	А	ⰰ
B	B	ᛒ	B	Б(v) В(b)	ⰱ ⰲ
C	Ↄ	ᚲ ᛦ ᚴ ᚵ	Γ Λ	Г	ⰳ
D	D	ᛞ ᛑ ᛑ (ᛗ)	D Δ ⊗	Д	ⰴ
E	Ɛ	ᛖ ᚿ ᛈ ᚿ	F	Е	ⰵ
F	F	ᚠ ᚡ ᚹ ᚹ	Φ F	Ф	ⰶ
G	G	ᚷ ᛜ			
H	H	ᚺ ᚺ ᚺ ᚺ	H	Н (i)	ⰷ ⰸ
I	I	ᛁ	I	І (i)	ⰹ ⰻ (ⰺ)
J					
K	K	ᛉ	K	К	ⰽ
L	L	ᛚ	Λ	Л	ⰾ
M	M	ᛗ	M	М	ⰿ
N	N	ᚾ ᚾ ᚾ ᚾ	N	Н	ⱀ
O	O	ᛟ ᛟ ᛟ	O O	О	ⱁ
P	⅂	ᚴ M (ᛇ)	Γ	П	ⱂ
Q	Q		Q		
R	ꟼ	ᚱ ᚱ	P	Р	ⱃ
S	ꙅ	ᛊ	Σ	С	ⱄ
T	T	ᛏ	T	Т	ⱅ
U	V	ᚢ ᚤ	V		
V					
W		ᚹ ᚹ			
X	X	ᚷ (ᛉ)	X	Х (h)	ⱈ
Y	Y	ᛘ ᛣ ᛇ		Ү (u)	ⱊ
Z	Z	ᛇ ᛇ (ik, eo)	I	Ж(z) S(dz) З(dz)	ⱋ ⱌ ⱍ
NG		ᛜ ᛝ (ng)	Ξ +		
value	o sh ts ch sh ŭ	y	ĭ ě yu ya ye	ę (ye)	ǫ yę yǫ kh ps θ (th, f)
Cyrillic	Ш Ч Ц Ч Щ Ъ	Ы Ъ Н	Ь Ѣ Ю Ц Ѥ	Ѧ Ѩ Ѫ	Ѯ Ѡ Ѭ Ѱ Ѳ
Glagolitic	ⱁ ⱏ ⱎ ⱍ ⱋ ⱌ	ⱑ ⱒ ⱓ ⱔ ⱕ ⱖ ⱗ ⱘ	ⱙ ⱚ ⱛ ⱜ ⱝ ⱞ	ⱔ	ⱡ Ɫ Ᵽ ⱤⱣ ⱥ ⱦ ⱧⱣ

Figure 6: Comparison of Latin, runic, Greek, Cyrillic, and Glagolitic alphabets.

more easily and quickly written. In everyday life the cursive script—*i.e.,* the current hand—was developed with continuous modifications for greater speed. There were several varieties of it, such as those of Pompeii and Alburnus Major (a town in ancient Dacia, modern Roşia Montană, Romania). Between the monumental and the cursive scripts there was a whole series of types that had some of the peculiarities of each group. There were lapidary mixed scripts and book semicursive scripts, and there was the early uncial, or rather semiuncial, script of the 3rd century AD, which seems to have developed into the beautiful uncial script.

When the various European countries had shaken off the political authority of Rome and the learned communities had been dissolved and their members scattered, a marked change took place in the development of the Latin literary, or book, hand. Several national hands, styles of the Latin cursive, assumed different features. There thus developed on the European continent and in the British Isles the five basic national hands, each giving rise to several varieties: Italian, Merovingian in France, Visigothic in Spain, Germanic, and Insular or Anglo-Irish hands. At the end of the 8th century the Carolingian (Caroline) hand developed and, after becoming the official script and literary hand of the Frankish Empire, developed as the main book hand of western Europe in the following two centuries. The combination of the majuscules, or capital letters, and minuscules, or small letters, can be attributed mainly to the Carolingian script.

National cursive hands in Europe

In the course of the next centuries, various book hands or chart hands and other cursive scripts developed from the Carolingian style. In the late 12th century and during the next two centuries the letters gradually became angular in shape; this resulted from the pen being held in a position that made a slanting stroke. The new hand, termed black letter or Gothic, was employed mainly in northwestern Europe, including England, until the 16th century. It is still used, though rarely, in Germany, where it is called Fraktur script.

In Italy the black letter was also used, but the Italians preferred a rounder type, called *littera antiqua,* "old letter." During the 15th century the round, neat, humanistic or Renaissance hand was introduced in Florence and was employed for literary productions, while the needs of everyday life were met by an equally beautiful, though not as clearly legible, cursive hand. The two styles developed into two main varieties: (1) the Venetian minuscule, nowadays known as italic, traditionally (though wrongly) considered to be an imitation of Petrarch's handwriting; and (2) the Roman type, preferred in northern Italy, chiefly in Venice, where it was used in the printing presses at the end of the 15th and the beginning of the 16th centuries; from Italy it spread to Holland, England (about 1518), Germany, France, and Spain. The classical Roman character was adopted for the majuscules. This majuscule writing, along with the Roman-type minuscule and the italic, spread all over the world. In England they were adopted from Italy in the 16th century.

The survival of the black letter (Gothic) in Germany is attributed to the fact that it was the current style at the time of the invention of printing in Germany and was thus employed by Gutenberg. In Italy the *littera antiqua* was used by the German printers Konrad Sweynheym and Arnold Pannartz, as well as by Nicolas Jenson, the great Venetian printer who perfected the Roman type.

The modern national alphabets of the western European nations are, strictly speaking, adaptations of the Latin alphabet to Germanic (English, German, Swedish, Dutch, Danish, etc.), Romance (Italian, French, Spanish, Portuguese, etc.), Slavic (Polish, Czech, Slovak, etc.), Baltic (Lithuanian, Latvian), Finno-Ugric (Finnish, Hungarian, etc.), and other languages. The adaptation of a script to a language is not easy, especially when the language contains sounds that do not occur in the speech from which the script has been borrowed. There arises, therefore, the difficulty of representing the new sounds. This difficulty was met quite differently in various alphabets. For instance, the sound *shch* as in English "Ashchurch," which in Russian is represented by one sign (щ), is represented in Czech by two signs (*šč*), in Polish by four (*szcz*), in English likewise by four, though different ones, and in German by as many as seven (*schtsch*). Thus, in these instances, combinations of two or more letters were introduced to represent the new sounds.

Invention of new signs and diacritical marks

In other cases, new signs were invented; *e.g.,* in the early Greek alphabet and in the Anglo-Saxon adoption of the Latin alphabet. In more recent times the most common way of representing sounds that cannot be represented by letters of the borrowed alphabet has been to add diacritical marks, either above or under the letters, to their right or left, or inside. To this group belong the German vowels *ü, ä, ö;* the Portuguese and French cedilla in *ç;* the tilde on Spanish *ñ* and Portuguese *ã* and *õ;* the Italian *à, é, è, ì, ù,* etc.; the great number of marks in the Latin-Slavic alphabets (Polish, Czech, Croatian, etc.)—*ą, ę, č, ć, š, ś, ž, ż, ź,* and so on. The Latin-Turkish alphabet, introduced in 1928, became general throughout Turkey in 1930. It contains 29 letters, of which two vowels (*ö* and *ü*) and three consonants (*ç, ğ,* and *ş*) are distinguished by diacritical marks; in one instance there is a distinction in reverse—the dot from *i* is eliminated (*ı*) to represent a new sound.

(D.D./D.R.O.)

CHINESE WRITING

Chinese writing and Semitic writing constitute the two great writing systems of the world. Just as the Semitic writing system was fundamental to the evolution of modern writing systems in the West, Chinese script was fundamental to the writing systems in the East. Chinese writing, at least until relatively recently, was more widely in use than alphabetic writing systems, and until the 18th century more than half of the world's books had been written in Chinese, including works of speculative thought, historical writings of a kind, and novels, along with writings on government and law.

When China was united in the 3rd century BC, the first emperor, Shih huang-ti, ordered that the writing system be standardized throughout the empire. This common writing system bound the Chinese people together, forming a medium of communication that could be read by groups who spoke very different, often mutually incomprehensible dialects of the language. Chinese writing is the only form of writing that has been in continuous use from the time of the invention of writing down to the present time.

Chinese script is logographic. Characters or graphs represent not units of sound as in phonographic writing systems but rather units of meaning, morphemes. Chinese, like any other language, has thousands of morphemes, and, as one character is used for each morpheme, the writing system has thousands of characters. Two morphemes that sound the same would, in English, have at least some similarity of spelling; in Chinese they are represented by completely different characters. The Chinese words for "parboil" and for "leap" are pronounced identically. Yet there is no similarity in the way they are written.

The fact that the Chinese script is logographic and that its characters or graphs have a pictorial property has led some writers to conclude that it is less abstract than sound-based writing systems. However, recent scholars point out that all writing systems began with pictorial signs that lost their pictorial properties to the requirement of ease of writing; it is easier to draw an arbitrary sign than a realistic picture. And it is now recognized that a logographic script is a relatively optimal solution to the problem of representing the Chinese language.

The Chinese language has clearly distinguished syllables that are easily recognized in speech and hence easily represented by a sign. These syllables correspond to morphemes; each morpheme is one syllable long. In English one morpheme is often expressed by two syllables (*e.g.,* "balloon"), and two morphemes may be contained in one syllable (*e.g.,* "boys"). In Chinese, with a general correspondence between morpheme and syllable, each morpheme is easily represented by a sign for the corresponding syllable. Moreover, one morpheme in Chinese is more or less equivalent to a word. Unlike English, in which morphemes combine to make new words (*e.g., make* + past = *made*), Chinese is an isolating language, in which elements of meaning

are strung together as a series of isolated morphemes. Similarly, the pronunciation of a syllable is relatively uninfluenced by adjacent syllables, which, therefore, remain relatively invariant. It is these invariant units of sound and meaning that are represented by distinctive logographs.

The earliest characters of the Chinese script were "motivated"; that is, they resembled the things they represented (see above *History of writing systems*). With the adoption of the brush as the tool and of ink on paper as the medium for writing, graphs became essentially arbitrary, involving simple lines and shapes. The basic stock of characters are simple graphs, some of which represent the names for objects or parts of objects, such as river, fish, man, and woman, and others of which stand for more abstract terms, such as yield, love, quarrel, prince, and the like. There are approximately 1,000 of these simple characters or graphs.

These basic characters serve two other roles. First, they may double as loan words. Thus, the character representing the word "prince" doubles for "thin-sliced," "law," "beating the breast," "avoid," and others that were difficult to depict directly. The principle for borrowing the character was that the new word be pronounced in the same or a similar way as the word represented by the character. This acrophonic principle played a similar role in the development of hieroglyphic and cuneiform writing. Indeed, it has been suggested that if this principle had been applied consistently, the Chinese would have ended up with a syllabic rather than a logographic system. However, the writing system would then have been extremely ambiguous, with one character representing a dozen or more unrelated words as a consequence of the extreme homophony of the Chinese language. The logographic principle eliminates that ambiguity by providing one character for one meaning.

The second use of the basic characters was in combination with other characters to make up complex characters. Complex characters consist of one graph representing the pronunciation of the character—that is, a graph standing for a set of similar sounding words based on the acrophonic principle combined with a second graph indicating the semantic category of the word. One part represents the sound of the syllable, the other the semantic category of the morpheme; *e.g.*, the character for "foundation" is composed of the character for "winnowing basket," a word that sounds, in Chinese, similar to the word "foundation," together with the character for "earth," a word that is semantically related to the word "foundation."

其 "winnowing basket," combined with 土 "earth," gives 基 "foundation"

The process of combining simple graphs to make complex ones is enormously prolific and had been used to generate thousands of unique characters capable of representing the morphemes of the language. With some 40,000 graphs, the system comes close to the ideal of a fully explicit writing system that represents each distinctive unit of meaning with a distinctive unit of writing. But, of course, such a large number of graphs imposes a major obstacle to learning to read and write. The problem is made more complex by the fact that neither the sound property nor the semantic property of the characters is of much help in the recognition of a character. Because of changes in pronunciation of the language, the complex signs no longer reflect the sound pattern that they originally grew out of. Similarly, the semantic relations represented by the graph are no longer so clear. Consequently, as the relations between the characters and what they represent are largely unknown to readers and writers of the language, the graphs are seen as groups of lines and angles that make up repeated visual units, just as readers of English recognize whole words without analyzing them into their constituent letters. A literate Chinese person knows perhaps 4,000 of the most important characters. (D.R.O.)

Adjuncts to writing

PUNCTUATION

Punctuation is the use of spacing, conventional signs, and certain typographical devices as aids to the understanding and correct reading, both silently and aloud, of handwritten and printed texts. The word is derived from the Latin *punctus,* "point." From the 15th century to the early 18th the subject was known in English as pointing; and the term punctuation, first recorded in the middle of the 16th century, was reserved for the insertion of vowel points (marks placed near consonants to indicate preceding or following vowels) in Hebrew texts. The two words exchanged meanings between 1650 and 1750.

Since the late 16th century the theory and practice of punctuation have varied between two main schools of thought: the elocutionary school, following late medieval practice, treated points or stops as indications of the pauses of various lengths that might be observed by a reader, particularly when he was reading aloud to an audience; the syntactical school, which had won the argument by the end of the 17th century, saw them as something less arbitrary, namely, as guides to the grammatical construction of sentences. Pauses in speech and breaks in syntax tend in any case to coincide; and although English-speaking writers are now agreed that the main purpose of punctuation is to clarify the grammar of a text, they also require it to take account of the speed and rhythm of actual speech.

Syntactical punctuation is, by definition, bad when it obscures rather than clarifies the construction of sentences. Good punctuation, however, may be of many kinds: to take two extreme examples, Henry James would be unintelligible without his numerous commas, but Ernest Hemingway seldom needs any stop but the period. In poetry, in which the elocutionary aspect of punctuation is still important, and to a lesser degree in fiction, especially when the style is close to actual speech, punctuation is much at the author's discretion. In nonfictional writing there is less room for experiment. Stimulating variant models for general use might be the light punctuation of George Bernard Shaw's prefaces to his plays and the heavier punctuation of T.S. Eliot's literary and political essays.

Punctuation in Greek and Latin to 1600. The punctuation now used with English and other western European languages is derived ultimately from the punctuation used with Greek and Latin during the classical period. Much work remains to be done on the history of the subject, but the outlines are clear enough. Greek inscriptions were normally written continuously, with no divisions between words or sentences; but, in a few inscriptions earlier than the 5th century BC, phrases were sometimes separated by a vertical row of two or three points. In the oldest Greek literary texts, written on papyrus during the 4th century BC, a horizontal line called the *paragraphos* was placed under the beginning of a line in which a new topic was introduced. This is the only form of punctuation mentioned by Aristotle. Aristophanes of Byzantium, who became librarian of the Museum at Alexandria about 200 BC, is usually credited with the invention of the critical signs, marks of quantity, accents, breathings, and so on, still employed in Greek texts, and with the beginnings of the Greek system of punctuation. Rhetorical theory divided discourse into sections of different lengths. Aristophanes marked the end of the short section (called a *comma*) by a point after the middle of its last letter, that of the longer section (*colon*) by a point after the bottom of the letter, and that of the longest section (*periodos*) by a point after the top of the letter. Since books were still being written in tall majuscule letters, like those used in inscriptions and like modern capital letters, the three positions were easily distinguishable. Aristophanes' system was seldom actually used, except in a degenerated version involving only two points. In the 8th or 9th century it was supplemented by the Greek form of question mark (;). The modern system of punctuating Greek texts was established by the Italian and French printers of the Renaissance, whose practice was incorporated in the Greek types cut by Claude Garamond for Francis I of France between 1541 and 1550. The colon is not used in Greek, and the semicolon is represented by a high point. Quotation marks and the exclamation mark were added more recently.

Ancient Greek practices

In almost all Roman inscriptions points were used to separate words. In the oldest Latin documents and books, dating from the end of the 1st century BC to the beginning

Roman practice

of the 2nd century AD, words were divided by points, and a change of topic was sometimes indicated by paragraphing: the first letter or two of the new paragraph projected into the margin instead of being indented, as they have been since the 17th century. Roman scholars, including the 4th-century grammarian Donatus and the 6th-century patron of monastic learning Cassiodorus, recommended the three-point system of Aristophanes, which was perfectly workable with the majuscule Latin scripts then in use. In practice, however, Latin books in their period were written continuously—the point between words had been abandoned. The ends of sentences were marked, if at all, only by a gap (which might be followed by an enlarged letter) or by an occasional point. The only books that were well punctuated at that time were copies of the Vulgate Bible, for which its translator, St. Jerome (died 419/420), devised punctuation *per cola et commata* ("by phrases"), a rhetorical system, based on manuscripts of Demosthenes and Cicero, which was especially designed to assist reading aloud. Each phrase began with a letter projecting into the margin and was in fact treated as a minute paragraph, before which the reader was expected to take a new breath.

Early medieval practice

During the 7th and 8th centuries, which saw the transition from majuscule to minuscule handwriting (minuscule scripts were usually smaller than majuscule and had projections above and below the body of the letters, as in modern lowercase letters), scribes to whom the Latin language was no longer as well known as it had been, especially Irish, Anglo-Saxon, and German scribes, to whom it was a foreign language, began to separate words. It was only in the 13th century that monosyllables, especially prepositions, were finally detached from the word following them. The introduction of spaces between words was critical to the development of silent reading, a practice that began only about the 10th century. To mark sentences, a space at the end became the rule; and an enlarged letter, often a majuscule, generally stood at the beginning of sentences and paragraphs alike. The use of points was somewhat confused by St. Isidore of Seville (died 636), whose encyclopaedia recommended an aberrant version of the three-point system; but a point, high or low, was still used within or after sentences. The ends of sentences were often marked by a group of two or three marks, one of which might be a comma and not a simple point.

St. Jerome's concern for the punctuation of sacred texts was shared by Charlemagne, king of the Franks and Holy Roman emperor, and his Anglo-Saxon adviser Alcuin, who directed the palace school at Aachen from 782 to 796. An important element in the educational revival over which they presided was the improvement of spelling and punctuation in biblical and liturgical manuscripts. It is in the earliest specimens of the new Carolingian minuscule script, written at Corbie and Aachen (now in northern France and West Germany, respectively), about 780–800, that the first evidence for a new system of punctuation appears. It soon spread, with the script itself, throughout Europe, reaching its perfection in the 12th century. Single interior stops in the form of points or commas and final groups of stops continued in use; but they were joined by the mark later known as *punctus elevatus* (؟) and by the question mark (*punctus interrogativus*), of much the same shape as the modern one but inclined to the right. The source of these two new marks was apparently the system of musical notation, called neumes, which is known to have been used for Gregorian chant from at least the beginning of the 9th century. *Punctus elevatus* and *punctus interrogativus* indicated not only a pause and a syntactical break but also an appropriate inflection of the voice. By the 12th century another mark, *punctus circumflexus* (7), had been added to *elevatus* to indicate a rising inflection at the end of a subordinate clause, especially when the grammatical sense of the sentence was still not complete. Liturgical manuscripts in particular, between the 10th and the 13th century, made full use of this inflectional system: it is the origin of the "colon" still used to divide verses of the Psalms in breviaries and prayer books. In the later Middle Ages it was especially the Cistercian, Dominican, and Carthusian orders and the members of religious communities such as the Brethren of the Common Life who troubled to preserve a mode of punctuation admirably adapted to the constant reading aloud, in church and refectory, that characterized the religious life. The hyphen, to mark words divided at the ends of lines, appeared late in the 10th century; single at first, it was often doubled in the period between the 14th and 18th centuries.

Late medieval practice

Most late medieval punctuation was haphazard by comparison with 12th-century work—notably in the university textbooks produced at Paris, Bologna, and Oxford in the 13th and 14th centuries. In them, as elsewhere, a form of paragraph mark representing *c* for *capitulum* ("chapter") is freely used at the beginning of sentences. Within the same period the plain point and *punctus elevatus* are joined by the virgule (/) as an alternative form of light stop. Vernacular literature followed the less formal types of Latin literature; and the printers, as usual, followed the scribes. The first printed texts of the Bible and the liturgy are, as a rule, carefully punctuated on the inflectional principle. The profusion of points and virgules in the English books of the printer William Caxton pays remarkably little attention to syntax. Parentheses appear by about 1500. During the 15th century some English legal documents were already being written without punctuation; and British and American lawyers still use extremely light punctuation in the hope of avoiding possible ambiguities.

The beginnings of postmedieval punctuation can be traced to the excellent manuscripts of classical and contemporary Latin texts copied in the new humanistic scripts by Italian scribes of the 15th century. To about 1450 the point and the *punctus elevatus* seem to have been preferred for minor pauses; after that date they are often replaced by the virgule and what is now called the colon (:). The virgule, originally placed high, sank to the base line and developed a curve—turned, in fact, into a modern comma. The Venetian editor and printer Aldus Manutius (Aldo Manuzio; died 1515) made improvements in the humanistic system, and in 1566 his grandson of the same name expounded a similar system in his *Orthographiae ratio* ("System of Orthography"); it included, under different names, the modern comma, semicolon, colon, and full point, or period. Most importantly, the younger Aldo stated plainly for the first time the view that clarification of syntax is the main object of punctuation. By the end of the 17th century the various marks had received their modern names, and the exclamation mark, quotation marks, and the dash had been added to the system.

Punctuation in English since 1600. By the end of the 16th century writers of English were using most of the marks described by the younger Aldo in 1566; but their purpose was elocutionary, not syntactical. When George Puttenham, in his treatise *The Arte of English Poesie* (1589), and Simon Daines, in *Orthoepia Anglicana* (1640), specified a pause of one unit for a comma, of two units for a semicolon, and of three for a colon, they were no doubt trying to bring some sort of order into a basically confused and unsatisfactory situation. The punctuation of Elizabethan drama, of the devotional prose of John Donne or of Richard Hooker, and indeed of Bunyan's *Pilgrim's Progress* (1678) was almost wholly elocutionary; and it lacked the inflectional element that had been the making of 12th-century punctuation. It was Ben Jonson, in his *English Grammar,* a work composed about 1617 and published posthumously in 1640, who first recommended syntactical punctuation in England. An early example is the 1625 edition of Francis Bacon's *Essayes;* and from the Restoration onward syntactical punctuation was in general use. Influential treatises on syntactical punctuation were published by Robert Monteith in 1704 and Joseph Robertson in 1795. Excessive punctuation was common in the 18th century: at its worst it used commas with every subordinate clause and separable phrase. Vestiges of this attitude are found in a handbook published in London as late as 1880. It was the lexicographers Henry Watson Fowler and Francis George Fowler, in *The King's English,* published in 1906, who established the current British practice of light punctuation. Punctuation in the United States has followed much the same path as in Britain, but the rules laid down by American authorities have in general been more rigid than the British rules.

Shift to syntactical punctuation

The system of punctuation now used by writers of English has been complete since the 17th century. Three of its most important components are the space left blank between words; the indentation of the first line of a new paragraph; and the uppercase, or capital, letter written at the beginning of a sentence and at the beginning of a proper name or a title. The marks of punctuation, also known as points or stops, and the chief parts that they play in the system are as follows.

The end of a grammatically complete sentence is marked by a full point, full stop, or period. The period may also be used to mark abbreviations. The colon (:), which was once used like a full point and was followed by an uppercase letter, now serves mainly to indicate the beginning of a list, summary, or quotation. The semicolon (;) ranks halfway between a comma and a full point. It may be substituted for a period between two grammatically complete sentences that are closely connected in sense; in a long or complicated sentence, it may precede a coordinate conjunction (such as "or," "and," or "but"). A comma (,) is the "lightest" of the four basic stops. As the most usual means of indicating the syntactical turning points in a sentence, it is exposed to abuse. It may be used to separate the elements of a series, before a relative clause that does not limit or define its antecedent, in pairs to set off or isolate words or phrases, or in combination with coordinating conjunctions.

Other punctuation marks used in modern English include parentheses, which serve, like a pair of commas, to isolate a word or phrase; question, exclamation, and quotation marks; the hyphen; and the apostrophe.

Punctuation in French, Spanish, German, and Russian. Since the modern punctuation of all the western European languages stems from the practice of the great Italian and French printers of the 15th and 16th centuries, national differences are not considerable. In French, guillemets (<<>>) or dashes are used to mark quotations. In Spanish, since the middle of the 18th century, an inverted mark of interrogation or exclamation has stood at the beginning of sentences as well as the normal mark at the end; and quotations may be marked either as in French or as in English. German punctuation, which is still based on rules propounded in 1781, is more rigorously syntactical than the rest: all relative clauses and all clauses beginning with *dass* ("that") must be preceded by a comma. Quotations are marked either by pairs of commas („") or by reversed guillemets (>><<). Letter spacing, as well as italic type, is used for emphasis. Early Russian punctuation was based on Greek practice, since the Cyrillic alphabet is derived from the Greek; and by the 17th century several quite elaborate systems had evolved in different areas. Since the 18th century Russia has used a form of western European punctuation that has much in common with German practice: notably an even wider obligatory use of commas with subordinate and indeed coordinate clauses, and letter spacing (as well as italics) for emphasis. German quotation marks, French guillemets, and dashes may be used for direct speech.

Characteristics of German punctuation

Punctuation in Oriental and African languages. In Hebrew manuscripts written since the 9th century the main use of points is to indicate the vowel sounds, the alphabet being consonantal only. In biblical texts points and commas are used to mark the middle and end of verses; and in the commentaries points mark the end of sentences. Since the late 18th century, when Jews in Germany began to compose secular texts in Hebrew, the punctuation of such texts has been based on German practice. Early Arabic manuscripts had no punctuation, since the structure of the language ensured that the main and subordinate clauses were readily distinguishable without it. After Arabic began to be printed, European punctuation marks were gradually adopted. The first such mark was the reversed comma; it is now the most common and indicates a suitable point at which to pause and draw breath.

In Sanskrit, prose texts use one vertical stroke to mark the end of the sentence, and verse texts use one vertical stroke for the end of a line, two for the end of a couplet. In Bengali, Gujarati, Hindi, and Marathi, the vertical stroke is used as in Sanskrit, in conjunction with other marks borrowed from English. The diacritical signs and elements of punctuation found in Tamil were introduced early in the 18th century by a Jesuit missionary.

Before the modern period, the grammatical structure of written Chinese was such that no punctuation was required; but in the 19th century editors of texts began to add hollow circles, intended either to mark the ends of phrases or to emphasize particular passages. Since 1912 some of the European punctuation marks have been adopted, notably the marks of interrogation and exclamation and the comma (the hollow circle serves as full point). Direct speech is indicated either by double inverted commas or by an L-shaped mark placed at a corner of the first and last characters. Characters are capitalized by the addition of a straight or wavy line underneath or at the side, according to whether the text is written horizontally or vertically.

In Japan a complicated system of *kaeriten* and *kunten* marks was used from the 8th century onward to clarify the meaning and grammatical construction of texts in Chinese. As a result of contact with Europeans in the 15th and 16th centuries, a hollow point (○) and a reversed virgule (､) were used during the Edo period (1603–1868) as equivalents of the European full point and comma. Since 1868 they have been joined by the solid point (to separate items in a list), by the dash used as in English, and, finally, by the European marks of exclamation and interrogation.

The history of punctuation in Africa is part of the history of the scripts used in different parts of the continent: the Coptic script, based on the Greek alphabet with some additions from demotic writing, for the ancient language of Egypt; a derivative of South Semitic script, known as Ethiopic, for the languages of Ethiopia; Arabic script for speakers of Arabic, Berber, and Swahili; Latin—*i.e.*, European—script for the languages first recorded during and since the 19th century. (T.J.Br.)

SHORTHAND

Shorthand is a method of writing rapidly by substituting characters, abbreviations, or symbols for letters, words, or phrases. Other names for shorthand are stenography (close, little, or narrow writing), tachygraphy (swift writing), and brachygraphy (short writing). Because shorthand can be written rapidly, the shorthand writer is able to record the proceedings of legislative bodies, the testimony of law courts, or dictation in business correspondence. In addition, shorthand has been used through the centuries as a cultural tool: George Bernard Shaw wrote his plays in shorthand; Samuel Pepys recorded his diary in shorthand; Cicero's orations, Martin Luther's sermons, and Shakespeare's plays were all preserved by means of shorthand.

History and development of shorthand. Through the centuries shorthand has been written in systems based on orthography (normal spelling), on phonetics (the sounds of words), and on arbitrary symbols, such as a small circle within a larger circle to represent the phrase, "around the world." Most historians date the beginnings of shorthand with the Greek historian Xenophon, who used an ancient Greek system to write the memoirs of Socrates. It was in the Roman Empire, however, that shorthand first became generally used. Marcus Tullius Tiro, a learned freedman who was a member of Cicero's household, invented the *notae Tironianae* ("Tironian notes"), the first Latin shorthand system. Devised in 63 BC, it lasted over a thousand years. Tiro also compiled a shorthand dictionary. Among the early accomplished shorthand writers were the emperor Titus, Julius Caesar, and a number of bishops. With the beginning of the Middle Ages in Europe, however, shorthand became associated with witchcraft and magic, and disappeared.

"Tironian notes"

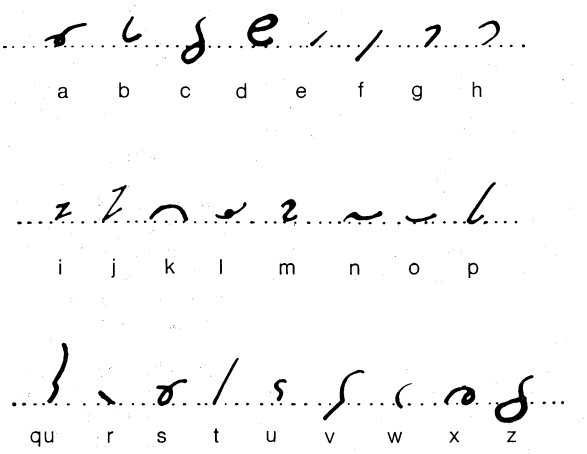
While he was archbishop of Canterbury, Thomas Becket (*c.* 1118–70) encouraged research into Tiro's shorthand. By the 15th century, with the discovery in a Benedictine monastery of a lexicon of Ciceronian notes and a Psalter written in Tironian shorthand, a renewed interest in the practice was aroused. Somewhat influenced by Tiro's system, Timothy Bright designed an English system in 1588 that consisted of straight lines, circles, and half circles.

(Tiro's method was cursive, based on longhand script.) Bright's system was called *Characterie: an Arte of Shorte, Swifte, and Secrete Writing by Character.*

The 17th century produced four important inventors of shorthand systems: John Willis, who is considered to be the father of modern shorthand; Thomas Shelton, whose system was used by Samuel Pepys to write his famous diary; Jeremiah Rich, who popularized the art by publishing not only his system but also the Psalms and the New Testament in his method of shorthand; and William Mason, whose method was used to record sermons and to translate the Bible in the years following the Reformation. Mason's system was later adapted and became the official system of the British Parliament.

Several other systems were invented in the next decades, but most of them were short-lived. One of the most successful was that of the British stenographer Samuel Taylor, who invented a system in 1786 that was based on that of one of his predecessors. Taylor's method was adapted into French, Spanish, Portuguese, Italian, Swedish, German, Dutch, Hungarian, and other languages.

The Industrial Revolution brought a demand for stenographers in business. Because the geometric systems then in use required a high level of education and long training, a need existed for a method that would be easier to learn. The German Franz Xaver Gabelsberger (1789–1849) turned away from geometric methods and developed a simple cursive system. His system, which he called "Speech-sign art," was based on Latin longhand characters and had a neatness and beauty of outline that is unsurpassed. It enjoyed a spontaneous success and spread to Switzerland, Austria, Scandinavia, Finland, and Russia. Its simplicity made it an easy matter to translate it into other languages, and in 1928 it became the Italian national system.



Alphabet of Gabelsberger shorthand, 1834.
From Hans Glatte, *Shorthand Systems of the World*

Pitman's system

Modern symbol systems. Sir Isaac Pitman (1813–97), an educator who advocated spelling reform, was knighted by Queen Victoria for his contributions to shorthand. Pitman had learned Taylor's method of shorthand but saw its weakness and designed his own system to incorporate writing by sound, the same principle he advocated in phonetic longhand spelling. He published his system in 1837, calling it *Stenographic Sound-Hand.* It consisted of 25 single consonants, 24 double consonants, and 16 vowel sounds. Similar, related sounds were represented by similar signs, shading was used to eliminate strokes, the shortest signs were used to represent the shortest sounds, and single strokes were used to represent single consonants. At first, the principle of positioning to express omitted vowels—*i.e.,* writing the word above, on, or below the line of writing—was reserved until later lessons, after the theory had been presented. Later, positioning was introduced with the first lesson.

In 1852 Isaac Pitman's brother, Benn Pitman, brought the system to America, where, with several slight modifications, it became the method most extensively used in the United States and Canada. An investigation in 1889 stated that 97 percent of the shorthand writers in America used the Isaac Pitman system or one of its modifications. Pitman shorthand has been adapted to Afrikaans, Arabic, Armenian, Dutch, French, Gaelic, German, Hebrew, Hindi, Italian, Japanese, Persian, Spanish, and other languages.

The Irish-born John Robert Gregg (1867–1948) taught himself an adaptation of Taylor's shorthand at the age of 10. He then studied Pitman by himself but disliked its angles, shading, and positioning. Later, while in his early teens, he read a history of shorthand by Thomas Anderson, a member of the Shorthand Society of London. Anderson listed the essentials of a good shorthand system, stating that no method then in use possessed them: independent characters for the vowels and consonants, all characters written with the same thickness, all characters written on a single line of writing, and few and consistent abbreviation principles.

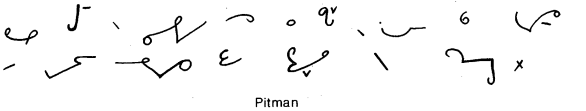
Gregg's light-line phonography

Gregg was 18 when he invented his own system and 21 when he published it in the form of a pamphlet, *Light-Line Phonography* (1888). The Gregg system was predominantly a curve-motion shorthand with circles, hooks, and loops. Based on the ellipse or oval and on the slope of longhand, its motion was curvilinear. Obtuse angles were eliminated by natural blending of lines, vowels were joined, shading was eliminated, and writing was lineal, or in one position.

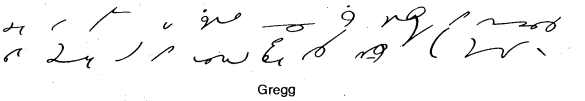
In 1893 Gregg took his system to the United States, and Light-Line Phonography became Gregg Shorthand. The inventor found that, except for the eastern coastal cities, shorthand was virtually unknown. At that time high schools began teaching shorthand, and Gregg traveled through the Midwest, the West, and the South, selling his system and demonstrating his teaching methods with great success. The Gregg system supplanted Pitman's as the predominant system taught in the United States. It also spread to Canada and to the British Isles. Gregg shorthand has been published in English, French, Spanish, Portuguese, Hebrew, Russian, Italian, Tagalog, Japanese, Thai, Chinese, Erse (Scottish Gaelic), Esperanto, Sinhalese, and Polish.

An early German system of importance was the Stolze-Schrey method. Wilhelm Stolze invented his system at about the same time as Gabelsberger and along similar lines. In 1885 Ferdinand Schrey, a Berlin merchant, attempted to simplify the Gabelsberger system. Sometime later the Stolze and Schrey methods were merged and became the leading system in Germany and Switzerland. Stolze-Schrey shorthand was also adapted to other languages, including Danish, Dutch, English, French, Italian, Norwegian, Polish, Russian, and Spanish.

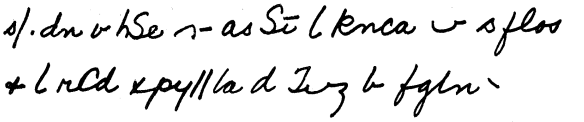
In 1924, after two decades of development, a new system based on the Gabelsberger and Stolze-Schrey methods was completed. As revised in 1936 and 1968, the Deutsch Einheitskurzschrift is the principal system now used in West Germany and in Austria. A modified version was introduced in East Germany in 1970.



Pitman



Gregg



Speedwriting

Modern systems of shorthand recording the same sentence: "Since the dawn of history man has strived to communicate with his fellows and to record experiences that would otherwise be forgotten."

Modern abbreviated longhand systems. The system of Speedwriting shorthand was created around 1924 by Emma Dearborn, an instructor at Columbia University. Her method was designed to be taken down on the typewriter; but in 1942 it was changed to be written by hand with pen or pencil. Speedwriting shorthand uses the letters of the alphabet and the known punctuation marks to represent sounds. For example, the sound of *ch* is written with a capital *C;* the word *each* is thus written *eC*. More than 20,000 words in the Speedwriting dictation can be written with a total of 60 rules and a list of approximately 100 brief forms and standard abbreviations. Speedwriting shorthand is taught in several languages—including English, Spanish, Italian, Portuguese, German, Flemish, and Afrikaans—in many countries.

Forkner Alphabet shorthand was first published in 1952 in the United States. The author, Hamden Forkner, spent 10 years in research before publishing the first edition of the new system, which uses a combination of conventional letters and a few symbols for the hard-to-write letters and sounds. For example, *H* is expressed by a short dash above the line. This same short dash through the letter *C* gives the *ch* sound, through the longhand *S* it gives *sh,* and across the *T* it designates *th*. Abbreviations are used for a number of common words.

Another American method, Hy-Speed Longhand, was first published under that title in 1932. Based on Andrew J. Graham's *Brief Longhand,* published in 1857, its principles include the omission of silent letters and most vowels, the substitution of letters, numbers, or signs, and the combination of certain letters.

Stenoscript ABC Shorthand is a phonetic system using only longhand and common punctuation marks. It originated in London in 1607 and was revised by Manuel Claude Avancena, who published a modern edition in 1950. Stenoscript has 24 brief forms that must be memorized; *e.g., ak* stands for *acknowledge, ac* for *accompany, bz* for *business,* and *gvt* for *government.*

Stenospeed originated in 1950 in the United States; the first publication was called Stenospeed High Speed Longhand, but in 1951 the system was revised under the name of Stenospeed ABC Shorthand. It is used by many schools as a standard text.

Other alphabetic or partially alphabetic systems have also been devised. Among these is Teeline, a system used extensively in Great Britain.

Machine shorthand. A method of recording speech by using machines became commercially feasible around 1906, when the Stenotype machine was invented by Ward Stone Ireland, an American stenographer and court reporter. At present, the Stenograph and Stenotype machines are used in offices to some extent, but they are principally employed for conference and court reporting. Both machines have keyboards of 22 keys. Because the operator uses all fingers and both thumbs, any number of keys can be struck simultaneously. The machines print roman letters on a strip of paper that folds automatically into the back of the machine. The operator controls the keys by touch and is thus able to watch the speaker. The fingers of the left hand control the keys that print consonants occurring before vowels. These keys print on the left side of the tape. The thumbs control the vowels, which are printed in the centre of the tape, and the fingers of the right hand control the consonants that follow the vowels, which are printed on the right side of the tape. There are not separate keys for each letter of the English alphabet; thus, those letters for which there are no keys are represented by combinations of other letters. Abbreviations are used for some of the most frequent words, giving the operator the ability to write two or three words in one stroke. (A.R.R./Ed.)

Stenograph and Stenotype machines

Handwriting and calligraphy

EARLY SEMITIC WRITING

During the 2nd millennium BC, various Semitic peoples at the eastern end of the Mediterranean were experimenting with alphabetic writing. Between 1500 and 1000 BC, alphabetic signs found in scattered sites showed a correspondence of form and provided material for sound translations. Bodies of writing from this period are fragmented: a few signs scratched on sherds or cut in stone. Few of these are celebrated in terms of aesthetic value.

One interesting set of Semitic inscriptions was discovered in 1905 at an ancient mining site on the Sinai Peninsula. A sphinx from that discovery yields the *taw, nun, taw,* or *t, n, t,* meaning "gift." It is evident that the *nun,* or *n,* sign is a rendering of a serpent. Most of the early Semitic alphabetic signs were similarly derived from word signs of more ancient vintage.

The several Semitic peoples in the Middle East area spoke languages that were closely related, and this enabled them to use the same set of alphabetic signs. After some experimentation the alphabet was reduced to 22 signs for consonants. There were no vowel signs. The tribes of Canaan (Hebrews, Phoenicians, and Aramaeans) were important in the development of alphabetic writing, and all seemed to be employing the alphabet by 1000 BC.

The Phoenicians, living along a 20-mile (30-kilometre) strip on the Mediterranean, made the great sea their second home, giving the alphabet to Greeks in the mutual trading area and leaving inscriptions in many sites. One of the finest Phoenician inscriptions exists on a bronze cup from Cyprus called the Baal of Lebanon (Louvre, Paris) dating from *c.* 800 BC. The so-called Moabite Stone (Louvre), *c.* 850 BC, has an inscription that is also a famous example of early Semitic writing.

Old Hebrew. Old Hebrew existed in inscription form in the early centuries of the 1st millennium BC. The pen-written forms of the Old Hebrew alphabet are best preserved in the 13th-century-AD documents of the Samaritan sects.

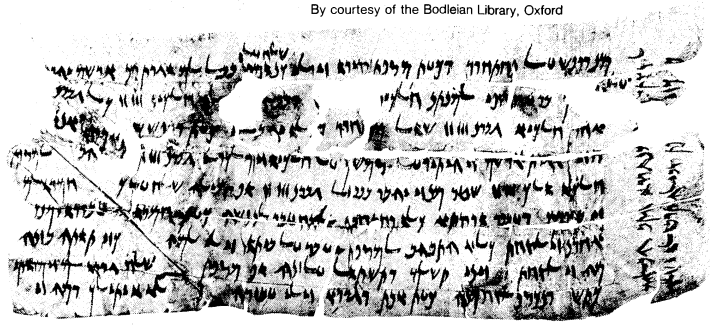
The exile suffered by the Israelites (586–538 BC) dealt a heavy blow to the Hebrew language, since, after their return from exile, Aramaic was the dominant language of the area, and Hebrew existed as a second and scholarly language. Aramaic pen-written documents begin to appear in the 5th century BC and are vigorous interpretations of inscription letters. As seen in the Aramaic document (MS. Pell. Aram. XIV) in the Bodleian Library at Oxford, the penman has cut the pen wide at the tip to produce a pronounced thick and thin structure to the line of letters. The penman's hand was rotated counterclockwise more than 45 degrees relative to vertical, so that vertical strokes were thinner than the horizontal ones. Then, too, there is a tendency to hold these strong horizontals on the top line, with trailing descenders finding a typical length, long or short on the basis of ancient habits. The *lamed* form, which has the same derivation as the Western L, resembles the latter and can be picked out in early Aramaic pen hands by its characteristic long ascender.

Merubba' script

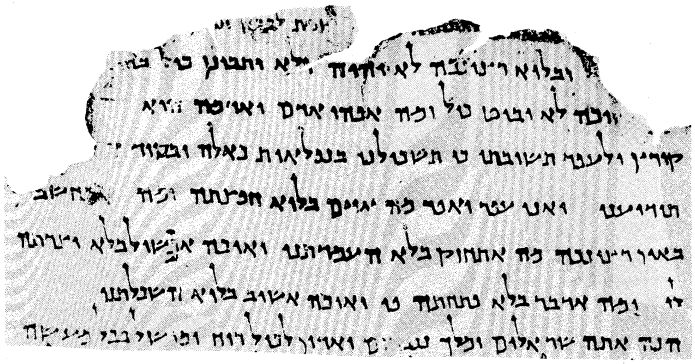
The traditional square Hebrew, or *merubba',* pen hand was developed in the centuries preceding the Christian Era. This early script may be seen in the famed Dead Sea Scrolls discovered in 1947. These scrolls are associated with a group of dissident Jews who founded a religious commune on the northwestern shore of the Dead Sea about 180 BC. The commune had an extensive library. Pens were the instruments of writing, and, as in earlier Aramaic documents, leather provided the surface. Again the *lamed* form is visually prominent.

There are no Hebrew manuscripts from the first 500 years of the Christian Era. Most of the development in the

By courtesy of the Bodleian Library, Oxford



Aramaic pen-written document, 5th century BC (Oxford, Bodleian Library, MS. Pell. Aram. XIV).



Merubba' pen hand. The *Thanksgiving Psalms*, portion of the Dead Sea Scrolls, c. 1st century AD (the Shrine of the Book, Jerusalem; vol. 10).

By courtesy of the Shrine of the Book, The Samuel and Jeane H. Gottesman Centre for Biblical Manuscripts, The Israel Museum, Jerusalem

square Hebrew script occurred between AD 1000 and 1500. The earliest script to emerge from the Dead Sea writing was the Early Sefardic (Spharadic), with examples dating between AD 600 and 1200. The Classic Sefardic hand appears between AD 1100 and 1600. The Ashkenazic style of Hebrew writing exhibits French and German Gothic overtones of the so-called black-letter styles (see below) developed to write western European languages in the late Middle Ages. German black letter, with its double-stroked heads and feet, was difficult for the scribe. Hebrew scripts from this period exhibit some of the same complicated pen stroking and change of pen slant within individual characters. Some decorative qualities of medieval French writing are seen in this Hebrew script.

Sefardic and Ashkenazic styles

Spread of Aramaic to the Middle East and Asia. Aramaic was the mother of many languages in the Middle East and Asia. Generally, the Canaanite-Phoenician influence went west from Palestine, while Aramaic became an international language spreading east, south, and north from the eastern end of the Mediterranean Sea. Never sponsored by great political power, the Aramaic script and language succeeded through inherent efficiency and because the Aramaeans were vigorous traders and extensive travellers in the millennium preceding the birth of Christ.

One of the important languages to derive from Aramaic was Syriac. It was spoken over large areas to the north and east of Palestine, but the literature emerged from a strong national church of Syria centred in the city of Edessa. The development of Syriac scripts occurred from the 4th to the 7th century AD.

Eastern Christendom was riddled with sects and heretical movements. After 431 the Syriac language and script split into eastern and western branches. The western branch was called Serta and developed into two varieties, Jacobite and Melchite. Vigorous in pen graphics, Serta writing shows that, unlike the early Aramaic and Hebrew scripts, characters are fastened to a bottom horizontal. Modern typefaces used to print Syriac, which has survived as a language, have the same characteristic. Eastern Syriac script was called Nestorian after Nestorius, who led a secession movement from the Orthodox Church of Byzantium that flourished in Persia and spread along trade routes deep into Asia. (D.An./Ed.)

Serta and Nestorian writing

By courtesy of the Biblioteca Apostolica Vaticana

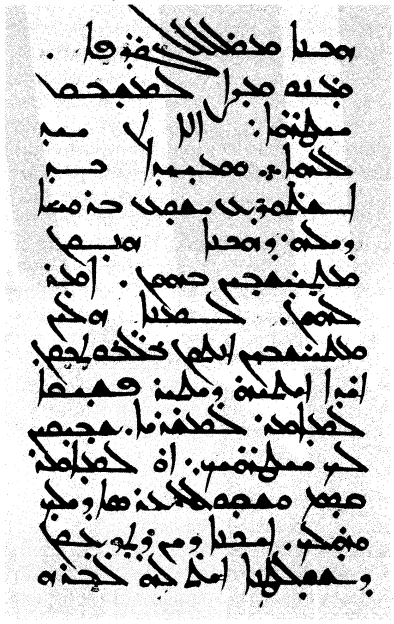
Medieval Hebrew scripts.
(Left) Sefardic script, before AD 1331; in the Biblioteca Apostolica Vaticana, Vatican City (7. Vat. heb. 12. Hagiographa). (Right) Ashkenazic script, AD 1295; in the Biblioteca Apostolica Vaticana (6. Urbinas heb. 1. Biblia).

ARABIC CALLIGRAPHY

In the 7th and 8th centuries AD the Arab armies conquered for Islām territories stretching from the shores of the Atlantic to Sind (now in Pakistan). Besides a religion, they brought to the conquered peoples a language both written and spoken. The Arabic language was a principal factor in uniting peoples who differed widely in race, language, and culture. In the early centuries of Islām, Arabic not only was the official language of administration but also was and has remained the language of religion and learning. The Arabic alphabet has been adapted to the Islāmic peoples' vernaculars just as the Latin alphabet has been in the Christian West.

The Arabic script was evolved probably by the 6th century AD from Nabataean, a dialect of Aramaic current in northern Arabia. The earliest surviving examples of Arabic before Islām are inscriptions on stone.

By courtesy of the Biblioteca Apostolica Vaticana



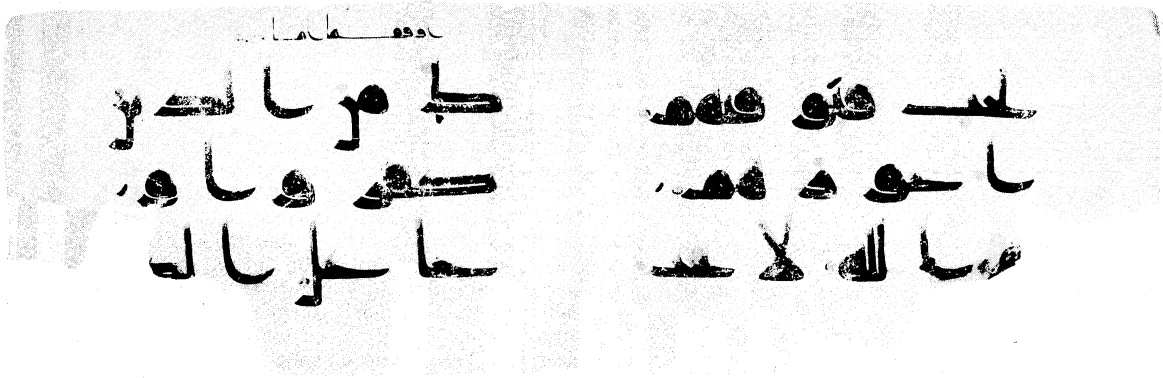
Jacobite script, 1481 (Vatican City, Biblioteca Apostolica Vaticana, 30.b Vat. syr. 18).

Arabic is written from right to left and consists of 17 characters, which, with the addition of dots placed above or below certain of them, provide the 28 letters of the Arabic alphabet. Short vowels are not included in the alphabet, being indicated by signs placed above or below the consonant or long vowel that they follow. Certain characters may be joined to their neighbours, others to the preceding one only, and others to the succeeding one only. When coupled to another, the form of the character undergoes certain changes.

These features, as well as the fact that there are no capital forms of letters, give the Arabic script its particular character. A line of Arabic suggests an urgent progress of the characters from right to left. The nice balance between the vertical shafts above and the open curves below the middle register induces a sense of harmony. The peculiarity that certain letters cannot be joined to their neighbours provides articulation. For writing, the Arabic calligrapher employs a reed pen (*qalam*) with the working point cut on an angle. This feature produces a thick downstroke and a thin upstroke with an infinity of gradation in between. The line traced by a skilled calligrapher is a true marvel of fluidity and sensitive inflection, communicating the very action of the master's hand.

Broadly speaking, there were two distinct scripts in the early centuries of Islām: cursive script and Kūfic script. For everyday purposes a cursive script was employed: typ-

Kūfic script



Kūfic script. Double page opening of a Qur'ān from Syria, 9th century AD. In the collection of R. Pinder-Wilson.
By courtesy of R. Pinder-Wilson

ical examples are to be seen in the Arabic papyri from Egypt. Rapidly executed, the script does not appear to have been subject to formal and rigorous rules, and not all the surviving examples are the work of professional scribes. Kūfic script, however, seems to have been developed for religious and official purposes. The term Kūfic means "the script of Kūfah," an Islāmic city founded in Mesopotamia in AD 638, but the actual connection between the city and the script is not clear. Kūfic is a more or less square and angular script. Professional copyists employed a particular form for reproducing the earliest copies of the Qur'ān that have survived. These are written on parchment and date from the 8th to the 10th century. They are mostly of an oblong as opposed to codex format. The writing is frequently large, especially in the early examples, so that there may be as few as three lines to a single page. The script can hardly be described as stiff and angular; rather, the pace is majestic and measured.

Kūfic went out of general use about the 11th century, although it continued to be used as a decorative element contrasting with those scripts that superseded it. About AD 1000 a new script was established and came to be used for copying the Qur'ān. This is the so-called *naskhī* script, which has remained perhaps the most popular script in the Arab world. It is a cursive script based on certain laws governing the proportions between the letters. The two names associated with its development are Ibn Muqlah and Ibn al-Bawwāb, both of whom lived and worked in Mesopotamia. Of the latter's work a single authentic example survives, a manuscript of the Qur'ān in the Chester Beatty Library, Dublin.

Establishment of *naskhī* script

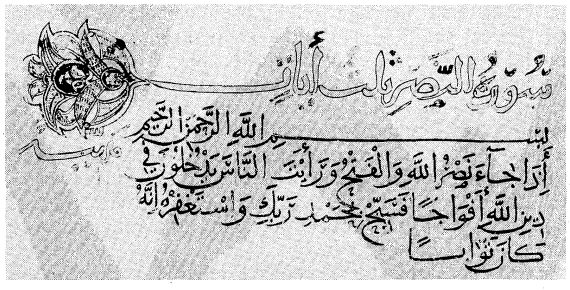
Distinctive scripts were developed in particular regions. In Spain the *maghribī* ("western") script was evolved and became the standard script for Qur'āns in North Africa. Derived ultimately from Kūfic, it is characterized by the exaggerated extension of horizontal elements and of the final open curves below the middle register.

Both Persia and Turkey made important contributions to calligraphy. In these countries the Arabic script was adopted for the vernacular. The Persian scribes invented the *ta'līq* script in the 13th century. The term *ta'līq* means "suspension" and aptly describes the tendency of each word to drop down from its preceding one. At the close of the same century, a famous calligrapher, Mīr 'Alī of Tabriz, evolved *nasta'līq,* which, according to its name, is a combination of *naskhī* and *ta'līq.* Like *ta'līq,* this is a fluid and elegant script, and both were popularly used for copying Persian literary works.

Development of *ta'līq* and *nasta'līq* in Persia

A characteristic script developed in Ottoman Turkey was that used in the chancellery and known as *divani.* This script is highly mannered and rather difficult to read. Peculiar to Turkish calligraphy is the *tuğra* (*ṭughrā*), a kind of royal cipher based on the names and titles of the reigning sultan and worked into a very intricate and beautiful design. A distinctive *tuğra* was created for each sultan and affixed to imperial decrees by a skilled calligrapher, the *neshani.*

By courtesy of the Chester Beatty Library, Dublin



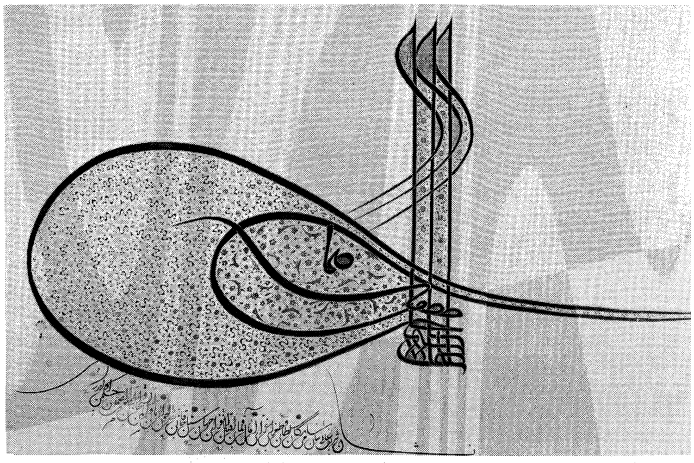
Naskhī script. Baghdad Qur'ān copied by Ibn al-Bawwāb c. 1000 (Dublin, Chester Beatty Library, MS. 1431, fol. 283).

By courtesy of the trustees of the British Museum

Maghribī script. Qur'ān from northwest Africa or Spain, 13th–14th century. In the British Museum, London.

There has always existed in the Islāmic world a keen appreciation of fine handwriting, and, from the 16th century, it became a practice to assemble in albums specimens of penmanship. Many of these assembled in Turkey, Persia, and India are preserved in museums and libraries. Calligraphy, too, has given rise to quite a considerable literature such as manuals for professional scribes employed in chancelleries.

In its broadest sense, calligraphy also includes the Arabic scripts employed in materials other than parchment, papyrus, and paper. In religious buildings, verses from the Qur'ān were inscribed on the walls for the edification of the faithful, whether carved in stone or stucco or executed in faience tiles. Religious invocations, dedications, and benedictory phrases were also introduced into the decoration of portable objects. Generally speaking, there is a



Tuğra of Süleyman the Magnificent, 16th century. In the British Museum, London.
By courtesy of the trustees of the British Museum

close relationship between these and the scripts properly used on the conventional writing materials. It was often the practice for a skilled penman to design monumental inscriptions. (R.H.P.-W./Ed.)

INDIC CALLIGRAPHY

The most important examples of calligraphy to develop from Aramaic writing in its dissemination through southern and Central Asia were the scripts of India, especially of Sanskrit. Indic writing first appeared in the 3rd century BC during the reign of Aśoka (*c.* 265–238 BC). Leader of a great empire, Aśoka turned from military success to embrace the arts and religion. Aśoka's edicts were committed to stone. These inscriptions are stiff and angular in form. Following the Aśoka style of Indic writing, two new calligraphic types appear: Kharoṣṭī and Brāhmī. Kharoṣṭī was used in the northwestern regions of India from the 3rd century BC to the 4th century of the Christian Era, and it was used in Central Asia until the 8th century. It is characterized by a vigorous pen letter, reflecting the influence of Middle East calligraphy.

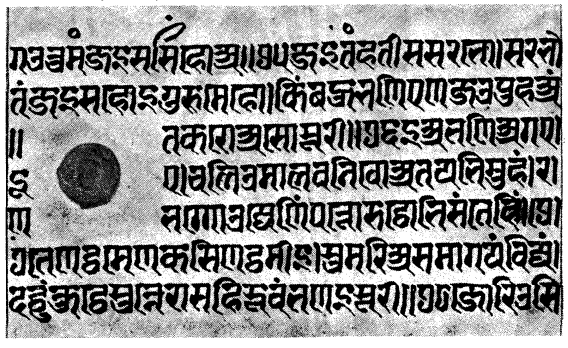
Kharoṣṭī script

Copper was a favoured material for Indic inscriptions. In the north of India, birch bark was used as a writing surface as early as the 2nd century AD. Many Indic manuscripts were written on palm leaves, even after the Indian languages were put on paper in the 13th century. Both sides of the leaves were used for writing. Long rectangular strips were gathered on top of one another, holes were drilled through all the leaves, and the book was held together by string. Books of this manufacture were common to Southeast Asia. The palm leaf was an excellent surface for penwriting, making possible the delicate lettering used in many of the scripts of southern Asia.

Devanāgarī script

Visually, Sanskrit is associated most closely with the alphabetic form named Devanāgarī. In a 15th-century pen-written manuscript in the Freer Gallery at Washington, D.C., it can be observed that the pen's nib is cut wide, giving a considerable difference in thick and thin strokes.

By courtesy of the Smithsonian Institution, Washington, D.C., Freer Gallery of Art



Sanskrit pen-written document, 15th century AD (Washington, D.C., Freer Gallery, MS. 23.3).

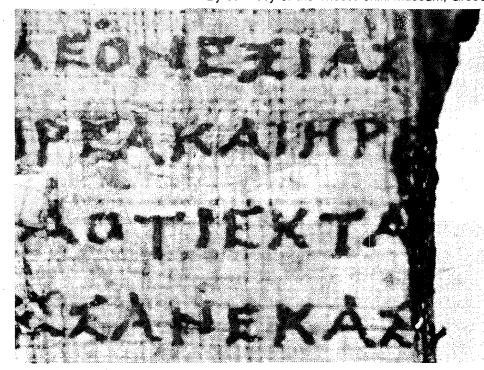
The alphabetic signs hang down from a strong horizontal top line that may become connected. Through the years the strong horizontal and vertical emphasis of inscription writing has been preserved in the Devanāgarī script, and modern typefaces and teaching manuals stress this stiffness of execution. In informal documents this historical script can have more warmth and grace. (D.An.)

GREEK HANDWRITING

Origins to the 8th century AD. The oldest Greek writing, syllabic signs scratched with a stylus on sun-dried clay, is that of the Linear B tablets found in Knossos, Pylos, and Mycenae (1400–1200 BC). Alphabetic writing, in use before the end of the 8th century BC, is first found in a scratched inscription on a jug awarded as a prize in Athens. The consensus is that the Homeric poems were written down not later than this time; certainly from the time of the first known lyric poet of ancient Greece, Archilochus (7th century BC), individuals committed their works to writing. But the vehicles of literary writing have perished. Scratchings on pottery or metal and then texts deliberately cut in bronze or marble or painted on vases are, until *c.* 350 BC, the only immediate evidence for the way the Greeks wrote, and their study is normally treated as the province of epigraphy. A find in 1962 at Dervéni (Dhervénion), in Macedonia, of a carbonized roll of papyrus (Archaeological Museum, Thessaloníki, Greece) offers the oldest example of Greek handwriting and the only one preserved in the Greek peninsula (end of the 4th century BC). From then until the 4th century AD, there are countless texts, especially on papyrus. Found in Egypt and, with a few exceptions, written there, these texts have given a firm foundation for knowledge. From outside Egypt there is a Greek library buried in Herculaneum, AD 79; and papyri and parchments from Owrāmān, Kurdistan, 1st century BC; from Doura-Europus on the Euphrates, 3rd century BC to 3rd century AD; from Nessana, 6th century AD; and from the Dead Sea area (Qumrān, 1st centuries BC and AD; Murabba'at and 'En Gedi, 2nd century AD). A number of original vellum manuscripts have survived from the 4th century AD onward, preserved in libraries such as at the monastery of St. Catherine at Mt. Sinai. These materials of diverse origin suggest that the forms and shape of Greek handwriting were remarkably constant throughout the Greek world, wherever writing was practiced and whatever the material used; within this consistent framework it is occasionally possible to distinguish local variations (as between the contract hands of 1st-century-AD Doura and of Egypt).

Oldest Greek handwriting

By courtesy of the Thessaloniki Museum, Greece



Papyrus from Dervéni, Macedonia, 4th century BC (Archaeological Museum, Thessaloniki, Greece).

The principal vehicles for writing were wax tablets incised with a stylus or a prepared surface of skin, such as leather and vellum, or of papyrus written on with a pen. Other surfaces—*e.g.*, broken pieces of pottery, lead, wood, and even cloth—were also used. To some extent the forms of letters were affected by the resistance of the material to the writing instrument. It is likely that the use as a pen of a hard reed, split at the tip and cut into a nib (which must be constantly sharpened), is an invention of the Greeks. Egyptian scribes used a soft reed, with which ink was brushed on.

Writing materials

Until about AD 300, ink was normally made of a carbon such as lampblack, mixed with gum and water, which even today retains its black lustre. After that time, because of the increasing popularity of vellum, iron inks (*e.g.,* of oak galls), better suited to vellum, tended to replace carbon. The iron inks have faded with age and often have eaten by chemical action into the vellum. Erasures could be made on wax with the blunt end of the stylus and on papyrus by wiping with a sponge; but, on vellum written in iron ink, erasures could be made only by rubbing with pumice or scraping with a knife. Texts from which a previous writing was deliberately erased to provide writing material are termed palimpsests.

Papyrus was normally sold in rolls (*volumina*) made up of 20 or 50 glued sheets: the horizontal fibres of the papyrus are placed on the inside of the roll, on which side (the recto) each gummed sheet overlaps the next when the roll is held horizontally. Leather, similarly, was for long made into rolls (the Dead Sea Scrolls). Shortly after the beginning of the Christian Era, the custom began of folding a single sheet (or several superposed sheets, a quaternion or quire) down the middle and stitching the quires into a binding case to give a book of modern form (codex, originally a set of wax tablets coupled with a thong). Tradition associated this invention with Pergamum.

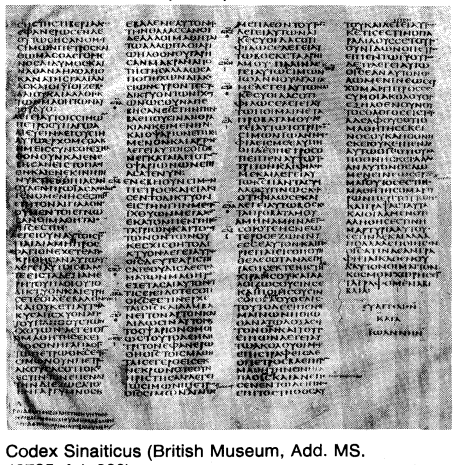
Development of the book

The decisive impetus to the use of this form came from the early Christians, who deliberately chose the commercial vellum notebook (*membranae*) in which to circulate the Christian Gospels in preference to the traditional Jewish roll. Almost without exception the earliest texts of the New Testament are in codex form, even though written on papyrus, which is less able than vellum to bear repeated bending. In the 2nd century AD, pagan works of literature also appeared in this format. By the 4th century it became the predominant form, and codices with handsome margins, of dazzling white vellum, and of sufficient size to contain the whole Bible (*e.g.,* the Codex Sinaiticus) were being produced.

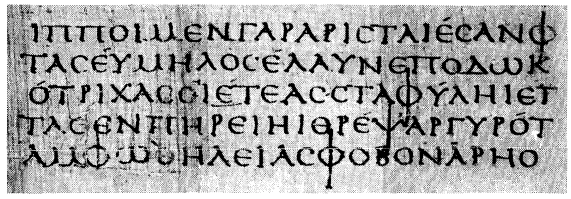
The fundamental distinction in types of handwriting is that between book hands and documentary hands. The former, used especially for the copying of literature, aimed at clarity, regularity, and impersonality and often made an effect of beauty by their deliberate stylization. Usually they were the work of professionals. Outstanding calligraphy is not common among papyrus finds, perhaps because they are mainly provincial work. But the British Museum Bacchylides or the Bodleian Library Homer can stand comparison with any later vellum manuscript from outside Egypt. Book texts are written in separately made capitals (often called uncials, but in Greek paleography, except for the time-hallowed class of biblical uncials, the term is better avoided) in columns of writing, with ample spaces between columns and good margins at head and foot. Punctuation (usually by high dot) is minimal or completely absent; accents are inserted only in difficult poetic texts or as practice by schoolboys; and letters are not grouped into separate words.

Book hands

By courtesy of the trustees of the British Museum



Codex Sinaiticus (British Museum, Add. MS. 43725, fol. 260).



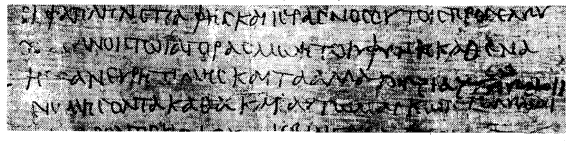
Greek book hand. Bodleian Library Homer, passage from the *Iliad*, copied in the 2nd century AD (Oxford, Bodleian Library, MS. Gr. class A.1 (P)).

By courtesy of the curators of the Bodleian Library, Oxford

Documentary hands show a considerable range: stylized official "chancery" hands, the workaday writing of government clerks or of the street scribes who drew up wills or wrote letters to order, the idiosyncratic or nearly illiterate writing of private individuals. The scribe's aim was to write quickly, lifting his pen very little and consequently often combining several letters in a continuous stroke (a ligature); from the running action of the pen, this writing is often termed cursive. He also made frequent use of abbreviations. When the scribe was skillful in reconciling clarity and speed, such writing may have much character, even beauty; but it often degenerates into a formless, sometimes indecipherable, scrawl.

Documentary hands

By courtesy of the trustees of the British Museum



Greek documentary hand. An authorization for the sale of slaves, late 1st century AD (British Museum, P. Oxy. 94).

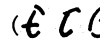
Both types of hands, in spite of the different styles they assume at different periods, show remarkable uniformity and continuity in the shapes of letters. Behind both lies an unvarying basic alphabetic form taught in the schools. The more skillful a book-hand scribe was, the harder it is to date his work. Documents in the ancient world carried a precise date; books never did. To assign dates to the latter, the paleographer takes account of their content, the archaeological context of their discovery, and technical points of book construction (*e.g.,* quires, rulings) or modes of abbreviation. But he finds of great service: (1) a stylistic comparison with those dated documentary hands that show resemblances to book hands, and (2) those cases where a roll was reused—*i.e.,* has a literary text on its recto and a dated document on its verso (in which case there is a *terminus ante quem* for the literary text, often estimated at 50 years before the date of the verso) or has a dated document on the recto and a book hand on the verso (which gives a *terminus post quem* for the literary text, not more than 25 years after the document). The number of illustrated manuscripts of this period is small; their quality is varied; and there is no agreement among specialists about the sources from which illustrations were taken.

Dated texts

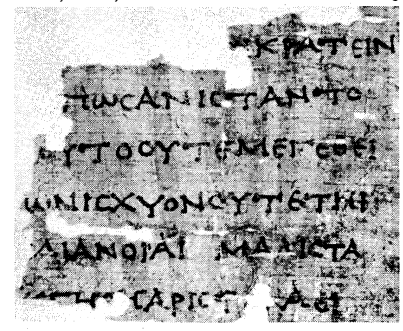
Any historical sketch is bound to be a simplification. At certain epochs several different styles of handwriting existed simultaneously, so that there is no straight line of development. Moreover, owing to the arbitrariness of finds, generalizations are based mainly on provincial work; and, even in that, examples of book hand belonging to the 2nd century BC and the 5th century AD are still relatively rare.

Ptolemaic period. In the roll from Dervéni, Macedonia, dated on archaeological grounds to the 4th century BC, lines and letters are well spaced and the letters carefully made in an epigraphic, or inscription, style, especially the square E, four-barred Σ, and arched Ω; the whole layout gives the effect of an inscription. In the Timotheus roll in Berlin (dated 350–330 BC) or in the curse of Artemisia in Vienna (4th century BC), the writing is cruder, and ω is in transition to what is afterward its invariable written form. Similar features can be seen in the earliest precisely dated document, a marriage contract of 311 BC. It has been argued that a documentary hand of cursive type had not yet been developed and that it was a creation of the

Alexandrian library. Plato, however (*Laws* 810), speaks of Athenian writing whose aim was speed; later on, when a cursive hand had certainly been developed, documentary scribes often used separate capitals.

Characteristic of its period is the contrast of size between the long letters (*e.g.*,) and narrow letters (). And characteristic forms are to be seen in the letters (with its long crossbar, often with initial stroke); (upsilon) with long shallow bowl; or in three or four strokes; in three strokes; (alpha) raised off the line and its last vertical not finished; small round (with internal dot or tiny stroke); and broad epigraphic and . These same features, written with more regularity, appear in the contemporary book hand of a fragment of a Thucydides manuscript (Staats- und Universitätsbibliothek,

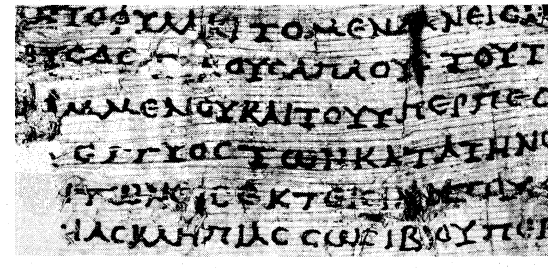
By courtesy of Staats- und Universitatsbibliothek, Hamburg



Thucydides manuscript, 3rd century BC (Hamburg, Staats- und Universitätsbibliothek, P. Hamburg 163).

Hamburg). In documentary cursive hands of this period, letters seem to hang from an upper line: (alpha) often turns into a mere wedge, and (nu) lifts its second vertical above the line. In the 2nd century BC the contrast between long letters and narrow letters disappears, the writing grows rounder, and letters are often linked by ligatures at the top of their last vertical (*e.g.*,). In a loan contract of 99 BC (The John Rylands University Library of Manchester),

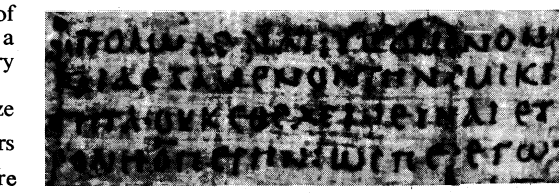
By courtesy of The John Rylands University Library of Manchester



Legal text of a loan contract, 99 BC (The John Rylands University Library of Manchester, Rylands Greek Pap. 586).

in which capitals and cursive are mixed, this irregular roundness is clearly seen. Note the ε with detached crossbar and the exaggerated serifs which have been elevated by some paleographers into a criterion of a special style, though in fact they are always apt to occur.

Roman period. Half a century or so passes after 30 BC before a definitely Roman manner is established. In documentary hands the tendency to roundness continues. Documentary cursive may be influenced in various ways (*e.g.*, by Latin forms such as those of *e* and *d*, or by the



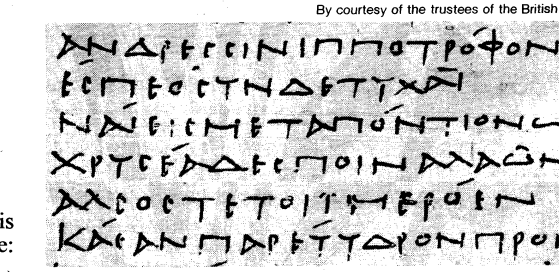
Phaedo, by Plato, copied in AD 100 (London, Egypt Exploration Society, P. Oxy. 1809).

By courtesy of the Egypt Exploration Society, London

exaggeration of verticals practiced by chancery scribes), may lean over in either direction, or may be reduced to tiny proportions. In the 2nd century the cursive hand tends to be round and sprawling, in the 3rd century to become more angular, and in the 4th century to become characterless and to combine letters into ligatures that distort the forms of the letters concerned. The book hand of a manuscript of Plato's *Phaedo* (*c.* AD 100; Egypt Exploration Society, London) shares its informality but regularizes the letter forms. Written on a larger scale and with more formality, this round hand can be very beautiful. In an example found at Hawara (2nd century AD), almost every letter (even ρ, τ, ι) would go into an identical square; only φ and ψ cross it above and below, μ, ω, and π horizontally.

If this writing is made to lean to the right and to revive the 3rd-century-BC distinction between narrow and broad letters, it takes on the aspect of the "severe" style of the Bacchylides roll in the British Museum (2nd century AD). If, however, the scribe makes his verticals or obliques

By courtesy of the trustees of the British Museum



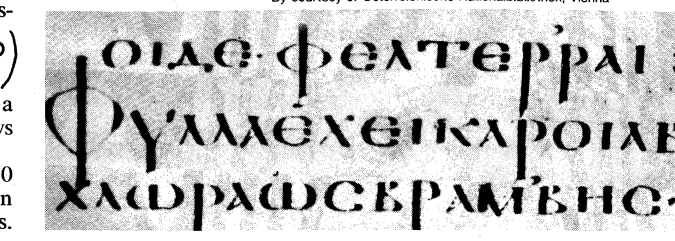
The "severe" style. Bacchylides roll, 2nd century AD (British Museum, P. 733).

thicker and his horizontals thinner, the hand is called biblical uncial, so named because this type is used in the three great early vellum codices of the Bible: Codex Vaticanus and Codex Sinaiticus of the 4th century and Codex Alexandrinus of the 5th century. It is now certain that this style goes back to the 2nd century AD. In the Dioscorides herbal in Vienna, written in AD 512, the writing is rigid and lumbering; the thick strokes are overdone; and blobs of ink terminate the horizontals of, for example, δ, ε, σ, τ. Such heavy decoration is also a feature of the Coptic style, of which there are examples as early as the 2nd century AD. This hand may be thought of as constituting a special case of biblical uncial.

Biblical uncial

Byzantine period. For the paleographer the significant division is not the founding of Constantinople in 330 but the 5th century, from which a few firmly dated texts survive. At its close a large, exuberant, florid cursive is found fully established for documents; in the 7th and 8th centuries it slopes to the right, becomes congested, and adopts some forms that anticipate the minuscule hand. A favourite informal type of the 6th century is shown in

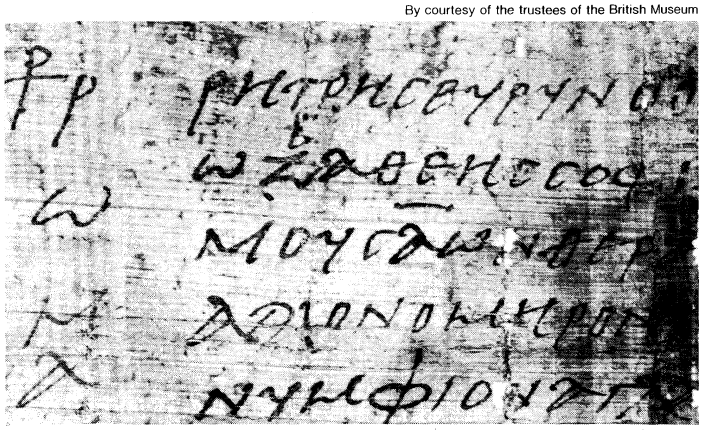
By courtesy of Osterreichische Nationalbibliothek, Vienna



Biblical uncial. Dioscorides herbal, AD 512 (Vienna, Österreichische Nationalbibliothek, Med. Gr. 1).

an acrostic poem by Dioscorus of Aphrodito; it bears a clear relationship to the Menander *Dyskolos* hand, which was probably written in the later 3rd century AD. Similar pairs could be found to illustrate the continuity in transformation of the biblical uncial and Coptic styles. The latest Greek papyrus from Egypt is not later than the 8th century. There is a considerable lapse of time before the history of Greek writing resumes at Byzantium. (E.G.T.)

By courtesy of the trustees of the British Museum



Informal Byzantine book hand of the 6th century. Acrostic poem by Dioscorus of Aphrodito, 6th century AD (British Museum, P. 1552).

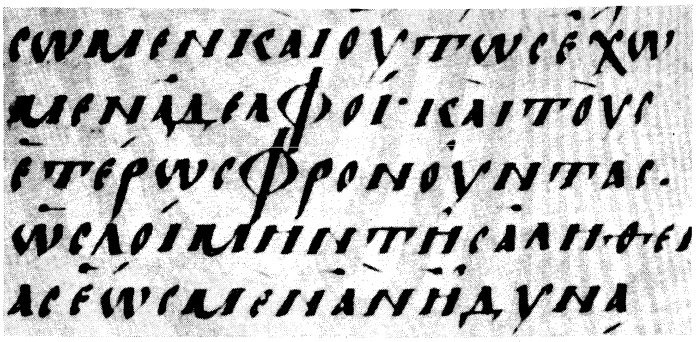
The 8th to 16th century. To judge when and where a Greek manuscript was written is as difficult in this as in the earlier period, but for different reasons. The material for study is admittedly more extensive; manuscripts produced in the Middle Ages and Renaissance have been preserved in very large numbers (more than 50,000 whole volumes survive, of which probably 4,000 or 5,000 are explicitly dated), and they include work from most parts of the Byzantine Empire as well as from Italy. The difficulty of the paleographer lies in the essential homogeneity of the material, which is largely the result of the conditions in which the manuscripts were produced.

The fully developed Byzantine Empire of the 8th to the mid-15th centuries was extraordinarily uniform in its culture. Its contraction in space after the Arab conquests of the 7th century, which cut off the more distant and ethnically differentiated provinces of Syria, Palestine, and Egypt, made it a relatively compact geographical entity. The continuity and comparative stability of a single empire not divided into distinct national states such as evolved in the West resulted in a strength and unity of tradition of which the Byzantines were always conscious and that shows in their habits of writing no less than in their literature and art. Distinct local styles and sharp breaks in ways of writing in different periods cannot, therefore, be looked for; characteristics that may be specially typical of one period come in gradually and disappear equally slowly. A more potent factor than date or place in producing divergencies in the style of writing is the purpose for which a manuscript was designed and what type of scribe wrote it.

Late uncial, 9th to 12th century. There is a gap in the evidence covering the 7th and 8th centuries, because of the Arab conquest of Egypt, the perpetual wars on all fronts in the 7th century, and the iconoclastic struggle among Eastern Christians during the 8th and early 9th centuries, so that no literary texts (and very few others) have survived that can actually be dated to this period.

During this time the evolution of writing in capitals (not very aptly named uncial) probably continued toward a greater formality and artificiality. But this natural tendency was hastened by the introduction and spread of minuscule as the normal way of writing, after which the purpose of uncial changed completely. From an everyday hand in which all books were naturally written, it became a ceremonial hand used only for special copies and therefore grew increasingly stylized and artificial. In the 9th century a still elegant style was used for both patristic and classical works in splendid volumes destined for the imperial library or for presentation copies, such as the copy

Change of purpose of uncial



Late uncial. Copy of Gregory of Nazianzus, AD 879–883 (Paris, Bibliothèque Nationale, Grec. 510, fol. 61v).

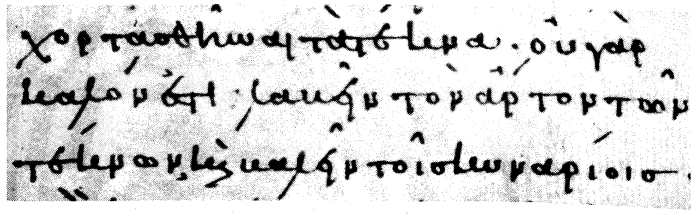
By courtesy of the Bibliotheque Nationale, Paris

of Gregory of Nazianzus (Bibliothèque Nationale, Paris) made for the emperor Basil I between 879 and 883. By the 11th and 12th centuries, capitals were used only for liturgical books, mainly lectionaries, which had to be read in dimly lit churches; but the increasing tortuousness of the style must in the end have reduced its usefulness, and by about 1200 uncial was dead.

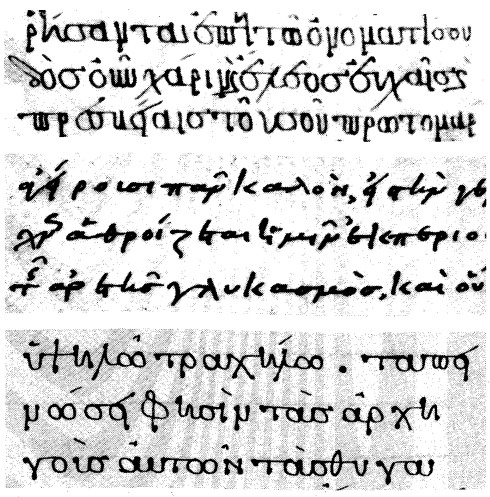
Earliest minuscule, 8th to 10th century. By far the most important development that took place during the 7th–8th-century gap was the introduction of minuscule. There is no incontrovertible evidence of how this came about, or where. What scraps of evidence there are (a few documents from the gap, a few sentences in lives of the abbots of Stoudion of that time, and the first dated manuscript written in true minuscule) point to its development from a certain type of documentary hand used in the 8th century and to the likelihood that the monastery of the Stoudion in Constantinople had a leading part in its early development. Though its origins are obscure, the reasons that led to its introduction and rapid spread are obvious: the state of poverty resulting from wars and persecutions coincided with a shortage of papyrus after the Arab conquest of Egypt in the middle of the 7th century, and these factors combined to induce a search for a more economical use of the relatively expensive vellum; the polemics of the iconoclastic controversy demanded a speedy, informal style of writing; and, finally, when peace was restored in the middle of the 9th century, the revival of learning, with the reorganization of the university, brought the need for multiplying plain workmanlike texts for scholarly purposes.

Reasons for the rapid spread of minuscule

The earliest dated example of true minuscule (and it is probably one of the oldest extant examples altogether) is a copy of the Gospels written in 835 (Leningrad Library), probably in the monastery of the Stoudion. Here are found all the characteristics of the earliest minuscule, which is called pure minuscule because there is as yet no admixture of uncial forms, except occasionally at line ends. The letters are even and of a uniform size; letters are joined or not joined to each other according to strict rules, sometimes by ligatures in which part of each letter is merged in the other, but not to the extent of distorting the shape of either letter. There is no division between words, for the divisions are only those that arise from the rules for joining or otherwise of individual letters, and at this stage any letter that can be joined to the next one nearly always is joined to it. Breathings are square, either ⊢ ⊣ or ⌊ ⌋, and accents are small and neat; abbreviations are very few, usually confined to the



Earliest dated true minuscule. Copy of the Gospels supposedly done at the monastery of Stoudion, Constantinople, AD 835 (Leningrad, Bibl. Publ. 219, fol. 124).



Greek hands, 9th–10th century.
(Top) Stoudion minuscule, AD 890 (Paris, Bibliothèque Nationale, MS. grec. 1470, fol. 168). (Centre) Commentary on Gregory of Nazianzus, AD 986 (Bibliothèque Nationale, MS. suppl. grec. 469 a, fol. 7). (Bottom) Commentary on Isaiah by Basil, AD 953 (Oxford, Bodleian Library, MS. Auct. E.2.12, fol. 80).

By courtesy of (top, centre) the Bibliotheque Nationale, Paris, (bottom) the curators of the Bodleian Library, Oxford

established contractions for *nomina sacra* (the names and descriptions of the Trinity and certain derivatives), omitted *ν* at line ends, a few of the conventional signs for omitted case endings, and ⸌S⸍ sometimes for *καὶ*. The writing stands on the ruled lines or is guided by them.

First changes in minuscule

Absolutely pure minuscule did not last long. Gradually, uncial forms of those letters that had specifically minuscule forms began to be used alongside the minuscule forms: *λ* was the first to appear, followed by *ξ* and then *κ*, all by the end of the 9th century. Then from about 900 onward, *γ*, *ζ*, *ν*, *π*, and *σ* were used regularly, while *α*, *δ*, *ε*, and *η* were used sometimes. Not before about 950 were *β*, *μ*, *υ*, *ψ*, and *ω* used, and still comparatively rarely. But by the end of the 10th century, the interchangeability of all uncial and minuscule forms was complete, though all the alternative forms are not necessarily found in any one manuscript. Perhaps the earliest dated manuscript with any uncial form in it is of 892/893 (Mt. Sinai, St. Catherine, MS. 375 + Leningrad, Bibl. Publ. MS. 343, Chrysostom), but pure minuscule continued to be used, in probably the majority of manuscripts, up to 900 and thereafter mainly in provincial manuscripts until the last dated example in 969 (Metéora, Metamorph. MS. 565, John Climacus). Besides the intrusion of uncial letters, some other characteristics of the earliest minuscule were modified during the 10th century. Rounded breathings, ᾽, are first found in manuscripts of the last half of the century, interspersed with square ones. From about 925 the practice of making the writing hang from the ruled lines began to prevail. Although in most manuscripts abbreviations were confined to a few forms used at line ends only, a few copies dated in the last part of the century used nearly all the conventional signs.

In spite of these developments—the gradual disappearance of pure minuscule and the other changes that accompanied it—the same general styles of writing persisted until about the end of the 10th century. Broadly considered, three styles can be distinguished during this period. There is a rather primitive-looking, angular, cramped style that may perhaps be associated with the Stoudion monastery, in which a certain number of mainly patristic texts were written *c.* 880–*c.* 980 (*e.g.,* Paris, Bibliothèque Nationale, MS. grec 1470, a Menologion of 890). Second, there is a plain, neat, workmanlike style (seen in a commentary on Gregory of Nazianzus copied in 986 that is preserved in the Bibliothèque Nationale at Paris), which continued in use at least until the end of the 10th century. In it were written several of the important manuscripts that are now the oldest texts of some ancient Greek authors (for example, Aeschylus, Sophocles, Aristophanes) but are unfortunately not explicitly dated. Third, a consciously elegant, even mannered, style was used in books produced for the imperial library or for wealthy dignitaries, but it is not found before the early years of the 10th century (as seen in a copy of Basil on Isaiah in the Bodleian Library that is dated 953). All of these styles, which have numerous variations and are by no means always distinct from one another, are found at least until the end of the 10th century. Their one common characteristic is a crispness and individuality that clearly distinguishes them from writing of the next period.

Formal minuscule, 10th to 14th century. From about the middle of the 10th century, a smoother, almost mechanical appearance can be noticed in an increasing number of manuscripts; the hands seem more stereotyped, less individual. They are not immediately distinguishable from the plainer styles of the earlier part of the century, and their evolution during the next four centuries was very gradual. A few distinct types can be singled out from time to time. A bold, round, heavy liturgical style, fully established in the 11th century, was one of the most enduring types (*e.g.,* British Museum, Add. MS. 19,352, a psalter of 1066); it became more and more stereotyped and mechanical until, in the 15th century, a branch of it was transplanted to Italy.

The style most widely used for biblical and patristic texts from the end of the 10th century, probably mainly in monastic houses in Constantinople, was one with plain, neat, rounded letters; this style became known as *Perlschrift* from its likeness to small, round beads strung together. A very plain, businesslike, rather staccato style was used in manuscripts with musical notation, most commonly in the 12th and 13th centuries (*e.g.,* Leningrad, Bibl. Publ. MS. 789, a sticheron of 1106).

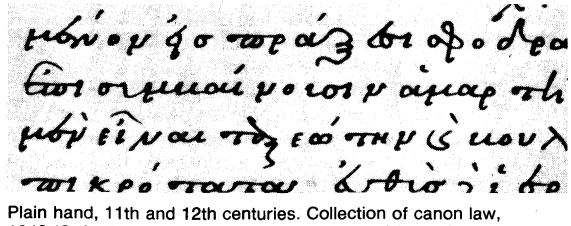
Manuscripts written outside Constantinople are recognizable, if at all, usually by a rougher, provincial appearance. Only two styles can be assigned with any certainty to a specific provincial centre. One, a small unpretentious hand used by St. Nilus, the founder of numerous monasteries in southern Italy at the end of the 9th century, was used for a time by others in that area (*e.g.,* a copy of the works of Dorotheus made by Nilus in 965). In the heyday of the reorganized Greek monasteries there in the 12th century, another elegant, rather mannered style, which almost certainly had its origin in Constantinople, is nevertheless found in manuscripts known to have been written in southern Italy and Sicily (*e.g.,* Paris, Bibliothèque Nationale, MS. grec 83; Gospels copied in Sicily in 1167).

These particular styles, however, are not really as typical of the period as the less distinctive plain hands in which the majority of the manuscripts are written, at least in the 11th and 12th centuries (*e.g.,* a collection of canon law copied in 1042 in the Bodleian Library).

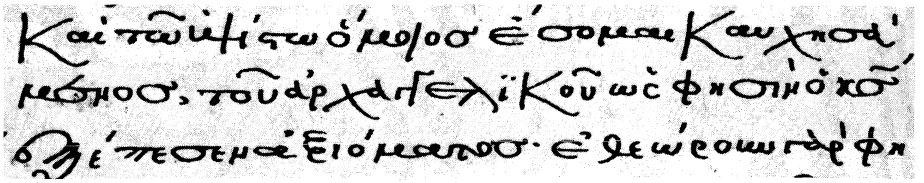
The comparatively uniform type of writing of which all these were minor variations was remarkably enduring and widely dispersed, but, from the 11th century onward, certain changes may be observed that help to date manuscripts written in all types of formal minuscule. One change in its general appearance may be noticed as the 12th century advances: an increasing lightness of touch and a lessening of the closely knit, rather thick appearance that is characteristic of the 11th century. But the most noticeable change in this period is the breakdown in the evenness and regularity of the writing, which is partly attributable to the influence of documentary and the later

Changes dating from the 11th century

By courtesy of the Bodleian Library, Oxford



Plain hand, 11th and 12th centuries. Collection of canon law, 1042 (Oxford, Bodleian Library, MS. Barocci 196, fol. 253).



Formal hand, 12th and 13th centuries. A synaxarion, probably 1329 (Oxford, Bodleian Library, MS. Auct. E.5.10, fol. 73v).

By courtesy of the Bodleian Library, Oxford

personal hands. It is not, however, entirely so attributable, for a tendency to enlarge some letters out of proportion to the size of the rest is seen in a small way in some of the more personal hands of the earliest period. But it is rare in formally written manuscripts, only gradually becoming more general until, in the 12th and 13th centuries, it is the most noticeable feature of even the most formal hands (*e.g.*, a synaxarion, or a short narrative of a saint's life, copied probably in 1329). In the 14th century and later there was a return to less flamboyant ways with the tendency to imitate earlier models more closely, but the habit of enlarging some and diminishing the size of other letters never died out.

In the actual forms of letters used in these formal styles, there was practically no change; very occasionally, from the end of the 10th century onward, one of the "modern" forms of letter normally confined to personal hands found its way into a formal manuscript. Much the same is true of ligatures. The tendency from the 11th century onward is to use ligatures and to join letters less automatically than in earlier times. The permissive rules and most of the forms remain unchanged, for, already in the 10th century, most of the distorting forms (notably those in which the ε is represented only by a C-shaped stroke; *e.g.*, for σε) were well established, and in formal manuscripts these, with the earlier forms, continued in use until they were illogically taken over by the first printers of Greek. Time did, however, gradually increase the tendency to join letters by insetting them in or superimposing them upon each other. Abbreviations were even more conservatively used, only the oldest conventional forms being admitted, and often only a very few and those only at line ends.

The rule that the writing should hang from the ruled lines, already applied in most manuscripts by the mid-10th century, became invariable by the middle of the 11th. Square breathings (used indiscriminately among the round ones) were gradually eliminated, though they did not completely disappear from formal manuscripts until the middle of the 12th century. The practice of joining accents with breathings and also with the letters to which they belonged spread from personal hands to formal writing in the 13th century, but it was far more often avoided altogether.

Apart from the actual writing, one development is common to all manuscripts written in this period: the use of paper instead of vellum, which occurred first perhaps in the late 11th century and was common by the 13th century whenever economy was a major consideration.

These are the main criteria by which a formally written manuscript can be assigned to an earlier or a later part of this period. But the problem of distinguishing different styles and their dates, and their places of origin, remains most difficult for these Greek manuscripts.

Personal hands, 12th to 14th century. From the beginning of minuscule, there were obviously educated individuals who occasionally copied texts for their own use in a formal hand that nevertheless had a distinctive personal flavour; indeed, professional scribes occasionally used a less formal style than usual. Several dated examples of this type of hand survive from the 10th, 11th, and early 12th centuries, but they are rarities. Toward the end of the 12th century, however, the prosperity and comparative stability of the Comnenan age (named from the dynasty of Byzantine emperors bearing the name Comnenus), with its brilliant literary and artistic achievements, gave way to increasing internal chaos and the hostile encirclement of Byzantium that was a prelude to the Fourth Crusade and the sack of Constantinople by the Western powers in 1204. Scholars perhaps already felt the pinch of poverty, which naturally grew greater during the exile of the Byzantium court (1204–61) and culminated in the economic crises of the 14th century.

Copying by scholars

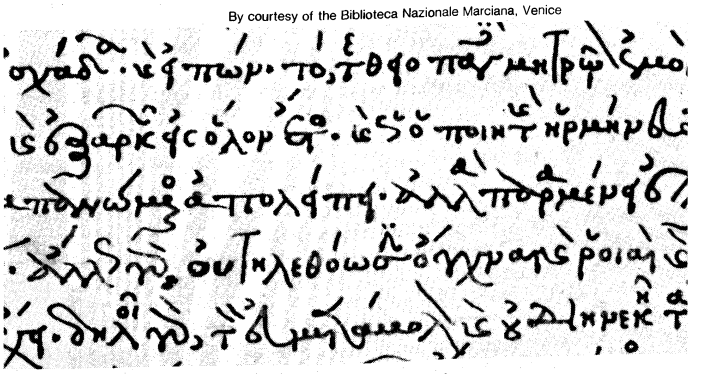
Certainly, a change in writing habits began slowly to take place. Instead of commissioning professional scribes to copy manuscripts, some scholars began to make copies for themselves, and, in place of the smooth, mechanical styles of the professionals, they used the sort of writing that they presumably already used for personal notes. This was an adaptation (for greater clarity) of the type of writing that had been standardized in official documents from the beginning of the Byzantine period. Its chief characteristic was the greatly exaggerated size of certain letters or parts of letters, particularly letters with rounded bows such as β, ε, ζ, θ, κ, ξ, ο, υ, φ, and ω, and the excessive size of these letters is made to look even more unbalanced by some exceptionally small forms of, for example, η, ι, ν, or ρ. This essentially unbalanced, "wild" look was transplanted to literary manuscripts written by scholars for their own use.

Along with this exaggerated contrast in size between letters, they took from the documentary hands several new forms of letters that had gradually evolved from the originally common forms of both hands. In the 12th century the new scholarly hands began to use with separate small bows; , with a broken back; , which had lost its high first stroke; and , which had dropped its first long downstroke; and, by the end of the 13th century, , with a short embryonic tail. The old forms of ligature were kept basically the same but, in some cases, were reduced to a barely recognizable minimum (*e.g.*, or for ει) and, in others, distorted by the general flourishing tendency of the script (*e.g.*, for επ). Abbreviations were naturally used with great frequency in all positions; the ancient conventional signs for suppressed syllables, which had acquired rounded and more flourished shapes, were used alongside a certain amount of "arbitrary" abbreviation in which a large part of a word was omitted and replaced simply by a general sign that some abbreviation had taken place.

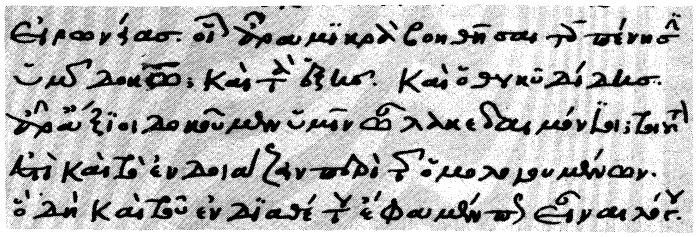
Use of abbreviations

Accents and breathings joined with each other, with letters, and with abbreviation marks are found earlier and

By courtesy of the Biblioteca Nazionale Marciana, Venice



Early medieval scholarly hand. Commentary on the *Odyssey*, written by Eustathius c. 1150–70 (Venice, Biblioteca Nazionale Marciana, MS. Marc. Gr. Z. 460 [coll. 330], fol. 79).



Late medieval scholarly hand. Grammatical works copied by Triclinius, 1308 (Oxford, New College, MS. 258, fol. 205).
By courtesy of New College, Oxford

more frequently in scholarly than in formal manuscripts. The only exception to the rule of round breathings in this type of manuscript is in cases of deliberate archaism such as practiced by Demetrius Triclinius (died *c.* 1340).

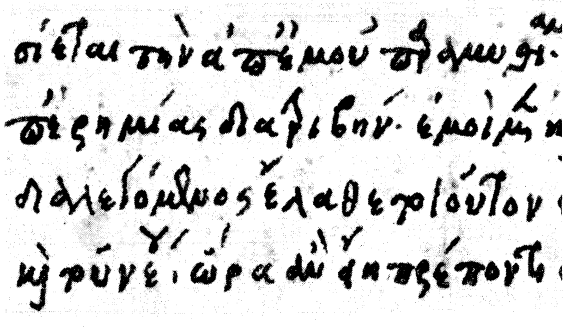
One of the earliest datable examples of these scholarly productions is the copy of his commentary on the *Odyssey* (Biblioteca Nazionale Marciana, Venice) written *c.* 1150–70 by Eustathius, the scholar-archbishop of Thessalonica. In the 13th century the exaggeration of specially round features reached its height (*e.g.,* in a copy of Euthymius Zigabenus on the Psalms, of 1279 [Oxford, Bodleian Library, MS. Roe 7]), while, in the 14th century, the tendency, as in the formal styles of writing, was toward less ebullience and exaggeration, and the writing of scholars such as Triclinius is compact and sober (*e.g.,* his copy of Aphthonius and other grammatical works, of 1308). For these hands the problem is not to discover centres of writing or styles for different uses but to identify the hands of individual scholars.

The Italian Renaissance. By the end of the 14th century, Italian scholars were learning Greek, and they were bringing back Greek manuscripts from Constantinople. At this time Greek scholars had also begun to teach in Italy. The Greek that the earliest Italians learned to write was a clear, simple style taught originally by Manuel Chrysoloras (died 1415). But, although they copied a number of manuscripts for themselves in this hand, the style had no influence beyond their small circle. Before long, Greek scribes began to go to Italy, and both scholars and scribes arrived in increasing numbers as the Turks pressed in around the Byzantine capital until it finally fell in 1453. They brought with them, naturally, the two styles of writing that had persisted throughout the history of the empire. On the one hand, professional scribes such as Joannes Rhosus (died *c.* 1500), the majority of them from Crete, copied an astonishing number of manuscripts in the formal, now glib and stereotyped "liturgical" style of writing (*e.g.,* British Museum, Harl. 5658, an *Odyssey* copied by Rhosus in Rome in 1479). On the other hand, scholars such as Janus Lascaris continued to write in a mannered personal style (*e.g.,* a letter of Demetrius Chalcondyles of 1488 in the Biblioteca Apostolica Vaticana, Vatican City).

Copies in "liturgical" style

It was on the scholarly hands that Aldus Manutius and other early Italian printers of Greek based their types. But perhaps the most enduring was that of a group of Cretan scribes who were employed by Francis I in his library at Fontainebleau. The writing of one in particular, Angelus Vergecius (*e.g.,* his copy of Manuel Philes made in 1564 in the Bodleian Library), was used as a model for the French Royal Greek type, which has influenced the form of Greek printing down to the present day. (R.Ba./Ed.)

By courtesy of the Biblioteca Apostolica Vaticana



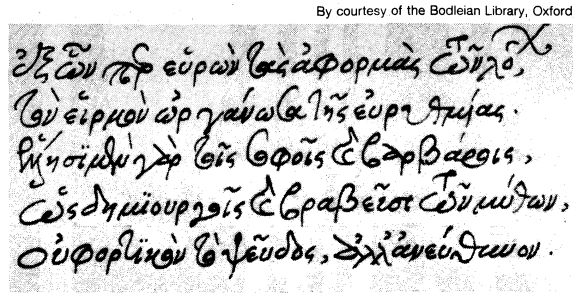
Renaissance personal hand. Letter by Demetrius Chalcondyles (autograph), 1488 (Vatican City, Biblioteca Apostolica Vaticana, Lat. 5641, fol. 2).

LATIN-ALPHABET HANDWRITING

Ancient Roman styles. *Rustic capitals.* The Latin and vernacular handwriting of western Europe descends in an unbroken line to the present day from the point at which it is first observed, in the 1st century AD. The script used throughout the Roman Empire at that time for books and occasionally for formal documents is known as rustic capitals. The pen was cut with a broad end and held so that its thickest strokes fell at an oblique angle to the line of writing, and it was lifted several times in the formation of a single letter. The rustic capital alphabet is "majuscule," in that all the letters are contained between a single pair of horizontal lines. The use of this elaborate script, whose letter forms were the natural outcome of using a broad pen held obliquely, was extended to certain sorts of inscription on stone and other materials, and it is called rustic only by comparison with the magnificent square capitals characteristic of Roman imperial inscriptions, whose forms were governed by the use of the chisel. Square capitals were seldom used in manuscripts except for titles. Rustic capitals continued in use for literary manuscripts until the 6th century, especially for texts of Virgil, but thereafter they appeared only in titles, down to the 12th century.

Roman book hand

By courtesy of the Bodleian Library, Oxford



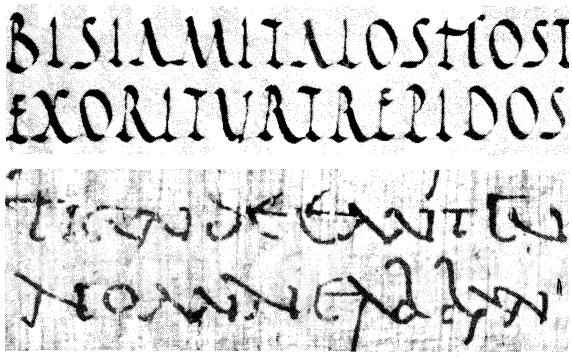
Model hand for the design of French Royal Greek type. Copy of Manuel Philes, made by Angelus Vergecius, 1564 (Oxford, Bodleian Library, MS. Auct. F.4.15, fol. 2).

Cursive capitals. The business hand of the 1st century, used for correspondence and for most documents, private and official alike, is known as cursive capitals. Here the pen, cut to a sharp point, was held at the same oblique angle but was lifted less often, and this "cursive" handling automatically produced new and simpler letter forms such as ∂ (two strokes) for D (three strokes) and ϵ (two strokes) for E (four strokes). Some of these new letter forms are "minuscule," in that parts of them ascend or descend beyond the body of the letter (h, q) instead of being confined between a pair of lines, as in the majuscule rustic capitals (H, Q).

Roman business hand

From the 2nd to the early 4th century, parchment was replacing papyrus as the standard writing material for books, and the codex was replacing the roll as their standard form. The evidence that survives from this period, during which biblical and other Christian literature was beginning to be copied extensively, is fragmentary, and its interpretation is still controversial. The main line of development, however, is clear enough. The elaborate letter forms of rustic capitals, with their numerous pen lifts, began to be abandoned, and experiments were made with new book hands in which the simplified letter forms of cursive capitals were written with a broad pen, sometimes held obliquely in the traditional way and sometimes held "straight," so that its thickest strokes fell at right angles to the line of writing. It was probably the use of a straight pen that produced, for example, the conversion of cursive capital ∂ (axis oblique) into the fully minuscule d (axis vertical).

Uncials, half uncials, and cursive minuscule. At the end of this period of transition, in the 4th and 5th centuries, when the evidence becomes more abundant, two new book



Early Roman capitals.
(Top) Rustic capitals. Codex Palatinus, Virgil, 4th–5th century AD (Vatican City, Biblioteca Apostolica Vaticana, Pal. Lat. 1631). (Bottom) Cursive capitals. Sale of a slave, AD 166 (British Museum, Pap. CCXXIX).

By courtesy of (top) the Biblioteca Apostolica Vaticana, (bottom) the trustees of the British Museum

hands and a new business hand are found in use. The older of the book hands, called uncials (the name dates only from the 18th century), was originally written with a slightly oblique pen; but, from the 6th century onward, a straight pen was used, and the hand began to look rounder and more contrived. Although it incorporates several of the cursive letter forms (δ, ε, h) of cursive capitals and has two forms peculiar to itself (ɑ, ɱ), it also preserves certain forms, such as B, N, R, S, which differ only a little from the forms of rustic capitals; and all three kinds of letters are treated as majuscules, being confined as far as possible between one pair of lines.

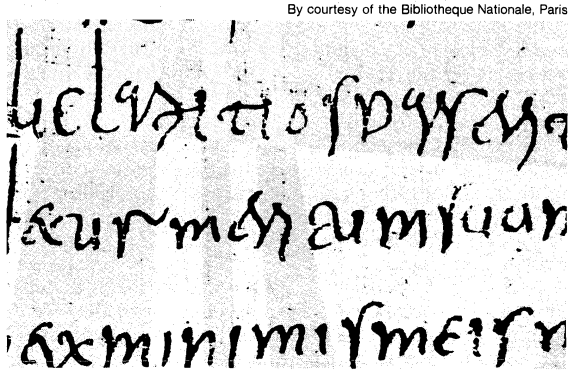
From the 4th to the early 7th century, most Christian books—biblical, patristic, and liturgical—were written in the uncial script, and even for pagan literature it almost entirely superseded rustic capitals. It survived the collapse of the Roman book trade. And, after the 6th century, when the production of all books, pagan as well as Christian, was taken over by the church—notably by the monasteries, such as the Vivarium founded in southern Italy by Cassiodorus, a scholar whose aim was to perpetuate Roman culture, and the houses that observed the Rule of St. Benedict—uncial script survived in many centres, especially for biblical and liturgical texts, down to the 9th century. Thereafter, like rustic capitals, uncials were used only for titles, and they, too, disappeared in the 12th century.

The younger of the two new book hands is called half uncial. This script was less popular than uncials and never broke their monopoly of biblical and liturgical texts, although like them it was still being written in the 8th century and even, as a display script for certain purposes, in the 9th. The artificial name half uncial tells nothing about the origin or nature of the script. It differs from early uncials in being written with a perfectly straight pen. One letter (N) remains more or less unchanged from the capital form, but the rest of the alphabet is cursive in origin. The letter forms that differ most from uncials are *a, b, d, g, m, r, s;* and the alphabet as a whole is frankly minuscule, since no attempt is made to confine it between a single pair of lines.

The new business hand of the 4th century and after is known as cursive minuscule. Like cursive capitals, it was written with a pointed pen, but the pen was held more or less straight. It, too, is a frankly minuscule alphabet and uses basically the same letter forms as half uncials, although the frequency in cursive minuscule of ligatures between letters tends to conceal the fundamental likeness between the two hands.

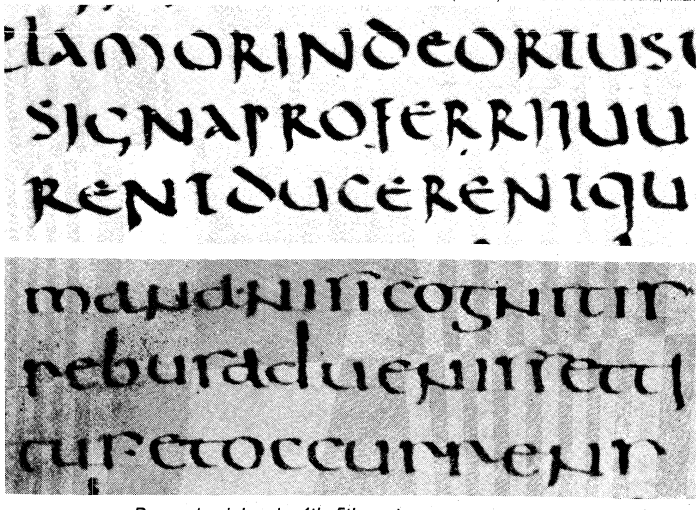
Cursive minuscule as a business hand

By courtesy of the Bibliotheque Nationale, Paris



Cursive minuscule. Avitus of Vienne, 6th century AD (Paris, Bibliothèque Nationale, Lat. 8913 and 8914).

By courtesy of (top) the Bibliotheque Nationale, Paris (bottom) the Biblioteca Ambrosiana, Milan



Roman book hands, 4th–5th century.
(Top) Uncials. Livy, 5th century (Paris, Bibliothèque Nationale, Lat. 5730). (Bottom) Half uncials. *De bello Judaico*, Hegesippus, 5th–6th century (Milan, Biblioteca Ambrosiana, C.105 inf.).

The letter forms that distinguish cursive minuscule and half uncials from rustic and cursive capitals and from uncials were evolved during the obscure period between the 1st and 4th centuries. The question of whether these forms were evolved in the sphere of the book hands or of the business hands is still undecided, but, whatever their origin, their importance for the subsequent history of European handwriting is paramount. They provided the material on which the Carolingian minuscule, first developed in the late 8th century, was based; and that script dominated Europe, in spite of severe modifications, until the end of the Middle Ages. Only in one other period were new letter forms evolved, between the 13th and the 15th centuries, in the group of scripts known as Gothic cursives; and the influence of these late innovations was ultimately cancelled out, thanks to the revival of Carolingian minuscule in a pure form by the Italian humanists at the beginning of the 15th century. (T.J.Br.)

The Anglo-Celtic and other "national" styles (5th to 13th century). From the 5th century the relaxation of imperial Roman authority brought on a reassertion and growth of native cultures—that is, wherever the people were not wholly occupied in a savage struggle against barbarians for mere existence. The most isolated places, such as the province of Britain, responded strongly to this opportunity and at the same time were able to conserve important elements of the Roman civilization. With Ireland, which was never under occupation by the legions, it offered during Europe's darkest age comparative peace and shelter for the development of the richest and most original of book styles.

The Insular manuscripts were produced at monasteries that were often on a barren rock in the sea or at an equally inaccessible site. According to tradition, the earliest centre of Christian learning in Ireland was established by the Romano-British apostle St. Patrick (fl. 5th century). A great successor, St. Columba, or Columcille (*c.* 521–597), whom legend credits with divine scribal powers, founded monastic houses at Derry and Durrow and then journeyed

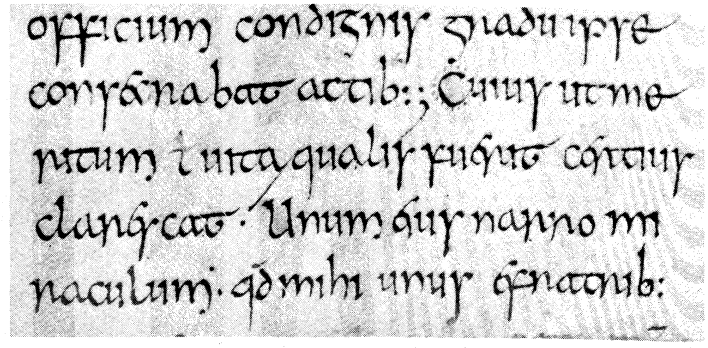
Anglo-Celtic monastic centres of manuscript copying and illumination

to the Inner Hebrides to found one on the lonely island of Iona *c.* 563. St. Columban, another Irish missionary, in much the same period was founding monasteries on the Continent: *c.* 590 in Gaul (modern France) the Burgundian centre Luxeuil, from which Corbie in Picardy was organized, and St. Gall in Switzerland and Bobbio in Italy about the years 612 to 614. From Iona a daughter house was founded in 635 on St. Cuthbert's holy isle of Lindisfarne just off the Northumbrian coast of England. To the south the Northumbrian monk, later abbot and saint, Benedict Biscop (*c.* 628–689/690) established the twin monasteries of St. Peter at Wearmouth in 674 and St. Paul at Jarrow in 682. He endowed them with splendid collections of books and pictures gathered during repeated visits to Rome, so that, in the late 7th and early 8th centuries, they constituted the most flourishing centre of Christian scholarship in western Europe and the meeting place of Hiberno-British and continental influences.

For the fine books made in the Anglo-Celtic centres, the majuscular script called Insular half uncial was deemed suitable rather than the pointed, more cursive Irish minuscule used for documents and vernacular texts. There is a high degree of conformity, attesting to their stylistic maturity, among such manuscripts as the Book of Kells (Trinity College, Dublin) and the Lindisfarne Gospels (British Museum), individual as they are in detail and ornament. After all, there is room for infinite variation where, in one-quarter of a square inch, 158 interlacements have been traced unerringly—by angels, it is said. The Book of Kells, Codex Cenannensis to paleographers, was probably produced at Iona around 800. It has 339 leaves, 13 by 10 inches (33 by 25 centimetres) of noble script in single column, jet black on well-made parchment, through which runs the most spirited and colourful of ornamentation, ranging from the red-dotted outlining of letters, which is as much a feature of the style as the wedge-topped ascenders, to the wildly extravagant full-page initials at the opening of Gospels. The other masterpiece of Anglo-Celtic calligraphy and illumination, the Codex Lindisfarnensis, was written in honour of St. Cuthbert shortly after his death in 687. It displays the same lively inventiveness, the love of fantastic animal and bird forms (zoomorphs), intricate interlacing, and even, rhythmic script, set off by generous margins.

The earliest of all extant manuscripts of the Insular style is the *Cathach* ("Battler") of St. Columba (Royal Irish Academy, Dublin), who, according to legend, wrote it himself and, in the judgment of scholars, may actually have done so. Housed in its *cumhdach* (a sort of ark), it was carried into battle to ensure victory.

Besides the proud witness of such books as these to the Anglo-Celtic contribution, there were also the productions of continental centres influenced by St. Columban and his disciples, as well as books mainly in the Roman tradition but carrying the unmistakable sign of Insular influence. For instance, there are three that scholars believe were written in the 7th century at Bobbio, Italy, in the monastery of St. Columban. They are Codex Usserianus Primus, now a treasure of Trinity College, Dublin, and two manuscripts preserved in the Biblioteca Ambrosiana, Milan, known as



Insular minuscule. *Historia ecclesiastica*, by Bede, 8th century (British Museum, Cotton Tiberius C.11).
By courtesy of the trustees of the British Museum

Codex Ambrosianus C.26 sup. and Codex Ambrosianus D.23 sup. There is another, Codex Amiatinus (Biblioteca Medicea-Laurenziana, Florence), of 1,030 leaves measuring 20 by 13½ inches (51 by 34 centimetres), made in Northumbria in the 8th century. It is continental Roman in style with no concession to the Insular habit of ornamentation. This is understandable, for it was designed for presentation to the pope.

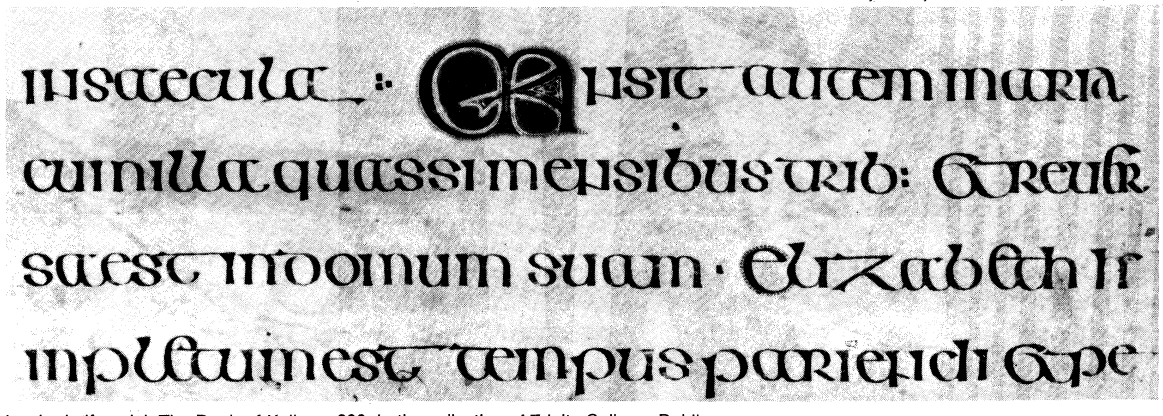
Though the Insular minuscule was ready to hand, the majuscular half uncial as the senior script was always given the place of honour and the preference for the fine Latin books of the Anglo-Celtic monasteries. Nevertheless, by the 8th century the minuscule was developing into a disciplined book hand, as seen in the copy of Bede's *Historia ecclesiastica* (*c.* 731; Cambridge University Library). The spiky, ligatured, compactly written style migrated early to the Continent and, by the beginning of the 8th century, was at home in the Anglo-Saxon foundation of Echternach, in Luxembourg. Fulda and Würzburg, in Germany, were other important centres abroad of Insular culture and book production in this style.

Merovingian and Visigothic scripts

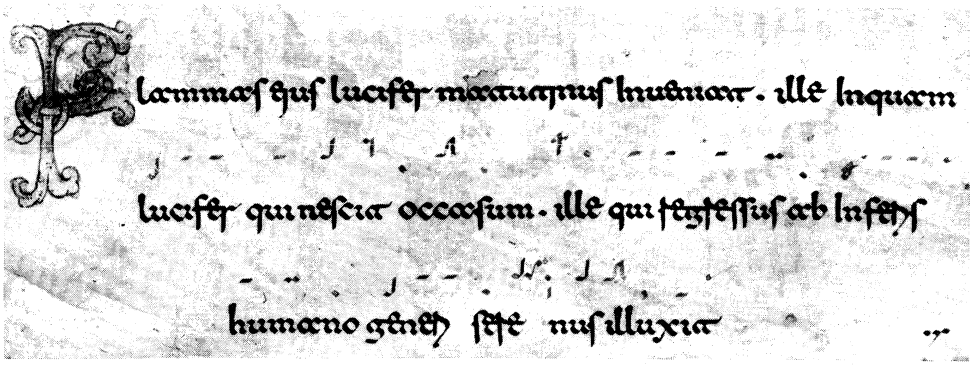
The Merovingian, in France, and the Visigothic, in Spain, are two more varieties of minuscular script that grew out of Latin cursive after the withdrawal of the Roman authority. In the Luxeuil monastery, in Burgundy, the minuscule attained in the 7th century the characteristics of a fine book hand but for only a short period, when the reforms under Charlemagne took effect. In the Iberian Peninsula the Visigothic style was in use from at least the 8th to the 12th century. It has the verticality of emphasis that is common to the other hands out of the same cursive background, when deliberately written, with weighted ascenders carefully topped by flat serifs.

The southern Italian script of the style called Beneventan, nurtured in the motherhouse of the Benedictine Order at Monte Cassino, was the "national" hand that rose to the status of calligraphy and held its position well into the 13th century, an active literary life of more than 500 years. This type of script has a peculiar jerky rhythm and retains individual cursive forms, which, together with the abundance of abbreviations and ligatures, make reading quite difficult.

By courtesy of The Board of Trinity College, Dublin



Insular half uncial. The Book of Kells, *c.* 800. In the collection of Trinity College, Dublin.



Beneventan script. Exultet Roll from Monte Cassino, Italy, late 11th or early 12th century (British Museum, MS. 30377).

By courtesy of the trustees of the British Museum

Carolingian reforms in the scriptorium (8th and 9th centuries). The literary reforms carried on in the latter part of the 8th century and early 9th century by order of the Holy Roman emperor Charlemagne set the highest of standards for the making of books throughout his Western empire. The extensive educational program, looking forward to new authorized versions of the Vulgate, the missal, and other liturgical works, he placed in the charge of the learned English cleric Alcuin of York. The Emperor persuaded Alcuin to leave his position at the head of the cathedral school of York and the excellent library he had gathered there, first to become master of the palace school at Aachen, then at Tours as abbot of St. Martin's to conduct the literary activities centred at the well-established scriptorium (writing room) there.

Contributions of Alcuin

Before taking up the abbacy in 796, Alcuin was responsible for, or at least inspirer of, the most precious of Carolingian codices, the so-called Golden Gospels. These were a series of illuminated masterpieces written mainly in gold and often on purple-stained vellum. The most famous is the Godescalc Gospels (Bibliothèque Nationale, Paris), written before 783 for Charlemagne, the body of the text in uncials and the dedication in Carolingian minuscule. The most luxurious is the Saint-Médard Gospel Book (Bibliothèque Nationale), written entirely in gold uncials and illuminated with full miniatures, initials, etc., in gold and silver on purple ground.

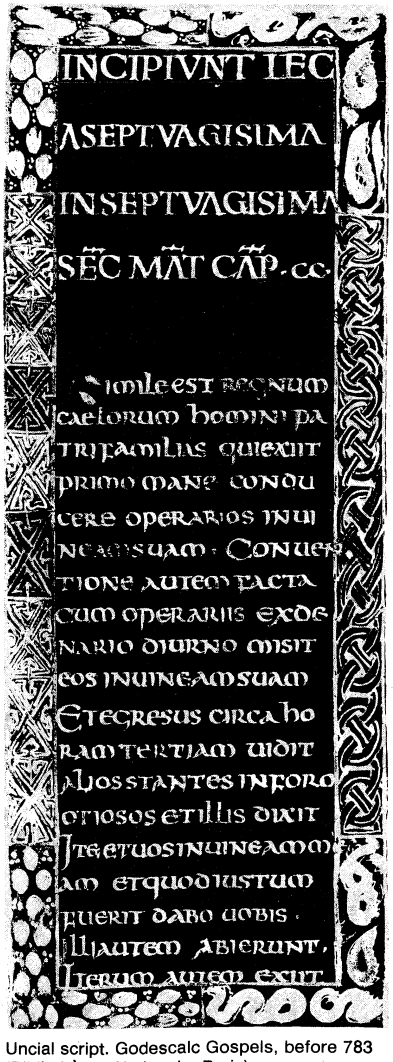
Alcuin carried forward the work of the St. Martin scriptorium in the spirit of a true classical renascence. Each variety of traditional letter form was studied with a view to finding its norm by careful comparison with archetypes in ancient monuments and books. Thus, the square capitals, at the top of the hierarchy of scripts, were modeled on Augustan inscriptions. For rustic capitals (*capitales rusticae*), the reformers adopted the style of those used for the text of such fine old codices as the Virgil fragment now in the Vatican. The roman uncial was restored to its simple dignity, as in well-made books of the 6th century, and the minor, or half, uncial was likewise restored to the plain elegance of that earlier period, after the degeneracy of the 7th century.

Carolingian minuscule

The model for the most valuable and characteristic of all the Alcuinian contributions, the Carolingian minuscule, has never been precisely determined. It may well have been a local variety of cursive or, more probably, a mixture of half uncial and cursive, in which Alcuin discerned the possibilities of his clear, round, flexible but disciplined script, comfortable to both scribe and reader. For, as regularized at the scriptorium of St. Martin, the minuscule was written with the shaft of the pen pointing somewhat to the right instead of straight back over the shoulder; though the letters were formed deliberately, even and round, stroke by stroke the same way every time by rule, the writing brought relative ease for the hand and eye. With the years some cursive features became more prominent—*e.g.*, a tendency nearly to join certain letters and an occasional hint of "italic" in the slightly sloping, even, well-spaced lines. The incipits (the opening words of the text) were celebrated by means of display letters and a decorative initial that might come from any one of a number of sources, including Insular, Byzantine, and Merovingian scripts. Otherwise, the classic calm was maintained. The crowning achievement of the Tours school of scholars, scribes, and artists was reached in the mid-9th century, under Alcuin's successors, in the Gospels of Lothair (Bibliothèque Nationale).

There are observable variations among the different Carolingian schools, but these are generally in small details. The most surprising departure is the Utrecht Psalter (University Library, Utrecht), written at Rheims, in rustic capitals and illustrated with fluent pen drawings in the Hellenistic fashion. Apparently, the whole work was devotedly copied in the 9th century from an old model.

By courtesy of the Bibliotheque Nationale, Paris



Uncial script. Godescalc Gospels, before 783 (Bibliothèque Nationale, Paris).

Gospels of Lothair, *c.* 850 (Paris, Bibliothèque Nationale, MS. Lat. 266, fol. 19).

By courtesy of the Bibliotheque Nationale, Paris

The black-letter, or Gothic, style (9th to 15th century). Carolingian minuscule remained the unrivaled book hand of western Europe through the 9th century, or nearly so. Then a trend away from the official imperial standards set in. It can be observed progressing in the manuscripts written at St. Gall, in Switzerland, near the end of the 9th century and during the 10th. There is a tendency toward lateral compression of the letters. This begins as the natural result of an easier motion of the pen held with the shaft out to the side rather than pointing back over the right shoulder. Scribes learned to cut the pen's writing edge obliquely so that it would be parallel to the top of the page even though the shaft was held in the slanting position, yielding a perpendicular stroke of maximum width. They were led on by the attractive novelty of bolder and bolder contrasts that eventually were to appear to the eyes of the Italian humanists in the Renaissance so brutal as to deserve the bad name of Gothic. Nevertheless, the more condensed, compact writing allowed significant economies in the amount of time taken in writing the books and the quantity of materials used and, therefore, in the cost of finished manuscripts.

Black-letter style of increasing density deepened the "colour" of the page and imparted to the northern, advanced, formal variety of book hand a matted aspect, or fanciful likeness to woven fabric, that gave rise to the name Textura. It is called by paleographers *littera textualis formata* or *lettre de forme.* As the script developed through the 11th and 12th centuries in Germany, France, and England, its curves broke into angles. During the 13th and 14th centuries the size of writing was generally reduced. It became stiffer, and though in the following century (its last of undisputed sway) the script regained size, the proportions were disagreeably narrowed, and the letters, in lines as rigid and mechanically perfect as a picket fence, have angles sprouting hairlines apparently added with a crow quill afterward. Of this species of formal black-letter book hand, two kinds are distinguished by paleographers. One stands flat and unserifed on the base line and is known as *textus prescissus.* The other is completed with square or diamond-shaped feet and is referred to as *textus quadratus.* Though the former, and earlier, variety has superior claims for the calligrapher, the latter is the variety that was carried into printing types and hence into much wider use.

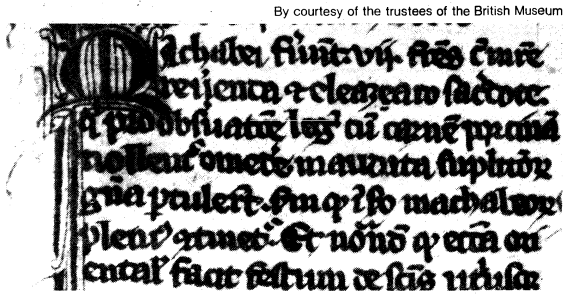
Lettre de forme

In Italy the writing during the same period took on weight, but the curves of comparable book script never became angles. The senior script there was the rotunda, heavy but not pointed. The form persisted for liturgical work both as writing and printing type until late in the Renaissance.

The north and the south had, of course, their other kinds of writing for court and business or personal uses. A cursive hand that flourished in France, Flanders, and England rose to favour in the 15th century as a vernacular book script. This *littera bastarda,* or *lettre bâtarde* as it is termed in the vexed nomenclature of paleography, for all its high style as attained in fashionable books, was close kin to the epistolary "running secretary" commonly written by northern Europeans and early American settlers until it grudgingly yielded in the 17th century to italic script.

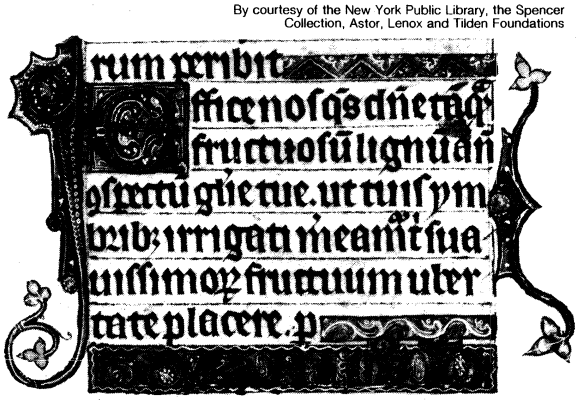
Lettre bâtarde

By courtesy of the trustees of the British Museum



Gothic book hand. *Legenda aurea* by Jacobus de Voragine, 1312 (British Museum, Add. 11,882).

By courtesy of the New York Public Library, the Spencer Collection, Astor, Lenox and Tilden Foundations



Textus prescissus by Brother Tickhill. "Beatus" page from the Tickhill Psalter, *c.* 1310. In the collection of the New York Public Library.

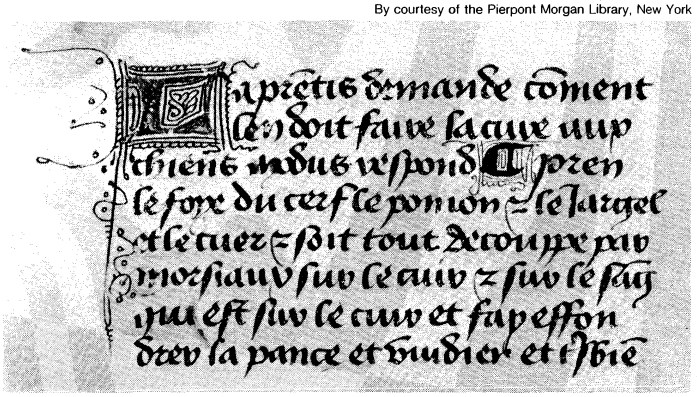
The scripts of humanism (14th to 16th century). Under the inspiration of the 14th-century Italian poet Petrarch, who started the collecting of ancient manuscripts, coins, medals, and other "antiques," the essentially literary movement called humanism engaged a group of scholars at Florence in the latter part of the 14th century and opening decades of the 15th. Their growing enthusiasm for antiquity led them on an ever-widening search for Latin books and ancient monuments, in order that they might restore the lost heritage of Rome. Classical authors written in the clear Carolingian minuscule with display lines in lapidary style (*i.e.,* in the style of ancient inscriptions cut on stone) appealed to them as though straight from the time of Cicero instead of from the 10th to 12th century. Reverently, Coluccio Salutati, the late 14th-century chancellor of Florence who followed Petrarch as leader of the movement, and his fellow humanists imitated the old script, which they spoke of as the *lettera antica* to distinguish it from the ordinary *moderna* black-letter style.

Lettera antica and *moderna*

Two protégés of Salutati are credited with developing,

on the basis of these studies and experiments among old manuscripts, the two fundamental scripts of humanism. Poggio, at the very beginning of the 15th century, produced the round, formal writing that, after polishing by a generation of scribes, served the new art of printing as prototype of "roman" fonts. He promptly followed up this achievement by introducing into his manuscripts the square capital letters as used on ancient monuments. Later in the century the rage for epigraphic inscriptions brought such enthusiasts as Cyriacus of Ancona, Felice Feliciano of Verona, and his fellow townsman Giovanni Giocondo into the field and, from Padua, Giovanni Marcanova, Bartolomeo Sanvito, and Andrea Mantegna. They compiled their researches in great *sillogi,* or anthologies, which, among other uses, gave calligraphers authoritative patterns for letter forms, ornament, and the correct spacing and placing of all display elements in their books. Feliciano, a calligrapher fond of ornament and fertile in invention, about 1460 first demonstrated how to construct the monumental roman capitals according to geometric rules.

By courtesy of the Pierpont Morgan Library, New York



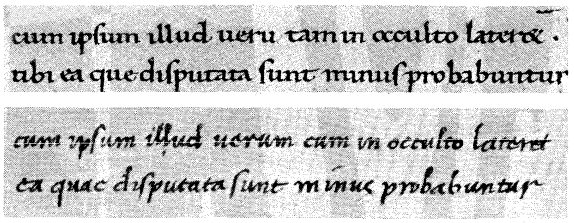
Lettre bâtarde by Henri de Ferrieres. *Livre du Roy Modus et de la Reine Ratio, c.* 1435 (New York City, Pierpont Morgan Library, M.820, fol. 16v).

The second style of humanistic script appears earliest in the writings of Poggio's friend Niccolò Niccoli, who was also an accomplished scribe. His slightly inclined cursive, written with a fairly narrow rounded nib at a good rate of speed, was to be to the printers' "italic" what the Poggian hand became to their "roman." Niccoli's innovation employed movements and rhythms close to those of the ordinary black-letter cursive familiar in everyday affairs. Indeed the script contains a sprinkling of current black-letter mannerisms and lends itself similarly to the joining or tying together of letters. The special character of this *antica corsiva* results from the narrowing of the bodies of letters due to the rapid up-and-down movement of the pen, facilitated by its being held with the shaft pointing at an angle away from the shoulder instead of straight back, producing this ✕, not +. As in italic fonts to this day, the form of *a* is distinctive; and *f, g, k,* and ſ (long *s*) are all more or less reminiscent of black-letter current habits. For his headings Niccoli preferred roman capitals "italicized" by a slight inclination to the right.

Antica corsiva

Both scripts were at once taken up and spread by other able scribes working at Florence in the first half of the 15th century, of whom the work of Giovanni Aretino, Giacomo Curlo, and Antonio di Mario, among others, is well recognized.

Poggio himself in 1403 had promptly carried his new script to Rome, where he later became papal secretary. Both scripts were devoted exclusively to the service of Latin literature, but there was a difference. Poggio used his hand as a calligrapher, while Niccoli used his as a copyist. The manuscripts of the former are set forth on fine parchment with meticulous care to formal details, such as even lines at the right-hand margin, and with handsome embellishment. Those of the latter are on paper, compactly and rapidly written, with attention to legibility and textual accuracy above all. There is an interesting parallel in the printed books of the following century. Ambitious



Humanistic scripts.
(Top) *De oratore,* by Cicero, calligraphy by Poggio, *c.* 1425 (Florence, Biblioteca Medicea-Laurenziana, MS. Laur, Plut. 50.31. C. 166). (Bottom) *Cicero,* calligraphy by Niccolò Niccoli, *c.* 1423 (Florence, Biblioteca Nazionale Centrale, Soppr., 1.1.14).

By courtesy of (top) the Biblioteca Medicea Laurenziana, Florence, (bottom) the Biblioteca Nazionale, Florence

Renaissance folios are set in fine roman types, while the well-edited but cheap little student books are just as naturally set in italic.

Typographic printing displaced the copyists. At the same time printing gave impetus and new significance to the work of the calligraphers. They accepted the challenge of mechanized writing and for a while turned out the finest of humanistic masterpieces. In the late 15th and early 16th centuries the Paduan Sanvito and Pierantonio Sallando of Mantua, for instance, not only wrote the round humanistic script in a fashion worthy of the richest miniatures and illumination but also honoured the slender proportioned "italic" script by promoting it to a place in some of their proudest and most precious manuscripts.

Influence of printing

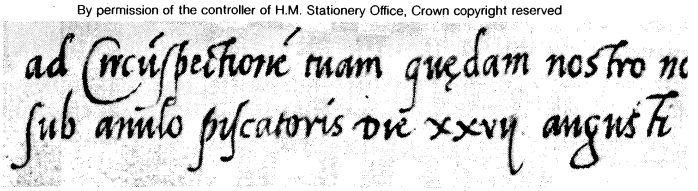
Sanvito's folio and octavo classics in the humanistic cursive are also celebrated for vellum pages stained purple, yellow, green, or salmon pink and for lines of inscriptional capitals alternating gold, blue, lake, purple, violet, and green. The *antica corsiva* perfected by 15th-century papal scribes for rapidly inditing briefs issued by the chancery also won its way as the chosen medium of polite correspondence. Thus, in the 16th century the versatile *lettera da brevi,* or *cancellaresca* (chancery cursive), lively yet disciplined, responsive to wide variety in cut of nib and speed of movement, attainable by the novice and gratifying to the adept either as book or epistolary hand, became a vehicle of the new learning throughout Christendom.

Cancellaresca, or chancery cursive

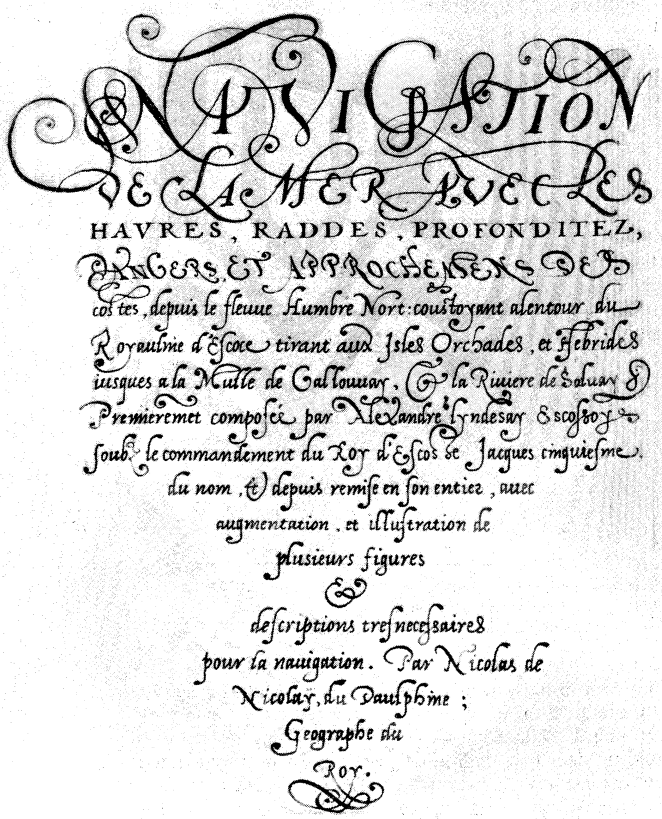
As written by the early 16th-century calligrapher and printer Lodovico degli Arrighi of Vicenza, the *cancellaresca* can range from eye-arresting contrasts of almost Gothic thick-and-thin strokes to a delicate, supple monotone tracery. The ascending letters, instead of terminating in serifs as with Sanvito or Sallando, now wave plumelike to the right, and the descenders are tending to balance them with a swing to the left. In 1522 Arrighi published at Rome his modest though revolutionary work *Operina da imparare di scrivere littera cancellaresca.* This, the earliest of printed writing manuals, or copybooks, held out to the public clear, simple directions with woodcut examples and invited everyone to learn in a few days how to write this fashionable hand for themselves. In effect, it announced the end of the era of the scriptorium and the beginning of the era of the writing master. Dependence upon attracting pupils and gaining a reputation with the public was thereafter reflected in the tendency to exploit novel or flashy scripts and extravagant ornament.

Writing manuals and copybooks (16th to 18th century). The Arrighi *Operina's* devotion to the chancery cursive is matched in single-mindedness only by Gerardus Mercator's *Literarum Latinarum* (1540). Arrighi's second publication, *Il modo de temperare le penne* (1523), is a more

By permission of the controller of H.M. Stationery Office, Crown copyright reserved



Littera cancellaresca. Brief of Pope Leo X, ascribed to Lodovico degli Arrighi, 1519 (London, Public Record Office, S.P. 1/19).


NAVIGATION
DE LA MER AVEC LES
HAVRES, RADDES, PROFONDITEZ,
DANGERS ET APPROCHEMENS DES
costes, depuis le fleuue Humbre Nort: coustoyant alentour du
Royaulme d'Escoce tirant aux Isles Orcbades, et Hebrides
iusques a la Mulle de Gallouuay, & la Riuiere de Saluay &
Premieremet composee par Alexandre Lyndesay Escosoy
soub le commandement du Roy d'Escos se Jacques cinquiesme
du nom, & depuis remise en son entiez, auec
augmentation, et illustration de
plusieurs figures
&
descriptions tresnecessaires
pour la nauigation. Par Nicolas de
Nicolay, du Dauphine;
Geographe du
Roy.

Chancery cursive attributed to Pierre Hamon. *Navigation,* title page from the Harleyian manuscript, *c.* 1560 (British Museum, MS. 3996, fol. 1).

By courtesy of the trustees of the British Museum

normal performance for the calligrapher, notary, printer, and erstwhile scriptor of apostolic briefs in the role of writing master. This book redresses the balance in favour of the popular scripts.

Arrighi's elder contemporary and possible mentor, Giovanni Antonio Tagliente, writing master to the Venetian chancery, published his *Presente libro* in 1524, two years after the *Operina;* and Giovanni Battista Palatino published in 1540 at Rome his *Libro nuovo d'imparare a scrivere.* Though Tagliente was master of an elegant *cancellaresca*—his "living hand" is displayed in a holograph supplication to the Doge and council of Venice, 1491—as well as of the black-letter hands, he was out-of-date and prey to the vices of the old professional; *e.g.,* his chancery specimens include a page of script leaning excessively to the right, and facing it is a page of similar script leaning away to the left, almost equally hard to read and useless. Unlike the author's own hand in his 1491 supplication, the woodcut models here are acutely pointed. Palatino at the same time exhibits a rigid, sharp-angled *cancellaresca formata* in his excellent manuscript specimen (Bodleian Library, Canon. Ital. 196 fol. 44r), while his popular manuals offer relatively easygoing models for learners. In particular, the Palatino *cancellaresca romana* is a normal even-toned script, modest as to bulbed ascenders (*testegiatta*) and angle of inclination, displaying in its forms hospitality for *bastarda* variations (*i.e.,* Gothicizing *g, h, k, r*). The fact that Palatino's *Libro nuovo,* besides being the most complete, was the most widely disseminated of mid-16th-century books on writing bespeaks certain latitudinarian qualities along with marvellous virtuosity.

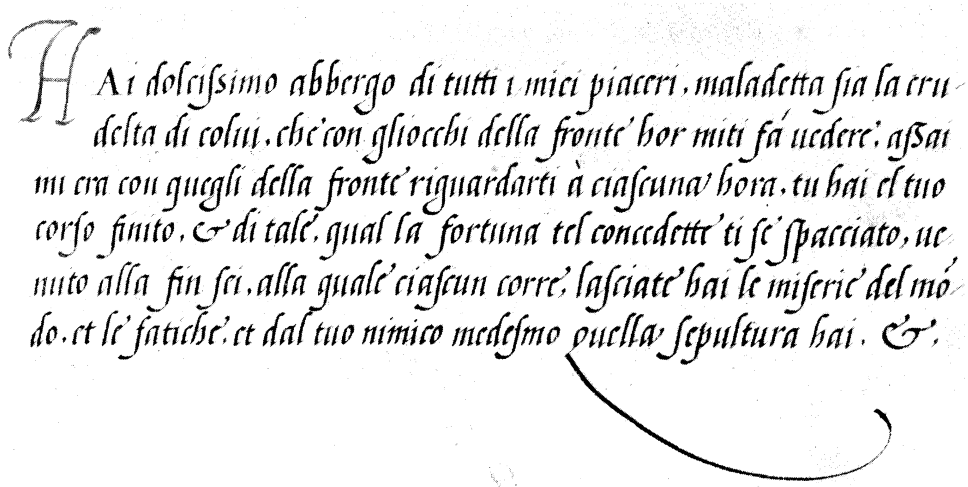
Spread of chancery cursive

The propagation of the chancery cursive abroad was furthered by native manuals too. The Latin letters called italics were introduced into Germany by Casper Neff's *Thesaurium artis scriptoriae* (1549); into Spain by a disciple of Palatino, Juan de Yciar, in his *Recopilacion subtilissima* (1548); into France by Pierre Hamon's *Alphabet de l'invention des lettres en diverses escritures* (1561); and into England by means of *A Booke Containing Divers Sortes of Hands* (1570) by Jean de Beauchesne and John Baildon.

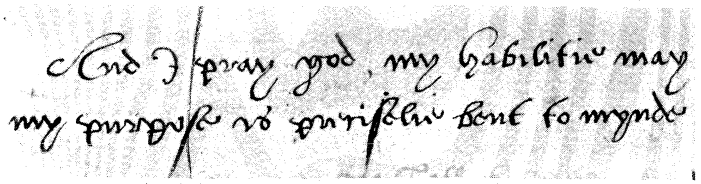
The pure Italian chancery hand was the favourite of court circles and humanist scholars through the second half of the 16th century. In England Roger Ascham, mid-16th-century schoolmaster to kings and queens, wrote and taught an exemplary *cancellaresca,* as did the late 16th-century Cambridge don Bartholomew Dodington. These and other educated Europeans, including Shakespeare, generally wrote the black-letter "running secretary." For their part, the writing masters, in striving to reach an ever larger public, increasingly emphasized the compromise script intended to bring learners already indoctrinated in the common Gothic cursive to command more readily the fashionable italic style. The *Opera nella quale si insegna a scrivere,* of Vespasiano Amphiareo (Albertacci), had already offered at Venice in 1554 models that combined the overdisciplined strokes of Palatino with elements of black-letter mercantile hands (*e.g.,* loops and running ligatures); to this hybrid *cancellaresca* that he claimed as his own invention, Amphiareo gave the accurate name of *bastarda.* Then Gianfrancesco Cresci of Milan, a Vatican scriptor, published his *Essemplare* (1560), to herald the oncoming Baroque and to reject vehemently the works and ways of Palatino and all his academic sort. He replaced their most stylish broad, chisel-bevelled nib and meticulous building up of disjoined strokes with a narrow, rounded-off pen flexible enough to respond to pressure and fluent in dashing off his much-inclined italic *bastarda* topped off

Importance of Cresci

By courtesy of the Bodleian Library, Oxford


Hai dolcissimo abbergo di tutti i miei piaceri, maladetta sia la crudelta di colui, che'con ghiocchi della fronte' hor miti fá uedere; assai mi era con quegli della fronte' riguardarti à ciascuna' hora, tu hai el tuo corso finito, & di tale' qual la fortuna tel concedette' ti se' spacciato, ue nuto alla fin sei, alla quale' ciascun corre' lasciate' hai le miserie' del mó do, et le' fatiche' et dal tuo nimico medesmo quella' sepultura hai, &'.

Cancellaresca formata by Giovanni Battista Palatino. Manuscript specimen book, *c.* 1550 (Oxford, Bodleian Library, MS. Canon. Ital. 196, fol. 44r).

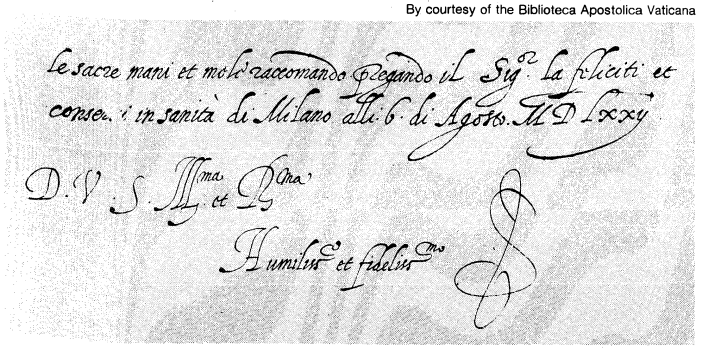


Running secretary hand. Letter by Roger Ascham, 1552 (British Museum, Landsdowne 3).
By courtesy of the trustees of the British Museum

dramatically with bulbous *testegiatta.* It is not too much to say that Cresci's script, with the handsome set of swash capitals (to borrow the typographic term for these forms), not only established a revolutionary front for 17th-century calligraphy but also contained the germs of 18th-century "copperplate" and even Spencerian scripts. Cresci was the first in Italy to take advantage of copperplate engraving, though Giuliantonio Hercolani's *Lo Scrittor Utile* (1574; plates engraved in 1571) was a close rival.

The principal French hands of the period were the national black-letter cursive called *lettre françois* or *financière,* which was commonly used in ordinary affairs, and the aristocratic *italienne bastarde.* In the Low Countries, examples of excellence are provided by Clement Perret in his *Exercitatio alphabetica* (1569), and by Jan van de Velde in his work *Spieghel der Schriftkonste* (1605), the fame of which was carried down through many generations of English penmen.

By courtesy of the Biblioteca Apostolica Vaticana



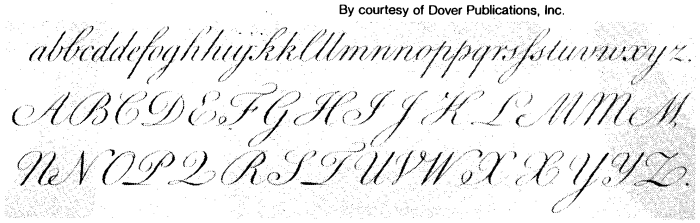
Italic *bastarda.* Letter by Gianfrancesco Cresci, 1572 (Vatican City, Biblioteca Apostolica Vaticana, Lat. 6185, fol. 135 R).

In Britain and its empire, the *italienne bastarde* was largely disseminated through Edward Cocker's engraved copybooks and manuals. Known variously as the "new mixt current" or the "speedy à la mode," its concessions to the black-letter running secretary include the *e* looped at the top like a latter-day Palmer Method *o* and what has been termed the upside-down *r,* which is easily misread as *u* by humanistic eyes. Nevertheless, in the second half of the 17th century, italic writing won decisively. By the opening of the 18th century, a chastened, businesslike version was developed in London by John Ayres and his younger contemporary writing masters eager to serve commercial demands. The small, narrow variety of Italian-French script, known as the Italian hand, was relegated to ladies' use as something suited to frail capacities, while the robust strain, called English round hand or "copperplate," was considered fit to dominate an epoch of trade supremacy.

Though the transition from black letter had been even more accelerated in the American colonies, Boston in New England was one of the last bastions of calligraphy in the trend toward countinghouse and commercial-college penmanship. The traditional foundations were laid solidly by the 18th-century writing master Abiah Holbrook, who raised a group of young Bostonian proficients at the South Writing School and left a fine manuscript, *The Writing Master's Amusement* (1767), in personal testimony.

Penmanship in colonial America

By courtesy of Dover Publications, Inc.



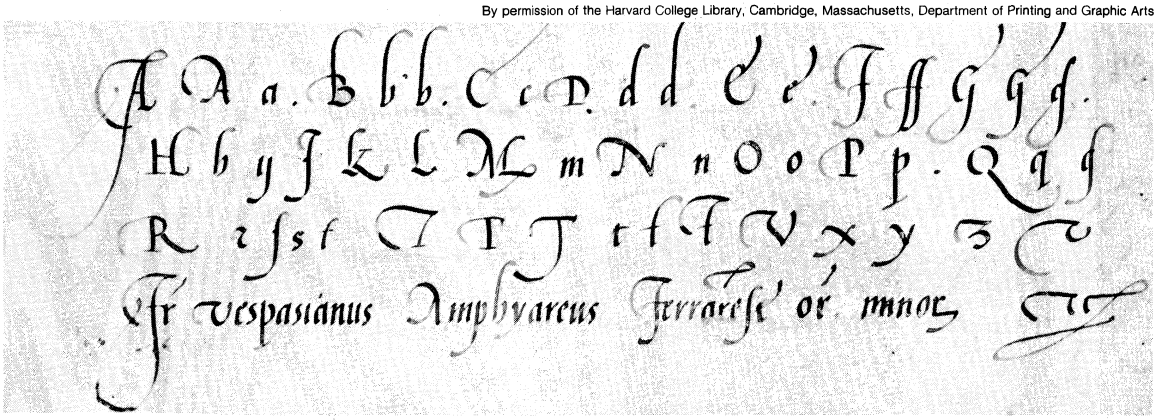
Copperplate, or English round hand. From *The Universal Penman,* Philip Hofer (ed.), engraved by George Bickham, 1743. Facsimile edition, Dover Publications, Inc., 1941.

As the 19th century advanced, the competition constantly mounted among systematists emphasizing the plain, practical business hand and decrying the "ornamental branches." Those who loved to flourish quill pictures of bounding deer or calligraphic portraits of national heroes were increasingly placed on the defensive. Certain religious sects clung to their heritage of individual styles of writing as an art. The Shakers, under the influence of visions, wrote and drew elaborate spiritual manuscripts. Most of these pieces that still remain are in the round hand and Spencerian scripts of their contemporary world. Other religious sects, such as the Moravians and Mennonites settled in Pennsylvania, produced out of their background old-country culture bold and handsomely coloured decorative pieces generally called Frakturs because of the script in which their so-called "Pennsylvania Dutch" is presented.

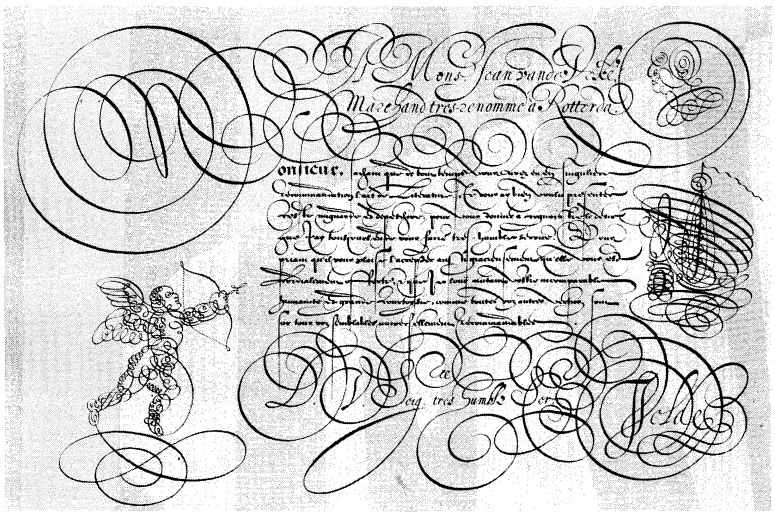
Revival of calligraphy (19th and 20th centuries). The revival of calligraphy at the end of the 19th century was part of an artistic revolt against the mechanization of life. About 1870 the English writer and artist William Morris had begun to concentrate attention on the ancient practice of scribes and to study with pen on parchment the means of achieving similar results. In this fashion he wrote out and illuminated a number of texts, humanistic and medieval, in the years before he took up similarly the study of 15th-century printing prior to establishing the Kelmscott Press. His searching inquiries and patronage led papermakers and ink suppliers, among others, back to forgotten standards.

Among those who heeded the message was Edward John-

By permission of the Harvard College Library, Cambridge, Massachusetts, Department of Printing and Graphic Arts



Bastarda script. Writing manual by Vespasiano Amphiareo (Albertacci), *c.* 1548. In the Harvard College Library, Cambridge, Mass.



Spieghel der Schriftkonste, calligraphy by Jan van de Velde, 1605. In the Columbia University Libraries, New York.

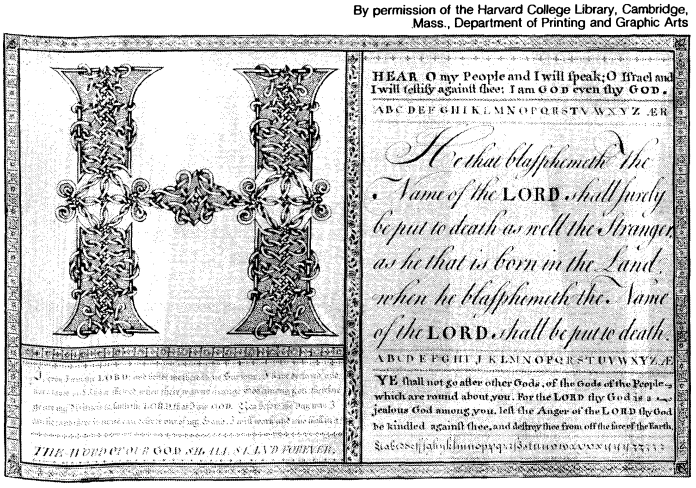
By courtesy of the Columbia University Libraries, New York

Influence of Edward Johnston

ston. On the basis of the Morris investigations and guided particularly by Sydney Cockerell, Morris' secretary, he rediscovered how to cut and sharpen reeds and quills, how to hold the chisel-nibbed pen so as to produce the kinds of writing that had become familiar in his study of old manuscripts, and every detail of how to make manuscripts. In 1899 he began teaching in London, first at the Central School of Arts and Crafts and, two years later, at the Royal College of Art. His early pupils soon became esteemed designers of lettering and influential calligraphers—*e.g.,* Anna Simons, Eric Gill, Noel Rooke, Graily Hewitt, and Percy Smith. In 1906 he published the basic manual *Writing and Illuminating, and Lettering,* 500 pages profusely illustrated with his own and Rooke's drawings and with reproductions from historic manuscripts. In 1910 the book was published in Germany, where it supported the work independently begun by Rudolf von Larisch in Vienna and carried on by Rudolf Koch and Anna Simons. Under the Nazi regime, the Gothic exponents enjoyed a resurgence, but, after the war, such skilled young German calligraphers as Hermann Zapf led the way back to roman and italic script styles. Jan Tschichold's work has been influential in the appreciation of classical letters for present-day uses.

Art schools everywhere gradually followed the Royal College of Art in offering courses in lettering and writing; these, along with sound forms, instilled a sense of responsibility to a rich heritage. Accordingly, type design passed from the engineers into the hands of artists and calligraphical scholars, such as Stanley Morison, Jan van Krimpen, Akke Kumlien, Bruce Rogers, Frederic W. Goudy, Victor Hammer, Berthold Wolpe, and Warren Chappell. The "freezing" of fine alphabets as font material stabilizes them and multiplies their influence. The effects were not limited to the printed or manuscript page or to book covers and jackets; all kinds of private and public markers and memorials, addresses, diplomas, and even the politicians' ubiquitous testimonial "scrolls" were touched by reform or at least by some consciousness of shortcoming.

By permission of the Harvard College Library, Cambridge, Mass., Department of Printing and Graphic Arts

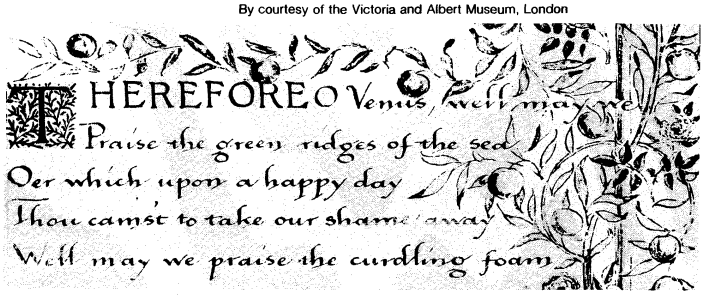


The Writing Master's Amusement, by Abiah Holbrook, 1767. In the Harvard College Library, Cambridge, Mass.

Though the United States was a decade behind Britain in organizing an arts and crafts movement, four years before Johnston's book appeared a Boston architect, Frank Chouteau Brown, brought out a surprisingly sophisticated book, *Letters and Lettering: A Treatise with 200 Examples* (1902). He praised Morris' initiative but wished for more classical balance; he pointed to the neglect of italics and gave them generous space, including specimens from the Lucas 1577 *Arte de escrivir,* "letra del Grifo," *bastarda,* and others. In 1905 *The Parable of the Prodigal Son,* fashioned like a humanistic manuscript by William Addison Dwiggins and printed from photoengraved plates, was offered for sale. Dwiggins, as illustration for a talk on early writing books, copied a page out of the 1542 Tagliente he had picked up in Europe and passed out prints of it at the Dec. 29, 1913, meeting of the Society of Printers—thereby leading the revival of *cancellaresca.*

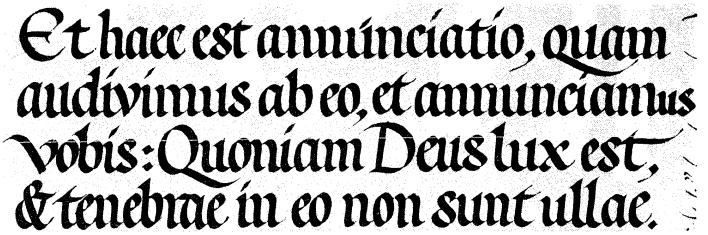
Revival of calligraphy in the United States

By courtesy of the Victoria and Albert Museum, London



"Praise of Venus," from *A Book of Verse,* calligraphy by William Morris, 1870. In the Victoria and Albert Museum, London.

Edward Johnston's pupils, and their pupils, organized in 1921 the Society of Scribes and Illuminators, "zealously directed toward the production of books and documents" by hand and the advancement of the crafts of member scribes, gilders, and illuminators. The program of this professional group based in London is conducted by means of lectures, publications, and exhibitions. In the 1930s exhibits traveled to five American centres, as well as to Paris and Copenhagen. Members representing the several crafts have collaborated in the making of the rolls of honour for the Royal Air Force and United States Air Force, at St. Clement Danes Church, London. The society has



Manuscript copy sheet by Edward Johnston, 1918. In the Newberry Library, Chicago.
By courtesy of the Newberry Library, Chicago, the Wing Collection

served as a model for similar activities abroad, notably the organization in The Netherlands under the leadership of Jan van Krimpen. In 1952 the British society, with Alfred Fairbank as president, recognized the rising popular interest in italic handwriting by instituting the Society for Italic Handwriting, which soon attracted a large international membership of teachers and amateurs.

In the United States individual enthusiasts and informal groups fostered calligraphy outside the art schools. Ernst F. Detterer in Chicago, who had lessons from Johnston in 1913, headed such a group at the Newberry Library for many years and, when he died in 1947, was succeeded by James Hayes. In Portland, Ore., the instruction and copybooks of Lloyd J. Reynolds, a professor of art in Reed College, had significant influence. For decades Paul Standard in New York practiced and preached *cancellaresca corsiva* and saw the italic reform gain.

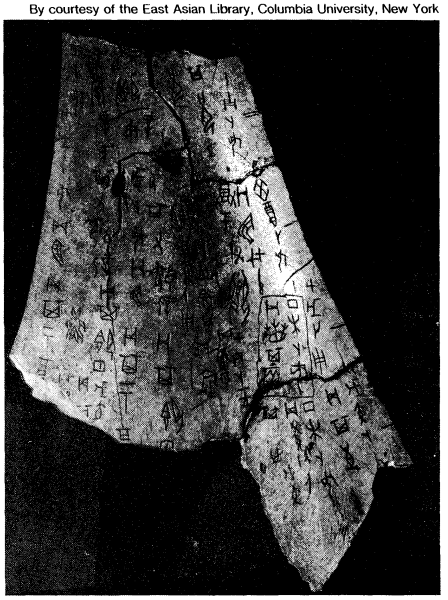
20th-century writing manuals and copybooks

Since Johnston a series of manuals and copybooks centring attention on handwriting improvement have proceeded from the espousal of the rather heavy humanistic hand he admired to an italic that at least implies more speed. The fullest and most practical work on the italic is by Alfred John Fairbank, *A Handwriting Manual* (1932). The author places before his book as frontispiece a page of the *bastarda* of Lucas, *Arte de escrivir,* 1577. J.H. Benson's *The First Writing Book* (1954) consists of the text and examples of the Arrighi *Operina* translated and admirably written out by the editor and furnished with practical clarifications and notes. Since Johnston calligraphical research and publishing activities have also produced a handsome and scholarly shelfful of books on related forms, most notably the Renaissance capitals. (Ra.N./Ed.)

EAST ASIAN CALLIGRAPHY

In China, Korea, and Japan, calligraphy is a form of pure art. Chinese, Korean, and Japanese calligraphy derive from the written form of the Chinese language. Chinese is not an alphabetical language; each character is composed of a number of differently shaped lines within an imaginary square. The early Chinese written words, like the Egyptian hieroglyphs, were pictorial images, though not so close to the objects they represented as in the ancient Egyptian writing. Rather, they were simplified images, indicating meaning through suggestion or imagination. These simple images were flexible in composition, capable of developing with changing conditions by means of slight variations.

By courtesy of the East Asian Library, Columbia University, New York



Chia-ku-wen (Chinese shell-and-bone script), Shang dynasty (18th–12th century BC). In the East Asian Library, Columbia University, New York.

Chinese calligraphy. The earliest known Chinese logographs are engraved on the shoulder bones of large animals and on tortoise shells. The piece illustrated contains a number of the early ideographs; each seems to have been carefully composed before being engraved on the bone. Although they are not entirely uniform in size, the variations are not great. The figures must have evolved from rough and careless scratches in the still more distant past. This *chia-ku-wen,* or shell-and-bone script (18th–12th century BC), is the earliest stage of development in Chinese calligraphy.

Shell-and-bone script

It was said that Ts'ang Chieh, the legendary inventor of Chinese writing, got his ideas from observing animals' footprints and birds' claw marks on the sand as well as other natural phenomena. He then started to work out simple images from what he conceived as representing different objects such as those that are given below:

Surely, the first images that the inventor drew of these few objects could not have been quite so stylized but must have undergone some modifications to reach the above stage. Each image is composed of a minimum number of lines and yet is easily recognizable. Nouns no doubt came first. Later, new ideographs had to be invented to record actions, feelings, and differences in size, colour, taste, and so forth. Something was added to the already existing ideograph to give it a new meaning. The ideograph for deer, for instance, is , not a realistic image but a very much simplified structure of lines suggesting a deer by its horns, big eye, and small body, which distinguish it from other animals. When two such simple images are put side by side, the meaning is "pretty," "prettiness," "beautiful," "beauty," etc., which is obvious if one has seen two such elegant creatures walking together. But, if a third image is added above the other two, as , it means "rough," "coarse," and even "haughty." This interesting point is the change in meaning through the arrangement of the images. If the three stags were not standing in an orderly manner, they could become rough and aggressive to anyone approaching them. From the aesthetic point of view, three such images could not be arranged side by side within an imaginary square without cramping one another, and in the end none would look like a deer at all.

After the shell-and-bone script came writing on bronze vessels, known as bronze script. In the early days of divination, when the kings of the Shang dynasty (18th–12th century BC) tried to solve their problems by consulting their ancestors and deities, the latter's answers were engraved on bones and on tortoise shells for perpetual preservation. Later, bronze was used to make cooking utensils and wine vessels for the special ceremonies of ancestral worship, raw or cooked food being offered up in them. So sacred were

Bronze script

these ancestor-worshipping ceremonies that the best types of bronze utensil were specially designed and cast for such purposes, and, in addition, inscriptions, from a few words up to several hundred, were incised inside the bronzes. The words of the engravings could not be roughly formed or even just simple images; they had to be well worked out to go with the decorative ornaments outside the bronzes, and in some instances they almost became the chief decorative design in themselves. Though they preserved the general structure of the bone-and-shell script, they were considerably elaborated and beautified. Each bronze or set of them may bear a different type of inscription, not only in the wording but also in the manner of writing. Hundreds of them were created by different artists. The bronze script represents another stage of development in Chinese calligraphy, known as *chin-wen* ("metal-script"), *ku-wen* ("ancient-script"), or *ta chuan* ("large-seal") style of writing.

Before long a unification of all types of the bronze script was enforced when China was united for the first time, in the 3rd century BC. The first emperor of Ch'in, Shih Huang-ti, gave the task of working out the new script to his prime minister, Li Ssu, and no other type was allowed to be used. Here are some common words that can be compared with similar words in bone-and-shell script mentioned above:

sun moon mountain water rain wood dog cattle horse

Hsiao chuan ("small-seal" style). Rubbing of a copy of Li Ssu's writing engraved on a stone epitaph, Ch'in dynasty (221–206 BC).
By courtesy of Chiang Yee

This was the third step forward in the development of Chinese calligraphy, known as *hsiao chuan* ("small-seal") style.

The *hsiao chuan,* or small-seal style

In the small-seal style of writing, all lines are drawn of even thickness, and more curves and circles are employed. Each word tends to fill up an imaginary square, and a passage written in small-seal style has the appearance of a series of equal squares neatly arranged in columns and rows, each of them balanced and well-spaced.

The uniformity of writing in China was established chiefly for the purpose of meeting the growing demands for documented records. Unfortunately, the small-seal style could not be written speedily and was therefore not entirely suitable. Another stage of development was needed—the fourth stage, which is called *li shu,* or official style. The Chinese word *li* here means "a petty official" or "a clerk"; *li shu* means a style specially devised for the use of clerks. If examined carefully, *li shu* is found to contain no circles and very few curved lines. Squares and short straight lines, vertical and horizontal, are chiefly used. Because of the speed needed for writing, the brush in the hand tends to move up and down, and an even thickness of line cannot be enforced. As the thickness varied, artist-writers could concentrate more on the artistic shaping of the lines.

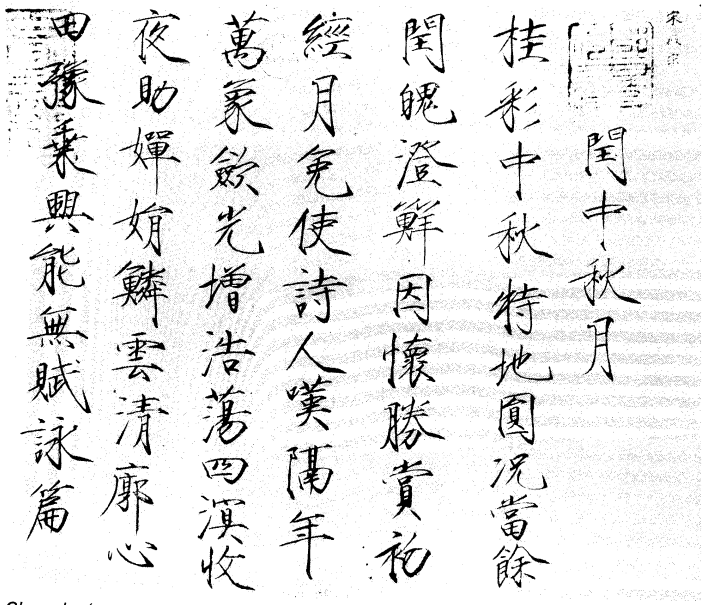
Li shu is thought to have been invented by Ch'eng Miao (240–207 BC), who had offended the First Emperor of Ch'in and was serving a 10-year sentence in prison. He spent his time in prison working out this new development, which opened up seemingly endless possibilities for later calligraphers. According to their own artistic insight, they evolved new variations in the shape of lines and in construction. The words in *li shu* style tend to be square or slightly rectangular horizontally. Though the even thickness of lines is relaxed, the rigidity in the shaping of them is still there; for instance, the vertical lines had to be shorter, and the horizontal ones longer. As this curtailed the freedom of hand for individual artistic taste, another stage of development came into being—the fifth stage, *chen shu* (*k'ai shu*), or regular style. There is no record of who invented this style, but it must have been in evolution for a long time, at least since the 1st century AD if not earlier. The Chinese still use this regular style of writing today; in fact, what is known as modern Chinese writing is almost 2,000 years old, and the written words of China have not changed since the first century of the Christian Era.

Development of the *chen shu,* or regular style

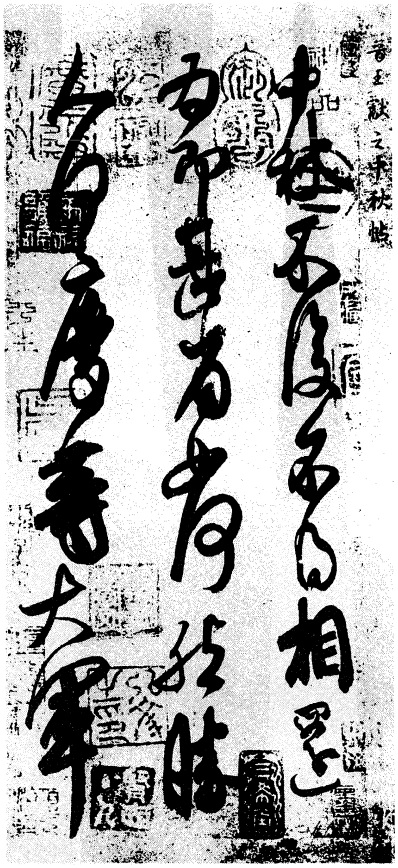
"Regular style" means "the proper style of Chinese writing" used by all Chinese for government documents, printed books, and public and private dealings in important matters ever since its establishment. Since the regulations for the civil service examination enforced in the T'ang period (AD 618–907), each candidate had to be able to write a good hand in regular style. This Imperial decree deeply influenced all Chinese who wanted to become scholars and enter the civil service. This examination was abolished in 1905, but most Chinese still try to acquire a hand in regular style even to the present day.

In *chen shu* each line, each square or angle, and even each dot can be shaped according to the will and taste of the calligrapher. Indeed, a Chinese word in regular style presents an almost infinite variety of problems of structure and composition, and, when executed, it presents to

By courtesy of the National Palace Museum, Taipei



Chen shu (regular style) written by Emperor Hui Tsung (1100–1125/26), Sung dynasty. In the National Palace Museum, Taipei.



Hsing shu (running style) and *ts'ao shu* (grass style) by Wang Hsien-chih (AD 344–386), Six Dynasties period. In the National Palace Museum, Taipei.

By courtesy of the National Palace Museum, Taipei

the onlooker a design whose abstract beauty can draw the mind away from the literal meaning of the word itself.

The greatest exponents of Chinese calligraphy were Wang Hsi-chih (died 379) and his son Wang Hsien-chih in the 4th century. Few of their original works have survived, but a number of their writings were engraved on stone tablets, and rubbings were made from them. Many great calligraphers imitated their styles, but none ever surpassed them.

Wang Hsi-chih not only provided the greatest example in the regular style of writing but also relaxed the tension somewhat in the arrangement of the strokes in the regular style by giving easy movement to the brush to trail from one word to another. This is called *hsing shu,* or running style, as if the hand were walking fast while writing. This, in turn, led to the creation of *ts'ao shu,* or grass style, which takes its name from its appearance—as if the wind had blown over the grass in a manner disorderly yet orderly. The English term cursive writing cannot describe the Chinese grass style, for even a cursive hand can be deciphered without very much difficulty. But Chinese words in grass style are greatly simplified forms of the regular style and can be deciphered only by those who have practiced calligraphy for years. It is not a style for general use but for the calligrapher who wishes to produce a work of abstract art.

The *hsing shu,* or running style, and the *ts'ao shu,* or grass style

Technically speaking, there is no mystery in Chinese calligraphy. The tools for Chinese calligraphy are very few—good ink, ink stone, a good brush, and good paper (some prefer silk). It depends on the skill and imagination of the writer to give interesting shapes to his strokes and to compose beautiful structures from them without any retouching or shading and, most important of all, with well-balanced spaces between the strokes. This balance needs years of practice and training.

The fundamental inspiration of Chinese calligraphy, as of all arts in China, is nature. In regular style each stroke, even each dot, suggests the form of a natural object. As every twig of a living tree is alive, so every tiny stroke of a piece of fine calligraphy has the energy of a living thing. Printing does not admit the slightest variation in the shapes and structures, but strict regularity is not tolerated by Chinese calligraphers, especially those who are masters of the *ts'ao shu.* A finished piece of fine calligraphy is not a symmetrical arrangement of conventional shape but, rather, something like the coordinated movements of a skillfully composed dance—impulse, momentum, momentary poise, and the interplay of active forces combining to form a balanced whole. (C.Y./Ed.)

Korean calligraphy. Koreans have used Chinese characters probably since the 2nd or 3rd century AD. Even after the invention of the Korean alphabet in 1447, Chinese was used as the official script until the 19th century.

A few inscribed stone monuments remain from the Three Kingdoms period (*c.* 57 BC–AD 668). Ancient Koreans, eager to adopt Chinese culture, developed a calligraphy reflecting Chinese styles. In the following Unified Silla dynasty (668–935), a devotion and adherence to the T'ang culture of China gave birth to such great masters of calligraphy in Korea as Kim Saing and Choi Ch'i-wŏn, whose styles of writing basically followed those of the Chinese calligraphers Ou-yang Hsün and Yü Shih-nan.

The angular, squarish style of Ou-yang Hsün, Yü Shih-nan, and Yen Chen-ch'ing, inherited from the Silla dynasty, continued in the Koryŏ period (918–1392) until around 1350, when the rounded, fluent style of the Chinese calligrapher Chao Meng-fu, of the Yüan dynasty, was introduced and became the vogue. Since that time the *chao* style has remained the basic undercurrent in Korean calligraphy.

The *chao* style

At first the calligraphy of the Yi dynasty (1392–1910) followed the *chao* style, but early in the 16th century a mannered, vulgar style began to be evident. The 19th century saw, however, the emergence of individual styles related to those of Chinese calligraphers. The new trend was the result of Korea's close cultural contacts with Ch'ing China.

The greatest master of the Yi period was Kim Chŏng Hi, who established the so-called *ch'usa* style. His calligraphy is derived from the *li shu* script of China, but his sense of pictorial composition, harmony within asymmetry, and animation by unmatched, forceful strokes gave him a style completely his own.

The influence of Japanese calligraphy began to be felt about 1920. Since World War II, calligraphy in both North and South Korea has been profoundly influenced by governmental decisions to replace all Chinese characters with words written in the native alphabet. As a consequence, modern Korean calligraphy has been developing along new lines. (W.-Y.K./Ed.)

Japanese calligraphy. The art of calligraphy has long been highly esteemed in Japan as in China. There is no definite record of when the Japanese began to use Chinese words—called *kanji* in Japanese. It is known that a Korean scribe named Wani brought some Chinese books of Confucian classics, such as the *Analects, Great Learning,* and *Book of Mencius,* to Japan near the end of the 4th century AD. From the 7th century onward, many Japanese scholars, particularly Buddhist monks, went to China, and some Chinese went to Japan. As Indian Buddhism reached Japan via Korea and China and took root there, the use of *kanji* in Japan gradually grew. Eventually, *kanji* became the official system of writing in Japan.

Kanji script

Most of the Chinese Buddhist monks who went to live in Japan were scholars and good calligraphers; their writings on the Buddhist scriptures and other subjects were admired and esteemed not only for their aesthetic value as calligraphy but also because they induced a sense of religious awe in the readers.

Many of the early Japanese emperors were ardent Buddhists and also acquired a masterly hand in *kanji* writings. So did many Japanese Zen priests, whose calligraphy tended to exercise a religious effect upon the Japanese mind. Theirs became a special type of calligraphy in Japan, namely, Japanese Zen calligraphy, or *bokuseki.*

Naturally, it was unsuitable for Japan to adopt an entire foreign language like Chinese, and Japanese thinkers

Ch'usa style written by Kim Chong Hi (1786–1856), Yi period.

began to devise a new, native script known as *hiragana,* which was often referred to as "women's hand," or *onnade* in Japanese. It was used particularly in the writing of Japanese poetry and had an elegant and graceful appearance.

Hiragana calligraphy

There are many outstanding pieces of Japanese calligraphy in *kanji,* but they are not distinctive when compared with their Chinese counterparts. Japanese *hiragana* calligraphy, however, stands out prominently and proudly, especially in the style of *remmen-tai,* in which the *hiragana* are written continuously and connected together without break, and in *chōwa-tai,* in which some *kanji* words join hands with the *hiragana.* Japanese calligraphy in *remmen-tai* or in *chōwa-tai* has some resemblance to the Chinese grass style, but the two are easily distinguishable. In Chinese grass style, although the words are greatly simplified and several words can be joined together with trailing strokes, each separate word normally still retains its regular spacing within an imaginary square, big or small. But Japanese *hiragana* cannot be spaced so separately and evenly. Therefore, a whole piece of *remmen-tai* calligraphy looks like a big bundle of beautiful silk strings hanging down confusedly yet artistically, as if the calligrapher had let his hand move swiftly of its own accord. The separate strokes and dots have no distinctive shape but join other strokes and dots in the following *hiragana.* The strokes or lines in *hiragana* are not shaped like living things, nor are they of even thickness; but there must be good spacing between the strokes or lines and between one *hiragana* and another, so that there is no confusion or blur in the completed piece. This is a highly demanding art, and the whole piece has to be executed with speed and without hesitation. *Hiragana* requires solid training and artistic insight. (C.Y./Ed.)

BIBLIOGRAPHY

General studies: GEOFFREY SAMPSON, *Writing Systems: A Linguistic Introduction* (1985), is the first systematic account of the relations between various writing systems and the linguistic structures they represent, with an especially good section on Chinese, Japanese, and Korean. ROY HARRIS, *The Origin of Writing* (1986); ALBERTINE GAUR, *A History of Writing* (1984); IGNACE J. GELB, *A Study of Writing,* rev. ed. (1963, reprinted 1969); DAVID DIRINGER, *The Alphabet: A Key to the History of Mankind,* 3rd ed. rev., 2 vol. (1968); and HANS JENSEN, *Sign, Symbol, and Script: An Account of Man's Efforts to Write,* 3rd rev. and enl. ed. (1968; originally published in German, 1935), are standard introductions to writing and the alphabet. ERIC A. HAVELOCK, *The Literate Revolution in Greece and Its Cultural Consequences* (1982); and JACK GOODY (ed.), *Literacy in Traditional Societies* (1968, reprinted 1975), examine the cultural, historical, and psychological dimensions of alphabetic writing systems. See also HARVEY J. GRAFF (ed.), *Literacy and Social Development in the West: A Reader* (1981). JACQUES DERRIDA, *Of Grammatology* (1976; originally published in French, 1967), examines the conceptual implications of writing. BRIAN STOCK, *The Implications of Literacy: Written Language and Models of Interpretation in the Eleventh and Twelfth Centuries* (1983), sets out the increasing uses of writing in the Middle Ages that presaged the Reformation. Good introductions to writing and its uses are provided by ROBERT PATTISON, *On Literacy: The Politics of the Word from Homer to the Age of Rock* (1982); DONALD JACKSON, *The Story of Writing* (1981); and MARCEL COHEN, *La Grande Invention de l'écriture et son évolution,* 2 vol. (1958). INSUP TAYLOR and M. MARTIN TAYLOR, *The Psychology of Reading* (1983), discusses the relation between various writing systems and the psychological processes involved in learning to read and write them. DAVID R. OLSON, NANCY TORRANCE, and ANGELA HILDYARD (eds.), *Literacy, Language, and Learning: The Nature and Consequences of Reading and Writing* (1985), gives a sampling of the interdisciplinary nature of recent work on literacy. See also the section on writing and orthography in *Linguistic Bibliography* (annual).

Reproductions of writings: FRIEDRICH BALLHORN, *Grammatography: A Manual of Reference to the Alphabets of Ancient and Modern Languages* (1861; reprinted with title *Alphabets of the World,* 1983); CHARLES FOSSEY (ed.), *Notices sur les caractères étrangers, anciens et modernes, rédigées par un groupe de savants,* new ed. (1948); *The Gospel in Many Tongues: Specimens of 872 Languages in Which the British and Foreign Bible Society Has Published or Circulated Some Portion of the Bible,* new ed. (1965); RICHARD LEPSIUS, *Standard Alphabet for Reducing Unwritten Languages and Foreign Graphic Systems to a Uniform Orthography in European Letters,* 2nd ed. (1863, reprinted 1981; originally published in German, 1855); and AKIRA NAKANISHI, *Writing Systems of the World: Alphabets, Syllabaries, Pictograms* (1982).

Subgraphemics or semasiography: GARRICK MALLERY, *Pictographs of the North American Indians* (1886), and *Picture-Writing of the American Indians* (1893, reprinted 1972); and HENRY R. SCHOOLCRAFT, *Historical and Statistical Information Respecting the History, Condition, and Prospects of the Indian Tribes of the United States,* vol. 1 (1851, reissued 1975).

Graphemics: C.F. VOEGELIN and F.M. VOEGELIN, "Typological Classification of Systems with Included, Excluded, and Self-Sufficient Alphabets," *Anthropological Linguistics,* 3(1):55–96 (1961); and I.J. GELB, "A Note on Morphographemics," in DAVID COHEN (ed.), *Mélanges Marcel Cohen,* pp. 73–77 (1970).

Alphabets: Almost all of the following works contain excellent, copious, and classified bibliographies: ARTHUR B. ALLEN, *The Romance of the Alphabet* (1937); DAVID DIRINGER, *Writing* (1962); ERNST DOBLHOFER, *Voices in Stone: The Decipherment of Ancient Scripts and Writings* (1961, reprinted 1973; originally published in German, 1957); WALTER DURFEE, *Alphabetics as a Science* (1956); CHARLES HIGOUNET, *L'Écriture* (1955, reprinted 1982); ALEXANDER HUMEZ and NICHOLAS HUMEZ, *A B C Et Cetera: The Life & Times of the Roman Alphabet* (1985); KEITH GORDON IRWIN, *The Romance of Writing, from Egyptian Hieroglyphics to Modern Letters, Numbers, and Signs* (U.K. title, *Man Learns to Write,* 1956); WILHELM H. LANGE, *Schriftbibel: Geschichte der abendländischen Schrift von den Afnangen bis zur Gegenwart,* 3rd rev. and enl. ed. (1951); SAMUEL A.B. MERCER, *The Origin of Writing and Our Alphabet* (1959); A.C. MOORHOUSE, *The Triumph of the Alphabet: A History of Writing* (1953); OSCAR OGG, *The 26 Letters,* rev. ed. (1971, reissued 1983); ALFRED PETRAU, *Schrift und Schriften im Leben der Völker: Ein Kulturgeschlichter Beitrag,* 2nd ed. (1944); PETER RUDLAND, *From Scribble to Script* (1955); ALBERT SCHMITT, *Untersuchungen zur Geschichte der Schrift: Eine Schriftentwicklung um 1900 in Alaska,* 2 vol. (1940, reissued with revisions, 1981); KURT SETHE, *Vom Bilde zum Buchstaben: Die Entstehungsgeschichte der Schrift* (1939, reissued 1964); MARTIN SPRENGLING, *The Alphabet: Its Rise and Development from the Sinai Inscriptions* (1931); and TOMMY THOMPSON, *The ABC of Our Alphabet* (1942). See also ALBERT KAPR, *The Art of Lettering: The History, Anatomy, and Aesthetics of the Roman Letter Forms* (1983; originally published in German, 1971).

Hieroglyphs: Books in English on hieroglyphic writing include ALAN H. GARDINER, *Egyptian Grammar,* 3rd ed., rev. (1957, reprinted 1978), with a list of hieroglyphs from the Middle Kingdom; F.LL. GRIFFITH, *A Collection of Hieroglyphs: A Contribution to the History of Egyptian Writing* (1898); and NINA M. DAVIES, *Picture Writing in Ancient Egypt* (1958), a collection of artistically valuable hieroglyphs.

Punctuation: G.V. CAREY, *Mind the Stop,* rev. ed. (1958, reissued 1976); and E.H. PARTRIDGE, *You Have a Point There: A Guide to Punctuation and Its Allies* (1953, reissued 1978), are the best guides to modern punctuation as practiced in Britain; the latter is more exhaustive and includes a chapter on American practice by JOHN W. CLARK. A comparable American work is the chapter on punctuation in WILMA R. EBBITT and DAVID R. EBBITT, *Writer's Guide and Index to English,* 6th ed. (1978). The following describe the practices of two famous presses, Oxford University Press and the University of Chicago Press: HORACE HART, *Hart's Rules for Compositors and Readers at the University Press, Oxford,* 39th ed. rev. (1983); and *A Manual of Style: For Authors, Editors and Copywriters,* 13th ed. rev. (1982). For punctuation in antiquity and the Middle Ages, see FRANZ STEFFENS, *Lateinische Paläographie,* 2nd ed. (1907, reissued 1964); and PETER CLEMOES, *Liturgical Influence on Punctuation in Late Old English and Early Middle English Manuscripts* (1952). For punctuation in and since the Renaissance, especially in Britain, see the relevant sections in A.C. PARTRIDGE, *Orthography in Shakespeare and Elizabethan Drama: A Study of Colloquial Contractions, Elision, Prosody, and Punctuation* (1964); and T.F. HUSBAND and M.F.A. HUSBAND, *Punctuation: Its Principles and Practice* (1905). ALEXANDER BIELING, *Das Princip der deutschen Interpunktion: Nebst einer übersichtlichen Darstellung ihrer Geschichte* (1880), is useful for German and for European punctuation in general since the 15th century.

Shorthand: EDWARD HARRY BUTLER, *The Story of British Shorthand* (1951), a complete history of early shorthand systems with emphasis on modern British systems; HANS GLATTE, *Shorthand Systems of the World: A Concise Historical and Technical Review* (1959), a short history that emphasizes the extent to which important systems spread to countries other than those in which they originated; LOUIS A. LESLIE (ed.), *The Story of Gregg Shorthand: Based on the Writings of John Robert Gregg* (1964), on the early life of John Robert Gregg, his invention of Gregg shorthand, his promotion of the system in the United States, and his method of teaching; ISAAC PITMAN, *A History of Shorthand,* 4th ed. (1918), a chronological history of shorthand systems with emphasis on British systems to the end of the 19th century; and ALLIEN R. RUSSON, *Methods of Teaching Shorthand* (1968), the history of shorthand instruction.

Aramaic and Hebrew calligraphy: G.R. DRIVER, *Semitic Writing from Pictograph to Alphabet,* rev. ed. edited by S.A. HOPKINS (1976), on the origins of Semitic writing; G.R. DRIVER (ed. and trans.), *Aramaic Documents of the Fifth Century B.C.* (1954, reprinted 1968); EUGÈNE TISSERANT (comp.), *Specimina Codicum Orientalium* (1914), reproductions of Semitic pen hands; CARLO BERNHEIMER, *Paleografia Ebraica* (1924), many Ashkenazic hands; MOSES GASTER, *Hebrew Illuminated Bibles of the IXth and Xth Centuries* . . . (1901); REUBEN LEAF, *Hebrew Alphabets, 400 B.C. to Our Days* (1950, reissued 1987), reproductions of manuscript styles; *The Book of Jonah,* woodcuts by JACOB STEINHARDT, calligraphy by FRANZISCA BARUCH (1953); and HENRI FRIEDLAENDER, *Die Entstehung meiner Hadassah-Hebräisch* (1967), on the relationship of Hebrew manuscript styles and types.

Arabic calligraphy: ANNEMARIE SCHIMMEL, *Islamic Calligraphy* (1970), a stimulating introduction with illustrations, including calligraphy in architecture and the decorative arts, and a useful bibliography, and *Calligraphy and Islamic Culture* (1984); YASIN HAMID SAFADI, *Islamic Calligraphy* (1978), on the development of Arabic writing from pre-Islāmic times through the 20th century; NABIA ABBOTT, *The Rise of the North Arabic Script and Its Kur'ānic Development* (1939), a study of the origins of the Arabic script and its development in the early Islāmic period; B. MORITZ (ed.), *Arabic Palaeography: A Collection of Arabic Texts from the First Century of the Hidjra till the Year 1000* (1905, reprinted 1986), a rich collection of texts on papyrus and paper; ANTHONY WELCH, *Calligraphy in the Arts of the Muslim World* (1979), essays on aspects of calligraphy in Islām; ARTHUR UPHAM POPE (ed.), *A Survey of Persian Art,* new ed. (1964), especially vol. 4 on Arabic calligraphy in the art and architecture of Persia; BASIL GRAY (ed.), *The Arts of the Book in Central Asia, 14th–16th Centuries* (1979), a work unique in both coverage and scholarship; and V. MINORSKY (trans.), *Calligraphers and Painters: A Treatise by Qādī Aḥmad* . . . (1959), an important work (written in 1606) that illustrates the Islāmic attitude toward calligraphy.

Western calligraphy: DOROTHY E. MINER, VICTOR I. CARLSON, and P.W. FILBY (comps.), *2,000 Years of Calligraphy* (1965, reissued 1980), a comprehensive and well-illustrated catalog of an exhibition devoted to regions using the Latin alphabet, with especially valuable notes and references from the 1st to the 19th century; EDWARD JOHNSTON, *Writing & Illuminating, & Lettering,* 14th ed. (1906, reprinted 1980), the gospel of the modern revival by its chief apostle, and *Formal Penmanship, and Other Papers,* edited by HEATHER CHILD (1971, reissued 1980); DAVID DIRINGER, *The Hand-Produced Book* (1953, reprinted with title, *The Book Before Printing: Ancient, Medieval, and Oriental,* 1982), a storehouse of information; JAN TSCHICHOLD, *An Illustrated History of Writing and Lettering* (1946, reissued 1948; originally published in German, 1940), a brief, perceptive, and personal account by an eminent designer; HERMANN DEGERING (ed.), *Lettering: A Series of 240 Plates Illustrating Modes of Writing in Western Europe from Antiquity to the End of the 18th Century,* 3rd ed. (1954, reprinted 1965; originally published in German, 3rd ed. 1952), a standard survey of scripts; and B.L. ULLMAN, *The Origin and Development of Humanistic Script* (1960, reprinted 1974). See also JOYCE IRENE WHALLEY, *The Pen's Excellencie: Calligraphy of Western Europe and America* (1980, reprinted 1982); MARC DROGIN, *Medieval Calligraphy, Its History and Technique* (1980); JAMES WARDROP, *The Script of Humanism: Some Aspects of Humanistic Script, 1460–1560* (1963); ALFRED FAIRBANK and BERTHOLD WOLPE, *Renaissance Handwriting: An Anthology of Italic Scripts* (1960); OSCAR OGG (ed.), *Three Classics of Italian Calligraphy: An Unabridged Reissue of the Writing Books of Arrighi, Tagliente, Palatino* (1953); AMBROSE HEAL, *The English Writing-Masters and Their Copy-Books, 1570–1800* (1931, reprinted 1962), the fundamental biographical and bibliographical work, well illustrated and with an important essay by STANLEY MORISON; RAY NASH, *American Writing Masters and Copybooks: History and Bibliography Through Colonial Times* (1959), and *American Penmanship, 1800–1850: A History of Writing and a Bibliography of Copybooks from Jenkins to Spencer* (1969), with miniature reproductions of title pages.

Greek calligraphy: E.G. TURNER, *Greek Manuscripts of the Ancient World,* 2nd ed. (1987), the best general work and also very well illustrated; BRUCE M. METZGER, *Manuscripts of the Greek Bible: An Introduction to Greek Palaeography* (1981); FRANCHI DE'CAVALIERI and JOHANNES LIETZMANN, *Specimina Codicum Graecorum Vaticanorum* (1910, reissued 1929), 50 Greek manuscript specimens in the Biblioteca Apostolica Vaticana, Vatican City; *The Codex Alexandrinus,* 5 vol. (1909–57), a facsimile in reduced size, with introductions by F.G. KENYON, H.J.M. MILNE, and T.C. SKEAT; *The Codex Sinaiticus and the Codex Alexandrinus,* prepared by H.J.M. MILNE and T.C. SKEAT (1951, reissued 1963), on the origins of the two great Greek uncial Bibles in the British Museum; *Facsimile of the Washington Manuscript of Deuteronomy and Joshua in the Freer Collection* (1910); and *Ilias Ambrosiana* (1953), a beautiful facsimile of the Homeric codex in the Biblioteca Ambrosiana, Milan (elegant uncial writing).

Latin calligraphy: The best short accounts in English are in B.L. ULLMAN, *Ancient Writing and Its Influence,* new ed. (1969, reprinted 1980); and EDWARD MAUNDE THOMPSON, *An Introduction to Greek and Latin Palaeography* (1912, reprinted 1973). All Roman books and documents are cataloged (in English) in, respectively, E.A. LOWE (ed.), *Codices latini antiquiores: A Palaeographical Guide to Latin Manuscripts Prior to the Ninth Century,* 12 vol. (1934–66, reprinted 1982); and ALBERT BRUCKNER and ROBERT MARICHAL (eds.), *Chartae latinae antiquiores: Facsimile Edition of the Latin Charters Prior to the Ninth Century* (1954–), 24 vol. of which had appeared to 1986. The most important monographs are JEAN MALLON, *Paléographie romaine* (1952); and essays in E.A. LOWE, *Palaeographical Papers 1907–1965,* 2 vol. (1972). LEONARD E. BOYLE, *Medieval Latin Palaeography: A Bibliographical Introduction* (1984), is a guide to further study of Latin scripts.

Chinese calligraphy: YU-HO ECKE TSENG, *Chinese Calligraphy* (1971); SHEN FU, *Traces of the Brush: Studies in Chinese Calligraphy* (1977, reissued 1980); LUCY DRISCOLL and KENJI TODA, *Chinese Calligraphy,* 2nd ed. (1964); YEE CHIANG, *Chinese Calligraphy: An Introduction to Its Aesthetic and Techniques,* 3rd ed. rev. and enl. (1973); WILLIAM WILLETTS, *Chinese Calligraphy: Its History and Aesthetic Motivation* (1981), and *Chinese Art,* 2 vol. in 1 (1958); CHIH-MAI CH'ÊN, *Chinese Calligraphers and Their Art* (1966); and SHEN FU, GLENN D. LOWRY, and ANN YONEMURA, *From Concept to Context: Approaches to Asian and Islamic Calligraphy* (1986).

Korean calligraphy: EUNG-HYON KIM, "Sang-ko eui Sŏye" (Ancient Korean Calligraphy), "Koryŏ eui Sŏye" (Calligraphy of the Koryŏ Dynasty), "Yi-cho eui Sŏye" (Calligraphy of the Yi Dynasty), and "Hyondae eui Soye" (Modern Calligraphy), all in *Han'guk yesul ch'ongnam,* edited by the Academy of Art, Seoul, Korea (1965), the best surveys on Korean calligraphy, by a noted calligrapher; and KI-SUNG KIM, *Han'guk Sŏye sa* (1966), a general survey of Korean calligraphy.

Japanese calligraphy: YOSHIAKI SHIMIZU and JOHN M. ROSENFIELD, *Masters of Japanese Calligraphy 8th–19th Century* (1984); YUJIRO NAKATA, *The Art of Japanese Calligraphy* (1973; originally published in Japanese, 1967); and HISAO SUGAHARA, *Japanese Ink Painting and Calligraphy* . . . , trans. from Japanese (1967).

(D.R.O./D.D./H.B./T.J.Br./A.R.R./E.G.T./R.Ba./Ra.N./D.An./R.H.P.-W./C.Y./W.Y.K.)

Yiddish Literature

Yiddish literature may be said to have been born twice. The earliest evidence of Yiddish literary activity dates from the 13th century and is found in southern Germany, where the language itself had originated as a specifically Jewish variant of Middle High German approximately a quarter of a millennium earlier. The Haskalah, the Jewish equivalent of the Enlightenment, effectively doomed the Yiddish language and its literary culture in Germany and in western Europe during the course of the 18th century. At the beginning of the 19th century, however, the Haskalah paradoxically promoted a renascence of Yiddish literature in those parts of eastern Europe to which the Yiddish language had been carried from the 13th century onward. The Haskalah therefore represents a watershed separating two essentially distinct cultural phenomena. Whereas in eastern Europe in the 19th and 20th centuries Yiddish literature eventually became an autonomous modern mode of literary expression fully comparable with parallel European literatures, the Yiddish literature of medieval Germany and the adjacent territories to which it spread remained in the shadow of the infinitely more prestigious Hebrew literature and, in theory at least, was addressed only to women (who were not taught Hebrew but learned to read and write Yiddish in the community schools) and to untutored men.

The low regard in which the Yiddish language and its culture were held conspired with the ravages of time and the turbulent vicissitudes of Jewish history in the German-speaking lands in such a way that all but a small proportion of medieval Yiddish texts have been lost, though important texts are still coming to light. From what survives, however, it is clear that the literature aimed predominantly at the edification of its readership and was either didactic in character or sought to entertain while serving as an antidote to the supposed moral dangers of Gentile literature. As a result of their traditional respect for learning, the efficacy of their community schools, and their predominantly urban way of life, literacy was much higher among the Jews of northern Europe during this time than among non-Jews living in the same area. Many works achieved such popularity that they were frequently reprinted over a period of centuries and enjoyed an astonishingly wide dissemination, with the result that their language developed into an increasingly ossified koine that was readily understood over a territory extending from Amsterdam to Odessa and from Venice to Hamburg. During the 18th century the picture changed rapidly in western Europe, where increasing cultural assimilation led to the abandonment of Yiddish in favour of the languages of the ambient societies. In eastern Europe, on the other hand, the Haskalah, as a result of the recognition that its mission to enlighten the Jewish masses could only be accomplished through the medium of Yiddish, unintentionally wrought a renewal of the language it disparaged. The resurgence of Yiddish literature in eastern Europe went hand in hand with the emergence of a new standard literary language based on the eastern dialects, which had been invigorated by contact with the languages of its Slavic environment.

This article provides a historical survey of the development of Yiddish literature. For a discussion of literature in Hebrew, see the article HEBREW LITERATURE.

For coverage of related topics in the *Macropædia* and *Micropædia,* see the *Propædia,* section 621.

The article is divided into the following sections:

The Middle Ages to the 18th century in western Europe 1077
  Biblical and religious literature 1077
  Secular literature 1078
Modern literature in eastern Europe and in emigration 1079
  The 19th century 1079
  The 20th century 1079
    To World War I
    After World War I
Bibliography 1082

The Middle Ages to the 18th century in western Europe

BIBLICAL AND RELIGIOUS LITERATURE

The earliest recorded sentence written in Yiddish is a blessing inscribed in a mahzor (prayer book for the Jewish holidays) written in Worms in 1272–73. Other rudimentary signs of Yiddish literary activity from the 13th century, or perhaps earlier, are interlinear and marginal glosses found in manuscripts of books of the Bible and of biblical commentaries. Glossaries were made from the end of the 14th century onward. The earliest known printed Yiddish book is a Hebrew–Yiddish glossary and concordance of the Hebrew Bible published in Kraków in 1534 and known as the *Seyfer shel Reb Anshel* ("Book of Rabbi Anshel"). The oldest known printed Yiddish text is a version of the Passover hymn *Addir Hu* ("Mighty is He"), which appears as *Almekhtiger Got* in a Haggadah (a book of Jewish lore) published in Prague in 1526. Actual translations of portions of the Bible appear from the end of the 15th century onward. The earliest example to which a definite date may be assigned is a 1490 Yiddish manuscript version of the Psalms. Printed Yiddish Pentateuchs (the first five books of the Old Testament) were first published in Constance and Augsburg in 1544, while a *Taytsh khumesh*—a Pentateuch accompanied by the appointed portions of The Prophets (Hebrew: Nevi'im) and the five Scrolls (Megiles; Hebrew: Megillot), together with extracts from the celebrated 11th-century commentator Rashi—appeared in Cremona in 1560.

Earliest printed Yiddish works

Considerable popularity was also attained by rhymed adaptations of parts or the whole of the Bible, many of which were embroidered with material from midrashim (anthologies of rabbinic sayings, parables, and tales arranged as expositions of books of the Bible, chapter by chapter). The earliest extant example is a manuscript dated 1382 that was found in Cairo at the end of the 19th century. It is housed in the Cambridge University Library and is known as the Cambridge Yiddish Codex. It includes the stories of Aaron, the Garden of Eden, the young Abraham, and of Joseph and Potiphar.

Greater originality was evinced by the biblical epics that survive from the 16th century, some of which may have been written earlier. Jewish audiences had long been familiar with more or less lightly expurgated German sagas and romances. The Cambridge Yiddish Codex includes a Hebrew-alphabet version of a section of the Gudrun cycle that is approximately 130 years older than the earliest recorded Middle High German version. There are also Hebrew-alphabet manuscript versions of the chivalric romance *Wigalois,* and evidence suggests that there were once Jewish versions of a number of other Middle High German epics.

Though perhaps these Hebrew-alphabet versions of German material should not strictly speaking be accounted a part of Yiddish literature, they certainly served as models for more authentically Jewish biblical epics. By far the most notable of these is the *Shmuel-bukh* ("Samuel Book"), published in Augsburg in 1544. Its anonymous author rendered the story of the prophet Samuel, of King

Biblical epics

Saul, and above all of King David into the rhymed stanzas of the *Nibelungenlied.* Using material from what at that time was still a single Book of Samuel copiously embellished with legends drawn from the Talmud and from midrashic sources, the author fashioned a dramatic panorama filled with realistically depicted battle scenes and heroic deeds narrated with an element of suspense, together with much broad humour and erotic incident, all centred on the somewhat idealized image of King David. A similar but slightly less successful treatment of the Books of Kings was the *Melokhim-bukh,* published in 1543.

Of all the various forms of Yiddish biblical literature the most popular was the *Tsenerene,* a paraphrase of the Pentateuch with other appointed portions of Scripture enriched with midrashic material. It was written at the end of the 16th century by Jacob ben Isaac Ashkenazi of Janów Lubelski. The work appears to have been published in Lublin some time between 1590 and 1600; it has gone through more than 200 editions and has remained constantly in print. The *Seyfer ha-maged* was almost as popular and took the form of a sequel to the *Tsenerene,* encompassing the Prophets and the Hagiographa in the original Hebrew text together with a paraphrase and Rashi's commentary in Yiddish.

Though the 16th century had seen numerous rhymed and more or less freely adapted versions of parts of the Bible, it was not until 1676–79 that Yekusyel ben Yitskhek Blits produced a translation of the Hebrew Scriptures in Amsterdam. By then there had been a reaction against the type of embroidered homiletic version used in the *Tsenerene,* and, in emulation of Martin Luther, an attempt was made to produce a translation close to the original that, nonetheless, eschewed word-for-word glossing. This approach was even more pronounced in the case of the second complete Yiddish Bible, which was translated by Yoysef Vitsnhoyzn and published in Amsterdam in 1679.

Freely adapted biblical material eventually found its way into drama around the end of the 17th century. Jewish tradition did not originally permit theatrical performances, and only at the time of the feast of Purim was this prohibition relaxed sufficiently to allow extempore enactment of the Megile of Esther, associated with the feast, and subsequently of other biblical stories. The oldest extant manuscript of such a Purim play (Purim-shpiln) is the *Akhashveresh-shpil,* named after the Persian king, Ahasuerus, in the Book of Esther. Other subjects of the early Purim plays that have survived are the story of the sale of Joseph into captivity and the story of David and Goliath.

Purim plays

Collections of prayers either with a Yiddish translation at the foot of the Hebrew text or in Yiddish only, whether for daily use or specially for the holidays, are found from the 16th century onward. The first printed Yiddish prayer book dates from 1544. More interesting, however, are the *tkhines,* which are collections of prayers for the use of women. Whereas the prayers for general use invoke the deity in the name of the community and closely follow the liturgical canon, the *tkhines* articulate the personal supplications of individual women and communicate far more about the joys and particularly the sorrows of everyday Jewish life in 16th-century Germany. This is also the case with the *minhagim,* or custom books, which recorded prevalent ritual usages as they had developed in particular communities. The *minhagim,* together with *musar* books, which prescribe norms of moral conduct, reflect prevailing ideals concerning everyday life. The earliest *musar* books in Yiddish were translations from Hebrew, but at the beginning of the 17th century a number were composed in Yiddish. Among these the *Seyfer brantshpigl* (1602; "The Burning Mirror") by Moyshe Khanekh Altshul was preeminent.

SECULAR LITERATURE

A genre that occupies an intermediate position between religious and secular writings is represented by collections of stories and fables, in which free adaptations of midrashic and talmudic material are found side by side with tales garnered from a variety of Gentile sources. The most notable example is *Eyn sheyn mayse-bukh* ("A Beautiful Story Book"), usually referred to simply as the *Mayse-bukh,* published in Basel in 1602 by Yankev ben Avrom of Międzyrzecz. Many of the stories are designed to illustrate an edifying moral or to provide models of behaviour.

By far the most accomplished Yiddish writer of this period was Elye ben Asher ha-Levi, known as Elijah Levita or Elye Bokher. Though he was born near Nürnberg, he lived and worked primarily in Padua, Venice, and Rome. He was a noted Hebrew grammarian, and his Yiddish translation of the Psalms, published in Venice in 1545, was the first to be printed and achieved great popularity. He also wrote the verse lampoons *Sreyfe fun Venetsye* ("The Fire of Venice") and *ha-Mavdil* ("Benediction"), but he made his reputation chiefly as the author of the Yiddish romances, the *Bove-bukh* and *Pariz un Viene.* The *Bove-bukh,* or *Bove d'Antona* as it was entitled in its first edition (1541), was composed in 1507–08 and takes the form of an abridged and skillful adaptation of one of the many Italian versions of the *Buève de Hantone,* an Anglo-Norman *chanson de geste.* The molding of the story to Jewish taste and the virtuoso performance in adapting the Italian *ottava rima* stanza to the Yiddish language (at much the same time and place as Giuseppe Sarfati was introducing this form into Hebrew poetry) are both evidence of exceptional literary talent. His *Pariz un Viene,* based on the Italian version of an originally Provençal source, was an even greater tour de force but does not seem to have achieved the same popularity and survives in only two incomplete copies printed in Verona in 1594.

Yiddish adaptation of the romance

The troubled social conditions in which the Jews of Germany lived during the 17th and 18th centuries are reflected in a class of verse narratives that served to disseminate news of blood-libels, expulsions, and pogroms. The narratives are similar to the rhymed chronicles or *historische Volkslieder* ("historical folk songs") that circulated as broadsheets and pamphlets in Germany at the same time. Indeed, the Yiddish songs often constituted rebuttals of the anti-Semitic libels in the German *Lieder.* The earliest and best known example is the *Sheyn lid megiles Vints,* by Elkhonen ben Avrom Heln, which tells of anti-Semitic rioting during a popular uprising led by Vincenz Fettmilch against the patrician oligarchy of Frankfurt am Main in 1612–14. The earliest extant copy is that printed in Amsterdam in 1648. Other examples are a *Meshiakh-lid* (1666; "Messiah Song"), about the false Messiah, Shabse Tsvi (Sabbatai Ẓevi); and a *Shvedish lid* ("Swedish Song"), concerning events in the Swedish War in 1648, at a time when Jewish refugees were fleeing into Germany as a result of the Khmelnitsky massacres in the Ukraine.

The most vivid reflection of Jewish everyday life in 17th- and early 18th-century Germany is found in the *Zikhroynes mores Glikl Hamil* (1896; *The Life of Glückel of Hameln*). These memoirs, written during the years 1691 to 1719, were not intended for publication but represented a kind of spiritual legacy to Glikl's 14 widely scattered children. The subjects on which she touches are extremely varied and highly informative concerning the social, cultural, and economic circumstances of the German Jews of her time. She tells of wars and plagues, of wedding feasts and bankruptcies, and of the impact on Germany of the hysteria aroused by Shabse Tsvi.

Yiddish literature in Germany and the West came to a virtual end with a handful of didactic comedies written by followers of Moses Mendelssohn and members of the Haskalah, who in principle were proponents of the repudiation of Yiddish in favour of standard German. Typical of this tendency was Arn Volfzon-Hale's *Laykhtzin un fremeley* (1796; "Frivolity and Bigotry"). It is subtitled a *Familyen-gemelde,* or "Family Portrait," and bears comparison with the bourgeois drama that had become fashionable on the contemporary German stage. The intention was to provide a modern secular alternative to the traditional Purim play, and this was certainly the first time that aspects of contemporary life had made an appearance in Yiddish drama.

Yiddish literature's demise in western Europe

Western Yiddish survived into the 19th century, and in certain areas immediately outside of the frontiers of Germany the language was still alive after World War II. No further literature of any significance was produced, however, in western Europe.

Modern literature in eastern Europe and in emigration

THE 19TH CENTURY

A resurgence of Yiddish literature in eastern Europe overlapped chronologically with the decline in the West. To a large extent it was shaped by the struggle between the Haskalah and Ḥasidism. The Ḥasidic religious revival, which originated in Podolia in the 18th century, sought communion with nature and spontaneity of worship in song and dance. The emphasis placed on individual fervour naturally favoured the free expression of feeling in Yiddish in contrast to formulaic liturgy in Hebrew. The main literary reflections of this ethos were the eulogies of the movement's founder, Israel ben Eliezer (known as the Ba'al Shem Ṭov); and above all the mystical tales of his great-grandson, Reb Nakhmen (Rabbi Naḥman) of Bratslav.

Struggle between Haskalah and Ḥasidism

The *maskilim,* adherents of the Haskalah, considered such Ḥasidic literature inimical to the enlightenment of the Jewish people and set about combating its influence by means of parody and satire. In particular it was the aim of the Haskalah to expose the supposed hypocrisy of the Ḥasidic rabbis and to ridicule the obscurantism of their tales. The Yiddish author most widely read during this period was Ayzik Meyer Dik, whose anti-Ḥasidic tone was less pronounced than that of his contemporaries, Yisroel Aksenfeld and Shloyme Etinger. Dik produced more than 400 sentimental and historical novels, many of which have disappeared because the cheap unbound editions in which they were produced were quite literally read to shreds.

The more committed maskilic writers were confined to parody by the very fact of their ideological alienation from their mass audience and from the world they described. This had the somewhat paradoxical effect of concentrating their attention on the use of colloquial language in dramatic dialogue, a fact that played a significant role in the gradual molding of the new eastern European standard literary Yiddish. At the same time, however, this restriction of scope severely limited the dimensions of their narrative perspective. A crucial step in overcoming this limitation and in the development of Yiddish literature in general was taken in 1864 when Sholem Yankev Abramovitsh (also spelled Shalom Jacob Abramowitsch) published *Dos kleyne mentshele* ("The Little Man"). It appeared anonymously in 12 installments in *Kol mevaser* ("The Herald"), the first successful Yiddish weekly journal. At the beginning of *Dos kleyne mentshele* Abramovitsh introduced the reader to his narrator, Mendele Moykher Sforim ("Mendele the Itinerant Bookseller"). Mendele's name, also spelled Mendele Mokher Sefarim, came to be regarded as the author's pseudonym, but Mendele is more accurately thought of as a brilliantly conceived narrative voice that permits his creator subtle ironic shifts of perspective. In 1865 there followed the first version of *Dos vintshfingerl* ("The Magic Ring"), which after complex revision was eventually to become his major novel. A play, *Di takse* (1869; "The Tax"), treated economic and class antagonisms within Jewish society for the first time. Abramovitsh's satirical allegory, *Di klyatshe* (1873; *The Nag*), represents the fate of the Jewish nation in the form of a prince transformed into a broken, maltreated horse.

Development of narrative technique

Yoyel Linetski benefited greatly from advances in narrative technique made by Abramovitsh. In *Dos poylishe yingl* (1869; *The Polish Boy*), which became immensely popular, Linetski adopted the device of the *faux-naïf* narrator.

When Shloyme Etinger's tragicomedy *Serkele* was performed by students at the Zhitomir rabbinical academy in 1862 it was the first time that a modern Yiddish play actually had been enacted. In 1876, however, Avrom Goldfadn (Abraham Goldfaden) traveled to Romania, where he joined forces with a group of singers to mount open-air performances. At first these were limited to a scenario-framework for his songs. Music played an important role in Goldfadn's work and many of his compositions acquired almost the status of folk songs. In 1877 the company performed his first play, *Shmendrik.* The play itself was a sentimental comedy attacking the custom of enforcing arranged matches particularly in Ḥasidic families. Its real significance lay, however, in the fact that it constituted the beginning of a professional Yiddish theatre, which was destined to play to packed houses in London, Buenos Aires, and New York City by the turn of the century, and which before it disappeared made a significant contribution to the Broadway and Hollywood industries. In response to the rising tide of anti-Semitism in the 1880s, Goldfadn's plays became more serious, more critical of superficial Haskalah rationalism, and increasingly nationalistic in tone. Though set in 14th-century Palermo, *Doktor Almosado,* which had its premiere performance in St. Petersburg in 1882, reflected the recent pogroms in Russia. Plays set in Palestine, such as *Bar Kokhba* (1883) or his final work, *Ben-Ami* (1907; "Son of My People"), were strongly Zionist in sympathy.

In the 1880s Yiddish literature as a whole saw a movement away from the militant educational campaign of the Haskalah and a corresponding tendency to espouse national values and to adopt a less ambivalent attitude toward Yiddish. This shift of emphasis was particularly evident in the works of Sholem Rabinovitsh, or Sholem Aleykhem (Sholem Aleichem), as he called his authorial persona from an early stage. With his arrival on the literary scene Yiddish literature made further strides toward self-conscious maturity. Having already shown early promise in a number of novels, stories, and feuilletons, he made a savage attack in *Shomers mishpet* (1888; "The Trial of Shomer") on the sentimental pulp fiction that in his view demeaned the status of Yiddish letters. Then in the two volumes of *Di yidishe folks-biblyotek* (1888–89; "The Yiddish Popular Anthology") he provided examples, drawn from authors such as Abramovitsh and Linetski, of writing he thought would stand comparison with works in other literatures.

The works of Sholem Rabinovitsh

In 1892 Rabinovitsh adopted the epistolary mode, in which he showed great virtuosity in his ironic handling of the constantly reborn but never gratified dreams of the average eastern European Jew. A series of monologues begun two years later featured the best known of all of Rabinovitsh's characters, Tevye der milkhiker ("Tevye the Milkman"), who narrates to Sholem Aleykhem the vicissitudes of his life. Tevye's tales epitomize the social strains and calamities suffered with faithful resignation by Jewish rural communities.

Rabinovitsh also published tales for the Jewish festivals and children's stories such as the well-known *Dos meserl* (1886; "The Penknife" in *Some Laughter, Some Tears*) and the *Motl Peyse dem khazns* cycle (1907–16; *Adventures of Mottel, the Cantor's Son*).

Returning to the novel in later life, Rabinovitsh dealt with the theme of the Jewish artist in *Blondzhende shtern* (1909–11; *Wandering Star*), giving a picaresque panorama of Yiddish theatre life stretching from eastern Europe to London's Whitechapel and Manhattan's Lower East Side. His comparatively few writings for the theatre include the posthumously published *Di goldgreber* (1908; "The Gold Diggers") and *Dos groyse gevins* (1916; "The Grand Prize"). In addition he adapted a number of his stories for the stage, including *Tevye der milkhiker.* (*Tevye* was made into a successful motion picture in 1939 and later was adapted as a musical comedy and motion picture under the title *Fiddler on the Roof.*)

THE 20TH CENTURY

To World War I. Together with Abramovitsh and Rabinovitsh, Yitskhek Leyb Perets is regarded as one of the three major classical writers in Yiddish literature. It was Perets who effectively ushered Yiddish literature into the modern era by exposing it to contemporary trends in western European art and literature. He was influenced by Polish neo-Romantic and Symbolist writings, and under their impact he lent new expressive force to the Yiddish language in numerous stories collected as *Khsidish* (1908; "Hasidic Tales") and *Folkstimlekhe geshikhten* (1908; "Folktales"). In these stories Ḥasidic material is viewed obliquely from the standpoint of a secular literary intellect and becomes the vehicle for an elegiac contemplation of traditional Jewish values.

Perets played an important moderating role as deputy

chairman at the Yiddish Conference that assembled at Czernowitz in 1908 to promote the status of the language and its culture. He is also remembered for the encouragement that he gave to a whole generation of younger writers who flocked to his home in Warsaw.

Perets and several other writers participated in the extraordinary burgeoning of literary activity that manifested itself above all in the foundation of literary periodicals with aesthetic programs. Particularly influential were the *Literarishe monatshriftn* ("Monthly Literary Review"), four issues of which appeared in Wilno (Vilnius) in the spring of 1908, carrying contributions by Perets, Der Nister (pseudonym of Pinkhes Kahanovich), Perets Hirshbeyn, and Sholem Ash (also spelled Sholem Asch).

Effects of World War I on Yiddish culture

World War I and its aftermath transformed Yiddish cultural geography. The war itself severely hampered contacts between the Jewish heartland in eastern Europe and nascent Yiddish intellectual life in the New World, while the Bolshevik Revolution, the Russian Civil War, and the ensuing Russo-Polish War severed the Soviet centres from literary activities in the resurgent Polish Republic. Furthermore, an ideological exchange took place, with Jewish intellectuals both fleeing from and migrating to the Soviet Union. Consequently, in dealing with Yiddish literature after World War I, it is appropriate to give separate attention to developments in the three major cultural centres.

After World War I. *The Soviet Union.* Of all the prewar Jewish parties in Russia, the most uncompromising in their support for Yiddish had been the members of the Jewish Labour Bund. Like the *maskilim* before them, the Bundists saw the language as a means to an end, in this case Socialism. The majority of the Bundists eventually joined forces with the Bolsheviks and took control of Jewish cultural institutions in the Soviet Union. While Hebrew literature and other supposedly chauvinist aspects of traditional Jewish culture were ruthlessly suppressed, Yiddish writers enjoyed unparalleled facilities. Before the war the Bundists had been bitterly opposed to the art-for-art's-sake trends in modernist Yiddish literature, but in the era immediately following the revolution much of what was published had, in fact, been written earlier and was Impressionist or Symbolist in style. Some of the best of this writing appeared in two important collections published in Kiev during the turbulent Russian Civil War era. *Eygns* (1918 and 1920; "Our Own") was edited by Dovid Berglson and Der Nister, while *Der oyfgang* (1919; "Ascent") was edited by Dovid Hofshteyn. In addition to works by the editors, these collections included significant contributions by Leyb Kvitko and Perets Markish that were stylistically innovative; at the same time they drew heavily on the themes of Russian Futurism and recapitulated Romantic and Expressionist developments in the European lyric that had previously been absent from Yiddish literature. All of these writers emigrated in 1920 or soon after but returned to the Soviet Union during the era of the New Economic Policy, when emphasis was placed on the need to win support among the more sympathetic of the émigré intellectuals.

Dovid Berglson (also spelled David Bergelson) had first made a name for himself with the short story *Arum vokzal* (1909; "At the Railway Station"; translated as "At the Depot" in *A Shtetl and Other Yiddish Novellas,* 1973) and with his masterpiece, *Nokh alemen* (1913; *When All Is Said and Done*), an impressionistic novel that brilliantly captures the atmosphere of unfulfilled longing in a Ukrainian provincial town. While in Berlin in 1924 he and Der Nister edited *Der Milgroym* ("Pomegranate"), one of the most impressive Yiddish literary journals ever produced. Increasingly, however, he thought of the Soviet Union as the only place where Yiddish literature could survive, and he went on to edit *In shpan* (1926; "In Harness"), a pro-Soviet journal that was intended to offer émigré writers an opportunity to align themselves with the supposed march of history.

Der Nister had been known as the author of brilliant Symbolist stories such as *Kleopatre* (1908; "Cleopatra"). In 1929 he published in *Di royte velt* ("The Red World") an extraordinary story entitled "Unter a ployt" ("Under a Fence"), a complex, phantasmagoric tale that expresses the agonies of artistic submission to external repression and perhaps best articulates the quintessence of Jewish intellectual life in the Soviet Union in those years.

In 1929 ideological rigidity once again gained the upper hand. Socialist Realism became the watchword and Yiddish writers increasingly found themselves accused of "bourgeois nationalism." During the 1930s the rate of Yiddish publishing actually increased, but it consisted largely of simplistic paeans to Stalin and to the system. The more sensitive writers felt themselves limited to the translation of foreign classics, to the production of "corrected" versions of earlier work, and to historical topics. Even within these restraints some remarkable works were produced, such as Berglson's semiautobiographical novel *Bam Dnyeper* (1932, 1940; "By the Shores of the Dnepr") or Der Nister's outstanding genealogical chronicle *Di mishpokhe Mashber* (1939; "The Mashber Family"), the action of which is set in the 1870s and which thereby sidesteps ideologically sensitive areas. Most remarkable perhaps among Yiddish publications of these years was *Zelmenyaner* (1929–30; "The Zelmenians") by Moyshe Kulbak, who was almost the only writer who dared to look at contemporary reality with a critical eye.

Political purges of Yiddish writers

Many writers were victims of the political purges of the 1930s; "proletarian" or Marxist orthodoxy was no guarantee against arrest. The World War II years brought with them a renewed freedom for Yiddish authors to write about Jewish aspirations and causes. Soon after the war, however, the Zhdanovite repression bore down with particular severity on Yiddish cultural life, and by 1948 virtually all of the leading figures of the Yiddish intelligentsia had been arrested. Some died in prison. Almost all those still alive on Aug. 12, 1952, were shot on Stalin's orders.

In 1955 and 1956 many of Stalin's victims were posthumously "rehabilitated," and since then Yiddish books and periodicals have appeared on a limited scale.

Poland. During the interwar years the Republic of Poland, independent once again after an interval of more than 120 years, together with Romania and Lithuania constituted the most fertile of the Yiddish cultural areas. An exponential increase in the number of talented writers at work produced a wealth of literary developments existing side by side. Naturalism was exemplified by Oyzer Varshavski's novel *Shmuglars* (1920; "Smugglers"), which depicts underworld life during the German occupation. Shloyme Zaynvl Rapoport, who wrote under the name of S. Ansky, was much influenced by the neo-Romanticism of Perets. His play *Der dibek* (1919; translated as "The Dybbuk" in *The Dybbuk and Other Great Yiddish Plays,* 1966) was performed by the celebrated Vilna Troupe in Warsaw and was received with great acclaim. It went on to attract international attention with performances in translation in Kraków, Berlin, Vienna, and New York City, all in 1925. (Mikhl Vashinski's screen version of *Der dibek* [1937] is perhaps the most successful of all the many Yiddish films made in Poland between the wars.)

The most self-conscious break with the past was made by the Expressionists. In 1919, writing in his short-lived but influential journal *Jung-yidish,* Moyshe Broderzon described the group of exuberant iconoclasts to which he belonged as a "*freylekhe . . . khalyastre*" or "merry gang," and this appellation became the title for the almanac *Khalyastre,* published in Warsaw and Paris in 1922–24, in which they proclaimed their modernist creed. The main protagonists in addition to Broderzon were Uri-Tsvi Grinberg, Melekh Ravitsh, and Perets Markish, who had gone to Warsaw from Kiev in 1921. Influenced both by the Futurism of the Russian revolutionary poet Mayakovsky and the Expressionism of the German Jewish writers Werfel and Else Lasker-Schüler, they sought to participate in the general European modernist movement and to respond to the anguish and chaos of the postwar world with universal images of apocalyptic pessimism. It was in Warsaw also that Markish first published his collection of poems entitled *Di kupe* (1921; "The Heap"), which is a Kaddish, a lament for the victims of a pogrom whose mutilated corpses lie heaped in a Ukrainian market square.

Khalyastre writers in Warsaw

Di kupe represented an important station in the history of Jewish artistic responses to anti-Semitism that

had included the 17th-century *lider* and Ḥayyim Naḥman Bialik's famous Hebrew poem on the Kishinyov pogrom of 1903. Grinberg, who in 1918 had himself narrowly escaped death in a pogrom in Lvov, voiced a more strident reaction in his *Uri-Tsvi farn tseylem* ("Uri-Ẓvi Before the Cross"), which he published in the second issue of his journal *Albatros* (1922). Grinberg eventually turned his back on both Europe and Yiddish in order to become a Hebrew poet in Palestine. Markish's Socialist ideals took him back to the Soviet Union, while Ravitsh left Poland and traveled widely before settling in Canada.

An associate of the group while it lasted was Yisroel-Yeshue Zinger (known in English as Israel Joshua Singer). He was joint editor of the first issue of *Khalyastre,* and he contributed to it an impressionistic story, *In der fintster* ("In the Dark"), written in Kiev during the Russian Civil War. Zinger, however, soon moved away from his neo-Romantic beginnings toward Realism. A semiautobiographical novel, *Shtol un ayzn* (1927; *Blood Harvest,* U.S. title *Steel and Iron*), reflects his wartime experiences. In *Yoshe Kalb* (1932) Zinger returned to his roots and wrote a novel of repressed sexual passion set amid the corruption and hypocrisy of a tyrannical Ḥasidic court. After the triumphant success of a stage version produced in New York City by Maurice Schwartz, Zinger emigrated to the United States. It was there that he wrote his masterpiece, *Di brider Ashkenazi* (1935; *The Brothers Ashkenazi*), which traces the antagonisms of two brothers against the background of the development of the textile industry in Łódź from the end of the Napoleonic era to the rebirth of the Polish Republic.

Itsik Manger was born in Austrian Bukovina and began writing in Romania, but he moved in 1928 to Warsaw, where a year later he published his first book of collected verse. His major achievements were *Khumesh lider* (1935; "Pentateuch Songs") and *Megile lider* (1936; "Songs of the Book of Esther"), in which he feigns the naïveté of an itinerant minstrel and draws on the Purim play tradition to give an anachronistic and vivid account of biblical stories.

The rich and varied Yiddish literary culture of Poland was brought to an abrupt and tragic end by the Nazi invasion in 1939. Nevertheless, poems, plays, and songs continued to be written in the ghettos and camps, and a small proportion survives.

The United States. With the mass migrations to the United States, and especially to the Lower East Side of New York City, in the 1880s, Yiddish literature made its first appearance in the New World. A volume of exhortatory Haskalah verse appeared as early as 1877. This early beginning was followed by a generation of "sweatshop poets." Moris Vintshevski represented a transition from the Haskalah to Socialism. He arrived in the United States in 1898 having already made a reputation for himself in London as a writer of propagandistic verse. Moris Roznfeld, like many of his readers, also came to New York City via the East End of London. He worked for many years in the tailoring shops of both cities and his unsophisticated verse eloquently expresses the drudgery and alienation experienced by the immigrant labourer.

"Sweatshop poets"

Avrom Lyesin and Avrom Reyzn made a more ironic and Romantic contribution to the development of Yiddish verse in America, while Yehoash Shloyme Blumgartn, who wrote under the name Yehoash, captured the sights and sounds of the metropolis and introduced the world of nature to the Yiddish lyric in the New World. His supreme achievement was a meticulous and scholarly translation of the Old Testament into modern Yiddish.

A significant change of direction occurred with the emergence of Di Yunge, a group of young poets associated with the journal *Di yugnt* ("Youth") that was published in New York City in 1907–08. They had been influenced by the Jewish Labour Bund and the revolutionary ferment in Russia, but when they went to the United States after the disappointments of the failed revolution of 1905 they no longer had any patience with the tendentious rhyming of their predecessors. The paradox of their position was that while their daily experience remained that of immigrant working-class life, they viewed it very largely from the perspective of European aestheticism. They were linguistic purists and also adept translators of the French, German, and Russian Symbolists who were their inspiration. The main protagonists of this movement were Mani Leyb, Zishe Landoy, and Moyshe Leyb Halpern. Another who was embraced by the movement was H. Leyvik (also spelled H. Leivick; pseudonym of Leyvik Halper, who was best known for his verse drama *Der goylem* (1921; "The Golem"). For the most part Di Yunge were unacquainted with English literature, but Yisroel-Yankev Shvarts (I.J. Schwartz) translated Walt Whitman's poetry and adopted something of his manner in the epic poem *Kentoki* (1925; "Kentucky"), in which the exoticism and the wide open horizons of the American South entered Yiddish literature for the first time.

Innovations of Di Yunge

After World War I a more radical rebellion challenged the innovations of Di Yunge. In the anthology *In zikh* (1920; "Introspection") Arn Glants-Leyeles, Yankev Glatshteyn, and Nokhem Minkov asserted that the world exists only insofar as it is reflected in subjective impressions, each of which demands its own unique formal expression, ideally in free verse. Unlike Di Yunge, the Inzikhistn, as they came to be called, had attended universities in London and New York City and were influenced by contemporary English and American literature, in particular by Ezra Pound, T.S. Eliot, Wallace Stevens, and the Imagists. In later years Glatshteyn became one of the outstanding figures in American postwar Yiddish literature. He no longer strove after universal themes, and his style became less experimental as he turned increasingly to elegiac verse mourning the destruction of traditional Jewish life in eastern Europe.

Works of Sholem Ash

Sholem Ash became the first Yiddish author to have a truly international following. He visited the United States for the first time in 1910 and spent most of the rest of his life there. He moved from genial portrayal of life in eastern Europe in *A shtetl* (1904) and in his colourful underworld novel *Motke ganev* (1917; *Mottke the Thief*) to American locations with *Onkl Mozes* (1918; *Uncle Moses*) and several historical novels. Ash became the subject of controversy when he chose to write novels about early Christianity, so much so that of his trilogy published in English as *The Nazarene* (1939), *The Apostle* (1943), and *Mary* (1949) only the first volume appeared in the original Yiddish, as *Der man fun Netseres* (1943). Critics remained hostile to his subsequent work, including his most successful novel, *Ist river* (1946; *East River*), which looks with sanguine idealism at interaction between Irish and Jewish immigrants of the Lower East Side at the turn of the century.

Lamed Shapiro (pen name of Levi Yeshue Shapiro) and Yoysef Opatoshu (pen name of Yoysef Meyer Opatovski) were prose writers of the same generation as Ash who migrated to the United States in their 20s. Shapiro achieved notoriety with a brilliant series of stories constituting a psychopathology of the pogrom, in which the action is seen through the eyes of the perpetrators of the violence. Opatoshu wrote realistic stories set in New York City and historical novels. Khayim Grade, who lived in Vilnius until World War II, became a leading figure in the Yung-Vilne movement of the late 1930s, and his work was primarily devoted to literary interpretation of the Jewish life of that city. By way of the Soviet Union and Paris he emigrated to the United States in 1948, and there he continued to re-create in fiction the lost world of his native city.

Works of I.B. Singer

Undoubtedly the most widely known Yiddish writer is Yitskhek Basheyvis, known to his English readers as Isaac Bashevis Singer, who was awarded the Nobel Prize for Literature for 1978. Singer, the younger brother of Yisroel-Yeshue Zinger, was born near Warsaw in 1904 and grew up in the Jewish quarter of Warsaw and in the Galician shtetl, Biłgoraj. His father, grandfather, and uncle were all Ḥasidic rabbis, and the atmosphere of traditional piety that surrounded them is directly reflected in his confabulation of reminiscence and imagination entitled *Mayn tatns bezdn-shtub* (1956; *In My Father's Court*). The work invites comparison with his brother's autobiographical volume *Fun a velt vos iz nishto mer* (1946; *Of a World That Is No More*).

Singer first won recognition in 1935 with his *Der sotn in*

Goray (*Satan in Goray*), a short novel depicting a 17th-century Galician shtetl in thrall to the Sabbatian cult. The underlying themes are constituted by the author's preoccupation with the preternatural and with the uneasy relationship between piety and sensuality. In the same year he followed his elder brother to New York City.

Broadly speaking, the fiction of Singer may be divided into three genres defined by their locales. In his historical novels like *Der sotn in Goray* his protagonists strive after equilibrium between instinct and intellect in a world that is governed by Kabbalah (Jewish mysticism) and occult folk beliefs. *Der kuntsnmakher fun Lublin* (1971; *The Magician of Lublin*) and *Der knekht* (1962; *The Slave*) illustrate this tendency. Another group of novels deals with Jewish life in Poland from the late 19th century to the Nazi invasion. These works are more realistic and present an ambivalent view of assimilation and loss of cultural cohesion. *Sonim: di geshikhte fun a libe,* which was serialized in *Forverts* and translated as *Enemies, A Love Story* (1972), has as its major theme deracination and moral disorientation suffered by Holocaust survivors in the New World, while in *Der bal tshuve* (1974; "The Penitent"), which is set in Israel, Singer probes further his recurrent themes of the Don Juan's yearning for female innocence and the envy of the cultivated freethinker for the moral certainties of the pious philistine.

Impressive as the novels of Singer are, it is in the short story that he developed his stylistic virtuosity to the full, often employing monologue techniques to produce a rich variety of narrative voices and perspectives in a manner that puts him in a direct line of descent from Mendele and Sholem Aleykhem. He published his first story in a Warsaw journal in 1925 and from that time wrote a large number of short stories for a variety of Yiddish periodicals. A small selection has been collected in book form as *Der sotn in Goray un andere dertseylungen* (1943; "Satan in Goray and Other Stories"), *Gimpl tam un andere dertseylungen* (1963; first published in English as *Gimpel the Fool and Other Stories,* 1953) and *Mayses fun hintern oyvn* (1971; "Stories from Behind the Stove"). A much larger selection of stories has appeared in English translation in numerous volumes.

Yiddish theatre in the United States

Yiddish theatre arrived in the United States by way of London, where it had enjoyed a brief heyday in the 1880s. The first professional production in the United States took place in 1882 (despite much outraged protest on the part of New York City's German Jews) when the young Boris Tomashevsky made his debut in Goldfadn's *Koldunye* ("The Witch"). On the whole, artistic standards were not high, but the efforts of Yankev (Jacob) Gordin brought a considerable improvement. His *Sibirye,* which opened in November 1891, and *Der pogrom in Rusland,* produced two months later, brought a purer Yiddish as well as more serious and realistic content to the Yiddish stage, though his work was often marred by melodramatic moralizing. In his prolific output he frequently contented himself with free adaptations from classical English, German, and Russian drama. His most successful works were *Der yidisher kenig Lir* (1892), based on Shakespeare's *King Lear,* and what is perhaps the most popular play in the entire Yiddish repertoire, *Mirele Efros* (1898), which reworks the same theme with a female protagonist.

The greatest achievements of Yiddish theatre took place in the years following World War I. Dramatists such as Sholem Ash, Dovid Pinski, and Perets Hirshbeyn arrived from Europe and began working with theatre groups subsidized by Yiddish cultural organizations. A major impact was also made by European touring companies such as the Vilna Troupe. Hirshbeyn wrote rural idylls such as *A farvorfn vinkl* (1912; "A Secluded Corner") and *Grine felder* (1916; "Green Fields"), which reflected his experience of Jewish farming life both in his native Byelorussia and in the Catskill Mountains. Both plays were successfully staged in New York City by Yankev Ben-Ami, who had worked in Hirshbeyn's own company in Odessa before the war. While studying in Berlin, Pinski had been influenced by the Naturalist drama of the prolific German writer Gerhart Hauptmann. Pinski arrived in the United States in 1899 and continued his studies of German literature at Columbia University. His *Der oytser* (1906; *The Treasure*) is an ironic comedy satirizing avarice in a Russian shtetl. It was given its premiere performance in German translation by the German-Jewish theatrical director Max Reinhardt in Berlin and became an international success. In *Dovid ha-meylekh un zayne vayber* (1914; "King David and His Wives") Pinski depicts the growing cynicism and hedonism of the King who, nonetheless, in old age achieves an almost Faustian insight into the elevation that comes from striving after an ideal. In 1949 Pinski, who had for decades been active in the Poalei Ẓion movement, moved to Haifa, where he became the focus of the Yung-Yisroel ("Young Israel") group.

Survival of Yiddish in Israel

Israel. Despite zealous promotion of Hebrew, the waves of eastern European immigrants reaching Israel also brought with them Yiddish literature and culture. Traces of Ashkenazic use of Yiddish in Palestine go back a long way, but it was only after World War I that the country became a significant, though still relatively marginal, centre for Yiddish literary culture. One of the earliest immigrant writers who chose to remain faithful to Yiddish was Yoysef Papyernikov. After contributing poems to Warsaw periodicals he moved in 1924 to Palestine. There he produced volumes of lyric poetry, such as *In zunikn land* (1927; "In the Land of the Sun"), that celebrate the landscape and the rebuilding of the land by the Jewish pioneers.

Like Khayim Grade, Avrom Sutskever played an active part in the Yung-Vilne movement. He pursued literary studies at the University of Vilnius and contributed to *In zikh.* During the war he was in the Wilno ghetto from which he escaped to fight with the partisans; he settled in Palestine in 1947. His experiences of the war and of the Holocaust are reflected in many of his works, including the long poems *Geheymshtot* (1948; "Secret City"), *Kol nidre* (1963), and *Sibir* (1952; *Siberia*). The quarterly journal *Di goldene keyt* ("The Golden Chain"), which he began in 1949, has increasingly assumed the role of the major forum for Yiddish cultural life.

BIBLIOGRAPHY. General studies of the literature and of periods in its development are: CHONE SHMERUK, "Yiddish Literature," in *Encyclopaedia Judaica,* vol. 16, cols. 798–833 (1971); DAN MIRON, *A Traveler Disguised: A Study on the Rise of Modern Yiddish Fiction in the Nineteenth Century* (1973); CHARLES A. MADISON, *Yiddish Literature: Its Scope and Major Writers* (1968, reissued 1971); SOL LIPTZIN, *The Flowering of Yiddish Literature* (1963), *The Maturing of Yiddish Literature* (1970), and *A History of Yiddish Literature* (1972); ISRAEL ZINBERG, *Old Yiddish Literature from Its Origins to the Haskalah Period,* vol. 7 of his *History of Jewish Literature,* 12 vol., trans. and ed. by BERNARD MARTIN (1972–78; originally published in Yiddish, 1955–71); N.B. MINKOFF and JUDAH A. JOFFE, "Old Yiddish Literature," and SHMUEL NIGER, "Yiddish Literature in the Past Two Hundred Years," both in RAPHAEL ABRAMOVITCH, *et al.* (eds.), *The Jewish People, Past and Present,* vol. 3, pp. 145–164 and 165–219, respectively (1952); LEO WIENER, *The History of Yiddish Literature in the Nineteenth Century* (1899, reprinted 1972 with a new introduction by ELIAS SCHULMAN); YITZHAK KAHN, *Portraits of Yiddish Writers,* trans. from the Yiddish by JOSEPH LEFTWICH (1979); A.A. ROBACK, *The Story of Yiddish Literature* (1940, reprinted 1974), and *Contemporary Yiddish Literature: A Brief Outline* (1957); and ISRAEL C. BILETZKY, *Essays on Yiddish Poetry and Prose Writers of the Twentieth Century,* trans. from the Hebrew by YIRMIYAHU HAGGI (1969). More specific topics are discussed in CHONE SHMERUK, "Yiddish Literature in the USSR," in LIONEL KOCHAN (ed.), *The Jews in Soviet Russia Since 1917,* 3rd ed. (1978), pp. 242–280; and RUTH R. WISSE, "Di Yunge: Immigrants or Exiles?", in *Prooftexts: A Journal of Jewish Literary History,* vol. 1, no. 1, pp. 43–61 (January 1981). Among anthologies are IRVING HOWE and ELIEZER GREENBERG (comp.), *Voices from the Yiddish: Essays, Memoirs, Diaries* (1975); JOSEPH LEFTWICH (comp. and ed.), *An Anthology of Modern Yiddish Literature* (1974); RUTH WHITMAN (ed. and trans.), *An Anthology of Modern Yiddish Poetry* (1966); and RUTH R. WISSE, *A Shtetl and Other Yiddish Novellas* (1973, reissued 1986). For a history of Yiddish theatre, see NAHMA SANDROW, *Vagabond Stars* (1977).

For further bibliography, see DINA ABRAMOVITCH, *Yiddish Literature in English Translation: Books Published 1945–1967,* 2nd ed. (1968), and *Yiddish Literature in English Translation: List of Books in Print* (1976); and LEONARD PRAGER and A.A. GREENBAUM, *Yiddish Literature and Linguistic Periodicals and Miscellanies: A Selective Annotated Bibliography* (1982).

(H.F.D.)

Yugoslav Literature

Just as the history of the various peoples of modern Yugoslavia has differed over the past thousand years, so the literatures of the Serbs, Croats, Slovenes, and Macedonians have developed independently. From an early date Croats and Slovenes formed part of the central European cultural heritage, whereas Serbs and Macedonians were part of the Eastern, or Byzantine, culture and, later, of the Turkish sultanate. Since 1918, when they were united, the Yugoslavs have had three literary languages—Serbo-Croatian, Slovene, and Macedonian—and two alphabets, the Latin and the Cyrillic. The Serbs and Croats speak a common language but use the Cyrillic and Latin alphabets, respectively. The Slovenes, with a language akin to Serbo-Croatian, use the Latin alphabet and the Macedonians the Cyrillic.

At times in the 20th century the literatures of the Serbs and Croats have been considered to be the expression of one national culture. This was particularly the case during the period immediately after World War I, when a widespread feeling of euphoria accompanied the creation of a unified Yugoslav state. At the same time, however, internal rivalries led to the resurgence of separatism among the country's major national groups. After World War II and the coming to power of the Communist Party of Yugoslavia, the modern tendency toward political unity vied with the historical legacy of differing cultural identities. Following Yugoslavia's expulsion from the Soviet bloc in 1948, official insistence on Socialist Realism in literature and the other arts began to decline, and from the mid-1950s authors reflected Western literary trends and styles. This cosmopolitan aspect of contemporary Yugoslavia is balanced by a continuing sense of attachment to the traditional cultural centres of Belgrade by the Serbs, Zagreb by the Croats, Ljubljana by the Slovenes, and Skopje by the Macedonians.

This article is divided into the following sections:

Serbian literature 1083
Croatian literature 1083
Slovene literature 1084
Macedonian literature 1085
Bibliography 1085

SERBIAN LITERATURE

Serbian literature originated in the use of the Old Church Slavonic language for the translation of religious works from Greek. From the 12th century, Serbian literature acquired its own characteristics, as in the illuminated Miroslav Gospel; biblical stories; hagiographies, notably that of St. Sava, patron saint of Serbia; and tales such as that of Barlaam and Josaphat. After the Turkish occupation of 1459, literature declined.

The most important Serbian representative of the Enlightenment period was Dositej Obradović, whose writings greatly influenced Serbian literary development. Many characteristics of European Romanticism could be observed in the literature of the period 1820–70, especially the cult of folklore and national self-assertion. A central figure was Vuk Stefanović Karadžić, a reformer of literary language who wrote a Serbian grammar and dictionary and collected Serbian folk poetry and stories. Petar II Petrović Njegoš, a Montenegrin ruler, was a poet whose gifts are best seen in his dramatic poem *Gorski vijenac* (1847; "The Mountain Garland"). The lyrical verses of Branko Radičević contributed to the break with earlier didactic-objective poetry. Among younger Romantic writers were the lyrical poets Jovan Jovanović (known as Zmaj), Đura Jakšić, and Laza Kostić.

From 1870 to 1900 there was a tendency toward Realism. The most significant narrative writing was by Laza Lazarević, a master of the short story; Simo Matavulj, a keenly observant novelist and short-story writer with a strong sense of humour; and Stevan Sremac, satirist and humorist. Vojislav Ilić was a fine descriptive poet, resembling the French Parnassians.

Toward the end of the 19th century Serbian literature was influenced by European currents, particularly the French Symbolists: outstanding figures were Jovan Dučić and Milan Rakić, a patriotic poet of a melancholy nature. Vladislav Petković (called Dis), a subjective poet, chose themes halfway between reality and the irrational. Prose changed from objective description to psychological narrative, often with a lyrical quality, as in the work of Borisav Stanković.

Serbian writers between the world wars

Serbian writers between World Wars I and II continued to follow major European literary movements. Miloš Crnjanski and Rastko Petrović were exponents of Modernism in the years immediately after World War I. The Belgrade Surrealist group introduced a note of radical, left-wing politics, and some of its members later turned to the style of Socialist Realism. The literature of the 1930s was shaped by the focus on political and social themes and attacks on the introspection of the earlier Modernists. World War II brought a break in literary output, but the end of the war was accompanied by the publication of three novels by Ivo Andrić: *Travnička hronika* (1945; *Bosnian Story*), *Na Drini ćuprija* (1945; *The Bridge on the Drina*), and *Gaspodjica* (1945; *The Woman from Sarajevo*). Andrić won the Nobel Prize for Literature for 1961. His works, usually placed in a historical setting, combine the realistic traditions of Serbian literature with more modern forms of expression.

The postwar period witnessed the continuation of the realistic trend in the war novels of Mihailo Lalić (*Lelejska gora* [1957; *The Wailing Mountain*]) and Dobrica Ćosić (*Daleko je sunce* [1951; *Far Away Is the Sun*]) and in the work of Meša Selimović, which has a strong emphasis on psychology. During the 1960s and '70s this trend turned to contemporary settings in a specifically Serbian milieu with strong local colouring, such as in the prose of Dragoslav Mihailović and Milisav Savić. More original forms of expression were introduced into prose from the 1950s by Oskar Davičo (*Pesma* [1952; *The Poem*]), who had links with the Surrealists before the war, and by the postwar writer Miodrag Bulatović (*Crveni petao leti prema neku* [1959; *The Red Cockerel*]). In poetry Serbia has been represented by Desanka Maksimović, Vasko Popa, Stevan Raičković, Miodrag Pavlović, and Ivan Lalić (*The Works of Love,* English translations published in 1981). Later developments have included novels with greater social and political comment, such as Danilo Kiš's *Grobnica za Borisa Davidoviča* (1976; *A Tomb for Boris Davidovich*), and experiments with the forms of literature by the Klokotrizam group, the intention of which is to defy the canons and aesthetic norms of art.

The 20th century has been fraught with polemics between different literary groups. The arguments from the turn of the century to the 1960s centred on the question of style. The tradition of realistic narrative has maintained a strong position in the Serbian literary establishment, which in many cases has been opposed to more innovative writers. Since the late 1960s the focus of debate has shifted away from style and toward questions of content and the place of political and social comment in the literature of a socialist society.

CROATIAN LITERATURE

The Croats were converted to Christianity at the turn of the 9th century by disciples of SS. Cyril and Metho-

dius. The first ecclesiastical works were written in the Glagolitic alphabet, and fragments have survived in the Kiev Missal pieces (11th century). From the 12th century the national language was widely used for inscriptions and legal documents, but from the 14th century Latin was increasingly used. Immediately before the age of Humanism (second half of the 15th century), literature consisted mainly of biblical stories, legends, folklore, and popular stories.

In the 15th and early 16th centuries the most outstanding Croatian writer was Marko Marulić, author of the epic *Judita* (1501), in which he encouraged his countrymen in their struggle against the Turks. Under the influence of classical and Italian literature and Croat folklore, the Petrarchan lyric was developed in Dalmatia and in Dubrovnik. One of the most prominent poets was Hanibal Lucić, author of *Robinja* ("The Slave Girl"), the first Yugoslav secular play. Among the best known writers of Old Croat literature was Marin Držić, who wrote pastoral dramas and comedies portraying Renaissance Dubrovnik. His comedy *Dundo Maroje* (written *c.* 1551) was performed throughout western Europe.

In the Croatian literature of the 17th and 18th centuries, the most important position belonged to Ivan Gundulić, author of a stirring epic, *Osman,* describing the Polish victory over the Turks in 1621. Andrija Kačić Miošić treated historical themes in the popular verse forms of the travelling minstrel.

Gundulić's epic

Romanticism in Croatian literature originated in a cultural and national reformation brought about by the Illyrian movement (1835–48), which aimed at a union of all south Slavs within the Habsburg federation. Ljudevit Gaj introduced the *štokavski* dialect and *ijekavski* speech as the literary language of Croatia and also a unified orthography. Personal, patriotic, and reflective lyrics were popular and were best represented by the sensitive, moving poems of Stanko Vraz, the patriotic songs and poetic drama of Petar Preradović, and the dramatic works of Dimitrije Demeter. Some of the best poetry includes *Smrt Smail-Age Čengića* (1846; *The Death of Smail-Aga Čengić*) by Ivan Mažuranić.

The popularity of patriotic verse

In the development of national characteristics of Croatian Romanticism an outstanding figure was August Šenoa, poet, dramatist, critic, journalist, and creator of the Croat historical novel. Šenoa was a subtle stylist and a champion of Realism, which was at its height in 1881–95. Writers were concerned with contemporary problems, particularly with conditions among the lower classes. Representatives of this movement were Evgenij Kumičić, a naturalist and poet; Ante Kovačić, satirist and author of a lively novel, *U registraturi* (1888; "In the Registry Office"); and Ksaver Šandor Gjalski, a short-story writer and novelist concerned with political and cultural problems. A poet, Silvije Strahimir Kranjčević, was a social rebel whose work was rich in imagery and rhythmic variety.

In the opening years of the 20th century poetry was the dominant genre in Croatian literature. Many poets reflected the style of the art-for-art's-sake trend, dwelling on the inner struggles of modern man with his world and the search for meaning in individual existence. These themes are found in particular in the work of Vladimir Vidrić and Vladimir Nazor. The leading figure of the early Modernist phase until World War I was Antun Gustav Matoš. He edited the anthology *Mlada hrvatska lirika* (1914; "The Young Croatian Lyric"), which marked the zenith of such verse. Between the wars avant-garde poetry continued to be expressed in the verse of such poets as Augustin Ujević, while Ivan Goran Kovačić in *Jama* (1943; *The Pit*), a long poem evoking the horror of war, retained a classical elegance in his verse. The short story was represented by the work of Slavko Kolar, who humorously depicted the life of the peasant in a changing world. The dominant prose writers of the interwar period were August Cesarec and Miroslav Krleža (*Povratak Filipa Latinovicza* [1932; *The Return of Philip Latinovicz*]; and the collection of English translations *The Cricket Beneath the Waterfall and Other Stories* [1972]). Both presented contemporary social problems as the result of class exploitation and deeply explored the psychology of their characters. Krleža is known not only for his creative writing, which spanned the century to his death in 1981, but also as an editor of literary periodicals, as an essayist, and as a critic who dominated Croatian cultural life for much of the century. Although a committed supporter of the Communists in Yugoslavia from the mid-1920s, Krleža always opposed the imposition of Socialist Realism as official party policy. His attitude was to support the freedom of the artist against all forms of social and political coercion in any type of society.

In the less restrictive atmosphere that followed the break in 1948 in relations with the Soviet Union, new prose writers included Ranko Marinković and Vjekoslav Kaleb (*Divota prašine* [1956; *Glorious Dust*]), who wrote on the war and contemporary society in Croatia. The younger prose writer Antun Šoljan took more cosmopolitan themes for his work, as did the poet Ivan Slamnig of the same generation. Aleksander Flaker of the University of Zagreb gained an international reputation as a scholar in the field of literary studies. Many writers gave their support to the national revival known as the Croatian Spring in 1970–71; demands included the recognition of Croatian as a separate literary language. The government stifled the movement, but it served as an example of the strength of feeling for cultural identity in Yugoslavia and as an indication of the potential political role of literary affairs. Later Croatian literature shared another concern with Western literature in the cosmopolitan choice of its themes, such as were found in the feminist writing of Dubravka Ugrešić.

SLOVENE LITERATURE

The dissolution in 821 of the independent state of Carantania (Carinthia; now Kärnten, Austria) and a merging of the upper classes into German feudal society retarded any development of Slovene literature. Among medieval manuscripts that have been preserved, the most remarkable are the *Brižinski spomeniki* (Freising manuscripts, *c.* 1000). These are the first written examples of the Slovene, and perhaps of any Slavic, language. Alongside this appearance of the language in its written form there flourished, as throughout the Balkans, a rich folk poetry. The use of written Slovene in the church liturgy was replaced by Latin at an early stage.

The Freising manuscripts

The demand for Slovene to be reintroduced into the church was made during the rise of Protestantism in Europe. The developing religious literature reached its peak in 1584 in a translation of the Bible by Jurij Dalmatin. Other developments of this period included the publication, also in 1584, of the first Slovene grammar, by Adam Bohorič. Most of this early work was undone, however, by the Counter-Reformation. It was not until the middle of the 18th century that interest in a Slovene literary language was revived under the influence of the Enlightenment.

The European Romantic movement, with its support of nationhood based on a common language and shared cultural background, encouraged philological and literary works and influenced Slovene intellectuals in the first half of the 19th century. Jernej Kopitar, a Slovene and censor for Austria-Hungary responsible for the Slav nations of the empire, initiated early attempts to standardize and codify the Slovene language. In the 1830s and '40s Matija Čop was very active in publishing literary works and literary periodicals in Slovenia. The greatest poet of those years up to the middle of the 19th century was France Prešeren. He raised the standards of literary Slovene to new heights and established the literary language as a vehicle capable of conveying nuances of thought and feeling in a modern idiom.

Standardization of Slovene

The second half of the 19th century witnessed the development of prose writing and the influence of Realism. A dynamic writer was Fran Levstik, a critic and dramatist and founder of a program for popular prose. The first Slovene novel, *Deseti brat* ("The Tenth Brother") by Josip Jurčič, was published in 1866. During this period Slovene writers began to reflect Western styles and trends, and publishing and other facilities were established to ensure the more secure development of a literature in the national language.

The emergence of a Slovene Modernist movement in

The Slovene Modernist movement

the early 20th century resulted in the second flowering of literature since the Romantics and Prešeren. The two main founders of this movement were Oton Župančič, a vigorous poet, and Ivan Cankar (*Hlapec Jernej in njegova pravica* [1907; *The Bailiff Yerney and His Rights*]; and *Hiša Marije pomočnice* [1904; *The Ward of Our Lady of Mercy*]), a masterful prose stylist. After World War I social and spiritual tension was evident in the predominantly Expressionist poetry of Tone Seliškar, Miran Jarc, and Anton Vodnik, as well as the plays of Slavko Grum. The novel and short story were revived during the 1930s, a decade marked by the appearance of Socialist Realism, especially in the works of Juš Kozak, Miško Kranjec, Ciril Kosmač, and Prežihov Voranc (*Samorastniki* [1940; *The Self-Sown*]), who turned to the everyday life of simple country people. The freer metres of Edvard Kocbek pointed the way to magically interpreted reality, and the caustic sonnets of Božo Vodušek manifested a sober disillusionment.

During and after World War II, writers who had been active before it remained rooted in the realistic tradition, although many deepened or broadened the scope of their themes. This is particularly true of the poet and dramatist Matej Bor (*Šel je popotnik skozi atomski vek* [1958; *A Wanderer in the Atomic Age*]), as well as the prose writers and dramatists Ivan Potrč and Mira Mihelič. Later developments in prose included, under Western influence, the appearance of science fiction by such writers as Miha Remec. Poetry has maintained its deeply subjective expression in the work of Cene Vipotnik, France Balantič, and especially Jože Udovič. This trend has been continued in the verses of the next generation, represented by Veno Taufer, Valentin Cundrič, and Dane Zajc. Slovene literature is identifiably part of mainstream European literature and is probably the most cosmopolitan of all the Yugoslav literatures.

MACEDONIAN LITERATURE

Like other Slavic literatures, literature in Macedonia began in the 9th century with the translations of religious works by SS. Cyril and Methodius and with the first Slavonic "university" that emerged at Ohrid (then in Bulgaria) under St. Clement (died 916), a disciple of SS. Cyril and Methodius. With this orthodox religious literature, there developed an apocryphal, heretic Bogomil literature. In the 16th century, with a fusion of Old Church Slavonic and vernacular languages, a popular Damascene literature developed: *Damaskini* were translations from the works of Damascenus the Studite. This lasted until the beginning of the 19th century, when books written in modern Macedonia were first published.

Early 19th-century Macedonian literature echoed the medieval religious enlightenment of the school of Kliment, and the only original contributions were from folk literature. Later, original lyrical poetry was written by Konstantin Miladinov, who, with his brother Dimitrije, compiled a notable collection of legends and folk songs contributing to the development of a national Macedonian literature.

The development of a Macedonian national consciousness, and of a Macedonian literature, was hampered in the 19th century by the continuing rule of the Ottoman Empire. The Balkan states finally made common cause against the empire in the Balkan Wars, 1912–13, only to fight among themselves over the division of Macedonia. Serbia received the lion's share of the northern part centred on Skopje and including Ohrid. After World War I the Macedonian language, most closely related to Bulgarian, was still considered by the government in Belgrade to be a southern dialect of Serbo-Croatian. The Macedonian language was not officially recognized until the establishment of Macedonia as a constituent republic in 1944. Despite these drawbacks some progress was made toward the foundation of a national language and literature, in particular by Kosta P. Misirkov in his *Za Makedonskite raboti* (1903; "In Favour of Macedonian Literary Works") and in the literary periodical *Vardar* (established 1905). These efforts were continued after World War I by Kosta Racin, who wrote mainly poetry in Macedonian and propagated its use through the literary journals of the 1930s. Some, such as Kole Nedelkovski, worked and published from abroad because of political pressure.

Official recognition of the Macedonian language

After World War II the Communist government of Yugoslavia officially recognized the Macedonian language, and encouragement was given to establish publishing facilities and provide educational services in the language. The scholar Blaže Koneski and others were charged with the task of standardizing Macedonian as the official literary language. With their new freedom to write and publish in their own language, and after years of struggle to defend a growing national consciousness and its expression as a separate cultural identity, the Macedonians produced many literary figures in the postwar period. Poetry has been represented in the work of Aco Šopov, Slavko Janevski, Blaže Koneski, and Gane Todorovski. Prewar playwrights, such as V. Iljoski, continued to write drama, and the theatre was invigorated by new dramatists, such as Kole Čašule and Tome Arsovski. Živko Čingo became one of Macedonia's best known writers of prose.

(V.Z./Em.Š./A.Gn./D.Še./D.A.N.)

BIBLIOGRAPHY

General: MILMAN PARRY and ALBERT B. LORD, *Serbocroatian Heroic Songs,* 2 vol. (1954); SVETOZAR KOLJEVIĆ, *The Epic in the Making* (1980), on the Serbian folk tradition, and *Yugoslav Short Stories* (1966); THOMAS BUTLER, *Monumenta Serbocroatica: A Bilingual Anthology of Serbian and Croatian Texts from the 12th to the 19th Century* (1980); ANTUN BARAC, *A History of Yugoslav Literature* (1955, reissued 1976); ANTE KADIĆ, *From Croatian Renaissance to Yugoslav Socialism* (1969); VASA D. MIHAILOVICH (ed.), *Modern Slavic Literatures,* vol. 2, *Bulgarian, Czechoslovak, Polish, Ukrainian and Yugoslav Literature* (1976), and *Contemporary Yugoslav Poetry* (1977). THOMAS EEKMAN, *Thirty Years of Yugoslav Literature: 1945–1975* (1978); SVETA LUKIĆ, *Contemporary Yugoslav Literature* (1972); BRANKO LENSKI (ed.), *Death of a Simple Giant, and Other Modern Yugoslav Stories* (1965); JANKO LAVRIN, *An Anthology of Modern Yugoslav Poetry in English Translation* (1962); BRANKO MIKASINOVICH, DRAGAN MILIVOJEVIĆ, and VASA D. MIHAILOVICH, *Introduction to Yugoslav Literature: An Anthology of Fiction and Poetry* (1973).

Serbia: ANNE PENNINGTON and PETER LEVI, *Marko the Prince* (1984); METEJA MATEJIC and DRAGAN MILIVOJEVIĆ, *An Anthology of Medieval Serbian Literature in English* (1978); GEORGE R. NOYES, *The Life and Adventures of Dimitrije Obradović* (1953); DUNCAN WILSON, *The Life and Times of Vuk Stefanović Karadzić* (1970); MIHAILO DORDEVIĆ, *Serbian Poetry and Milutin Bojić* (1977); CELIA HAWKESWORTH, *Ivo Andrić: Bridge Between East and West* (1984); MIODRAG PAVLOVIĆ, *Antologija srpskog pesništva* (1964); JOVAN SKERLIĆ, *Istorija nove srpske književnosti* (1967); JOVAN HRISTIĆ (ed.), *Srpska književnost u književnoj kritici,* 12 vol. (1965–66); ANTE KADIĆ, *Contemporary Serbian Literature* (1960); *Relations* (quarterly), translations of recent literature and criticism.

Croatia: JOSIP TORBARINA, *Italian Influence on the Poets of the Ragusan Republic* (1931); ELINOR M. DESPELATOVIĆ, *Ljudevit Gaj and the Illyrian Movement* (1975); EUGENE E. PANTZER, *Antun Gustav Matoš* (1981); ANTE KADIĆ, *Contemporary Croatian Literature* (1960); MIROSLAV BEKER (ed.), *Comparative Studies in Croatian Literature* (1981); FRANJO TROGRANČIĆ, *Storia della Letteraturea Croata* (1953); *Povijest hrvatske književnosti* (1978); NICOLA MILIČEVIĆ and ANTUN ŠOLJAN, *Antologija hrvatske poezije* (1966); MILIVOS SLAVIČEK (ed.), *A Collection of Modern Croatian Verse* (1965); *Most* (quarterly), translations of recent literature and criticism.

Slovenia: HENRY R. COOPER, *Francè Prešeren* (1981); ALOIS GRADNIK, *Selected Poems,* ed. by JANKO LAVRIN (1964); JANKO LAVRIN and ANTON SLODNJAK (eds.), *The Parnassus of a Small Nation,* 2nd enlarged ed. (1965); BRUNO MERIGGI, *Storia della letteratura slovena* (1961); ANTON SLODNJAK, *Geschichte der slowenischen Literatur* (1958); FRANCE SLOKAN, *Panorama littéraire slovène* (1972); JOŽE POGAČNIK, *Zgodovina slovenskega slovstva,* 8 vol. (1968–72); *Le livre slovene* (quarterly), translations of recent literature and criticism.

Macedonia: HORACE G. LUNT, *A Survey of Macedonian Literature* (1953); BLAŽE KONESKI, *Makedonska književnost* (1961); MIODRAG DRAGOVATS, *Contemporary Macedonian Writers* (1976); *Makedonskata književnost vo književnata kritika,* 5 vol. (1973–74); MILNE HOLTON (ed.), *The Big Horse and Other Stories of Modern Macedonia* (1974); MILNE HOLTON and GRAHAM W. REID (eds.), *Reading the Ashes: An Anthology of the Poetry of Modern Macedonia* (1977); *Macedonian Review,* English-language periodical published in Skopje, containing articles about Macedonia, including its literature.

(D.A.N.)

Yugoslavia

Yugoslavia (Socialist Federal Republic of Yugoslavia; Macedonian and Serbo-Croatian: Socijalistička Federativna Republika Jugoslavija; Slovene: Socijalistična Federativna Republika Jugoslavija), with an area of 98,766 square miles (255,804 square kilometres), is a state spanning parts of south central Europe and the Balkans. On its northwest–southeast axis, along the Dinaric Alps, it has a maximum length of 590 miles (945 kilometres), and it has a maximum width of about 450 miles from Žaga (Slovenia) in the east to Kostol (Serbia) in the west. The country borders Italy in the northwest, Austria and Hungary in the north, Romania in the northeast, Bulgaria in the east, and Greece and Albania in the south. To the west is the Adriatic Sea.

Yugoslavia's position places it at a point where the major cultural influences of Greece, Rome, and Byzantium meet and at the crossroads of the Roman Catholic and Eastern Orthodox churches and of Islām. Its multinational society, a confederation of South Slavs (Yugoslavs), bears the strong imprint of eastern and western Europe and Asia Minor in its various regions.

Modern Yugoslavia dates from 1918, when the Kingdom of Serbs, Croats, and Slovenes was proclaimed. In 1945, after the collapse of Nazi power in Europe, a federal people's republic was proclaimed, under the leadership of the Communist Partisan leader Josip Broz Tito.

Yugoslavia is a federal state comprising six republics that correspond to the main Slavic-speaking national groups. In order of population size they are Serbia (in which is located the federal capital, Belgrade [Beograd]), Croatia, Bosnia and Hercegovina, Macedonia, Slovenia, and Montenegro. Within Serbia are two autonomous provinces, Kosovo and Vojvodina.

This article is divided into the following sections:

Physical and human geography 1086
- The land 1086
  - Relief
  - Drainage
  - Climate
  - Plant and animal life
  - Settlement patterns
- The people 1090
  - Ethnicity, language, and religion
  - Demographic trends
- The economy 1091
  - Resources
  - Agriculture
  - Industry
  - Trade
  - Transportation
- Administration and social conditions 1093
  - Government
  - Education
  - Health and welfare
- Cultural life 1094

History 1095
- The prehistoric period 1095
- The Roman Empire 1096
- The barbarian invasions and the fall of Rome 1096
- Byzantium and the early Slav states 1096
- The Ottoman conquest 1099
- Society and economy: the 14th to the 18th century 1101
- The passing of the old order 1104
- Classes and states: 1830–1912 1105
- The Balkan Wars 1107
- World War I 1108
- The Kingdom of Serbs, Croats, and Slovenes 1109
- World War II 1111
- The second Yugoslavia 1113

Bibliography 1115

Physical and human geography

THE LAND

Relief. Yugoslavia is a predominantly hilly country. More than half of the land area is between 600 and 3,000 feet in elevation, and almost one-fifth is higher than 3,000 feet. The country lies in a zone of structural geologic weakness, making it vulnerable to earthquakes.

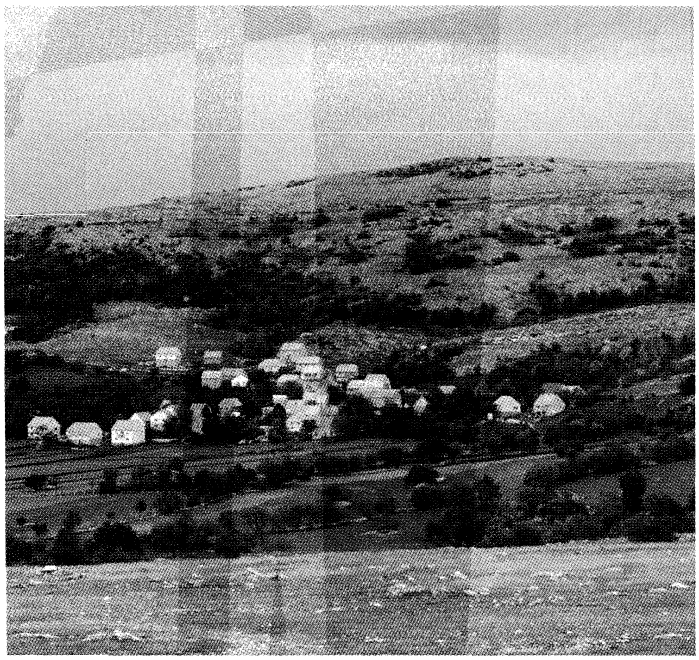
With the exception of the mountain ranges in the east and south, the main trend of the hills and mountains is northwest–southeast, with ridges and valleys running parallel to the Adriatic coast. In the northwest are the highest peaks of the Julian Alps, part of the main European Alpine system. They rise to 9,396 feet in Triglav, the highest mountain in Yugoslavia, and form the northern boundary of the country as far east as Maribor. To the south of the Julian Alps lie the Dinaric Alps. With elevations of more than 6,500 feet in some areas, the Dinaric Alps lie between the Adriatic and the Sava valley, extending south to Montenegro. This is a region consisting mainly of limestone plateaus, having features such as potholes and underground drainage, dry valleys, and extensive cave systems caused by the dissolving of the rock by water. The name for the area, Kras (Karst), has been adopted as the general term for any limestone area displaying such features. The uplands are often barren and denuded, the result of deforestation and thin soils, but the depressions between the ridges, called poljes, offer the best opportunities for agriculture. Poljes range in size from a few square yards to several hundred square miles. East of the Dinaric Alps and extending southward is a complex zone of different rocks forming the hills of Bosnia and Montenegro. Many are resistant to erosion, resulting in broken, rugged plateaus deeply dissected by rivers. The elevations range from 3,000 feet in the west to 6,000 to 8,000 feet in the south, close to the Albanian border. Farther east, in Serbia, the Carpathian and the Balkan mountains extend from the 70-mile stretch of four gorges on the Danube, which includes the Iron Gate, in the north to the valley of the Nišava River in the south. These mountains straddle the border with Romania and Bulgaria. In the south, Macedonia is also predominantly an upland area where the Šar Mountains rise to more than 9,000 feet.

The lowlands of the country lie mostly in the north and northeast, covering about a quarter of the area. The largest lowland area is the Pannonian Basin, which extends along the Danube and its major tributaries, the Sava, Drava, and Tisa. This floodplain is now protected by dikes, with gravel and loess terraces forming low hills. Rising beyond them are ancient outcrops of hard, crystalline rocks such as the Fruška Hills (1,765 feet) and Psunj Mountains (3,232 feet). There are narrow riverine lowlands in the valleys of the Morava and Vardar rivers. The latter valley widens toward the border with Greece in the south.

The Adriatic coastal zone is an extension of the parallel linear ridges and valleys of the Kras interior but has been submerged by geologically recent changes in sea level to form a series of long, narrow islands. The coastal area is known as Dalmatia.

Influence of the topography

The topography of the country has produced major problems. Mountains have been important in protecting the country from the northwest and the south and east, and in the past they were the strongholds of South Slav cultures. In modern times the mountains have limited access and made the exploitation of natural resources more difficult. Access to the sea is also a problem, for there are few



Small polje in the Kras region, near Livno, western Bosnia and Hercegovina.
Jerome Wyckoff

natural routeways from the interior to the Adriatic coast and none that are easy. The orientation of the country therefore has tended to be toward the interior rather than the Mediterranean. However, Yugoslavia lies across the historic routes from the eastern Mediterranean basin and Black Sea to the continental interior, and the valleys of the Vardar in the southeast and the Morava and Nišava in the east, in particular, have had an important role in channeling people and goods toward the Danube and the west and in providing routeways for modern communications by road and rail.

Drainage. The Danube runs through northeastern Yugoslavia for about 225 miles. It forms a part of Yugoslavia's border with Romania for another 142 miles. With its tributaries it drains some 69,196 square miles, about 70 percent of Yugoslavia's total area, to the Black Sea. Major northern tributaries of the Danube are the Mura, the Drava, which lies on the boundary with Hungary, and the Tisa. The longest river in Yugoslavia is the Sava (587 miles), which flows across the Pannonian Basin to its confluence with the Danube at Belgrade. It also drains a large part of southern Yugoslavia through its north-flowing tributaries, the Kupa, Una, Vrbas, Bosna, and Drina, which form a broken gridlike pattern superimposed on, and sometimes breaking through, the mountain ranges and valleys. In the east the major tributary of the Danube is the Morava. It forms a narrow watershed with the Vardar, which drains through Macedonia to the Aegean Sea. Only about 20 percent of the country drains to the Adriatic, and in the coastal region there are few rivers of any length, the longest being the Neretva (140 miles). Intermittent drainage, with disappearing rivers and streams and dry valleys, is a feature of the Kras. Many rivers follow turbulent courses through narrow gorges, making them less valuable for communications.

There are about 250 natural lakes in Yugoslavia. In the Alpine areas some occupy valleys dammed by glacial moraines. Others fill depressions left by glaciers and are known as "the eyes of the mountains" because of their clear blue waters. Another group fills hollows and depressions near the coast and includes the country's largest lake, Scutari (Skadarsko Jezero; 151 square miles), part of which is in Albania. Two other major lakes are on the southern border: Lake Ohrid, Yugoslavia's deepest lake, and Lake Prespa. The 16 lakes of the Plitvička system in Croatia are a major tourist attraction.

Climate. Although Yugoslavia occupies a southerly European location, the principal climatic influence of the country is the mountain ranges of the Adriatic coast, which limit the ameliorating influence of the Mediterranean Sea. A Mediterranean climatic regime is restricted to the narrow coastal strip and islands and to the Istria Peninsula in the north. The coastal region has mild and wet winters (January average 45° F [7° C]) and hot, dry summers (July average 75° F [24° C]) with cloudless skies. The northeastern plains of the Danube basin are open to the continental influence of eastern and central Europe and enjoy a moderate continental pattern of fairly cold winters. The January average is 32° F (0° C), but temperatures may drop to 14° F (−10° C) under the influence of Arctic airstreams. Summer temperatures range from 68° to 72° F (20° to 22° C), although they sometimes reach 86° F (30° C) in southern Yugoslavia. Temperatures are generally lower in the hills and mountains.

Annual precipitation averages 27 inches (680 millimetres) in Belgrade, falling mostly in late spring and early summer. Near the coast rainfall averages between 30 and 45 inches annually. Totals increase, however, in the Julian Alps and in the Velika Kapela and Velebit mountains to nearly 85 inches, with heavy mountain snowfall. Annual rainfall is about 30 inches in the northeastern plains. Crkvice, on the slopes of Mount Orjen in Montenegro, has Europe's highest recorded average annual precipitation (183 inches). Thunderstorms are frequent in summer, especially in the mountains.

Local winds

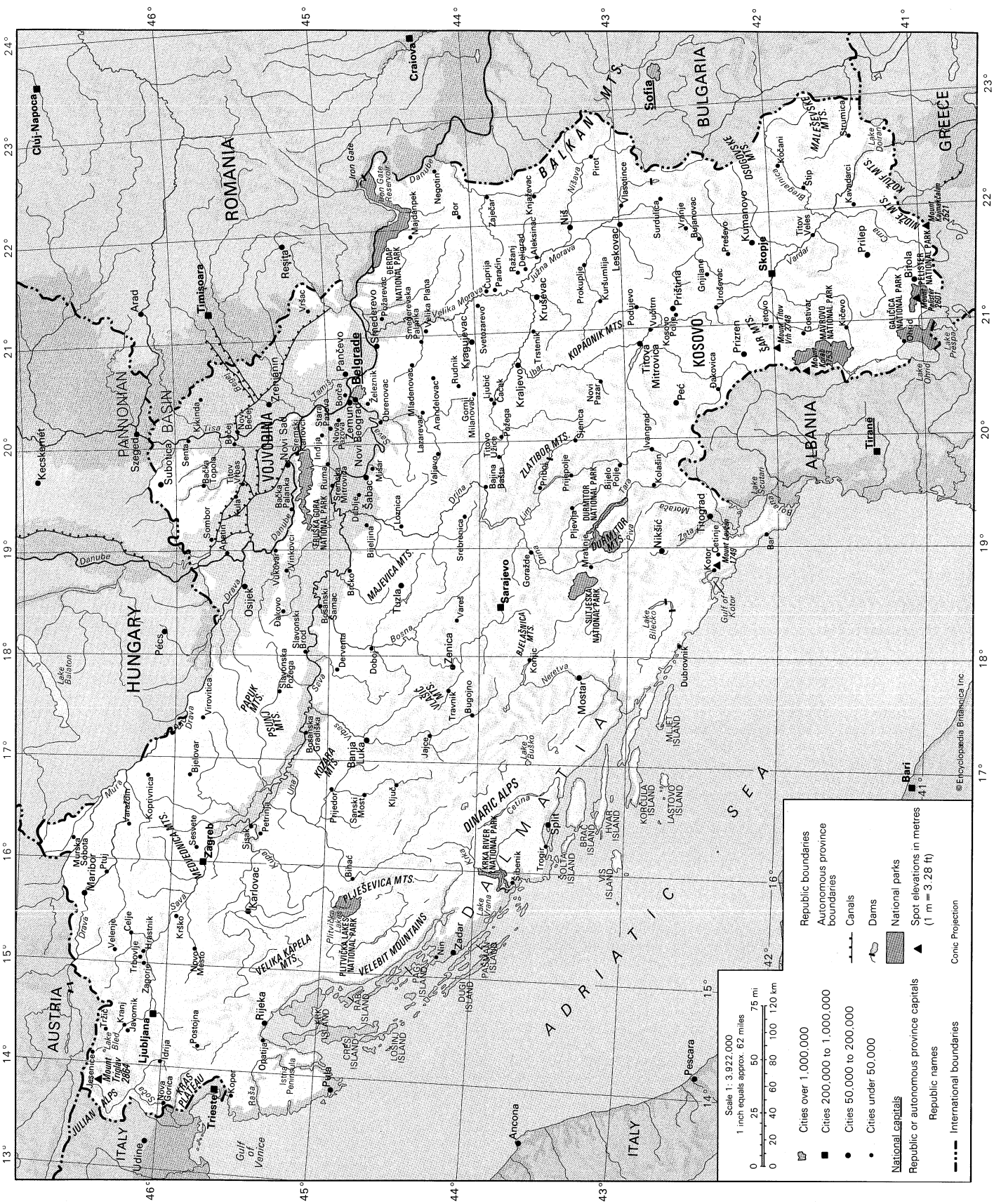
The cold *bura* wind blows from the interior to the warmer sea in winter, as does the *vardarac* in Macedonia and the *košava* in the northeast. The *jugo,* which blows north up the Adriatic, brings rainfall to the coastal zone, and in summer the *maestral* ameliorates temperatures along the coastal strip during the middle of the day.

Plant and animal life. There are more than 5,000 plant species in Yugoslavia, many of them native. More than a third of the country is covered by forests, three-quarters of which are deciduous. Evergreen species dominate the Mediterranean coastal regions, although much of the area has limited cover of maquis (such as laurel and myrtle). Subtropical plants such as cotton, rice, and poppies are cultivated in Macedonia. The major deciduous species covering most of the rest of the country are beech at higher altitudes (3,300 to 5,200 feet) and oak and hornbeam at lower altitudes (2,000 to 3,300 feet). Conifers such as mountain pine, firs, and spruce are found as high as 6,000 feet in certain areas. In the mountains the forest alternates with pastures with excellent summer grazing, while in the Kras they give way to bare, open country.

Animal life in the interior plains is characterized by central and eastern European influences. In the mountains larger European mammals such as deer, wild pig, wolf, fox, bear, lynx, and wildcat survive. Chamois and alpine rabbits live in the highest alpine areas.

Settlement patterns. Many factors have affected the impact of people on the territory of Yugoslavia; the most important of these is topography, which, together with the cultural diversity of Yugoslav peoples and their differing historical experience, has tended to produce different forms of settlement. Isolated uplands and mountains have remained relatively thinly settled, with only small villages and few towns. The plains of Slovenia, Croatia, and Serbia have enjoyed more vigorous agricultural growth, and from medieval times a hierarchy of small and medium-sized commercial, religious, and political centres developed in these regions. Under the Roman Empire favoured locations along the coast were able to exploit their positions, and they grew into large and powerful city-states. The Roman influence was also strong in the interior and led to the development of many settlements at crucial points on the overland trade routes of the empire. Under Ottoman rule towns such as Skopje and Priština grew in importance, becoming military garrisons and civil administrative centres. Belgrade, with its fortress Kalemegdan dominating the confluence of the Danube and Sava, has a strategic position and was fought over and held by every major group in the area, including the Romans, Huns, Byzantines, Avars, Slavs, Bulgars, Magyars, Turks, and Serbs.

Since World War II there has been rapid growth in the



MAP INDEX

Political subdivisions

Bosnia and Hercegovina . . . . 44 15 N 17 50 E
Crna Gora, see Montenegro
Croatia (Hrvatska) . . . . . 45 10 N 15 30 E
Kosovo . . . . . 42 35 N 21 00 E
Macedonia (Makedonija) . . . . 41 50 N 22 00 E
Montenegro (Crna Gora) . . . . . 42 30 N 19 18 E
Serbia (Srbija) . . . . . 44 00 N 21 00 E
Slovenia (Slovenija) . . . . . 46 15 N 15 10 E
Srbija, see Serbia
Vojvodina . . . . . 45 00 N 20 00 E

Cities and towns

Aleksinac . . . . . 43 32 N 21 43 E
Apatin . . . . . 45 40 N 18 59 E
Aranđelovac . . . . . 44 18 N 20 35 E
Bačka Palanka . . . 45 15 N 19 22 E
Bačka Topola . . . . 45 49 N 19 39 E
Bajina Bašta . . . . . 43 58 N 19 34 E

Banja Luka ..... 44 46 N 17 10 E
Bar ..... 42 05 N 19 06 E
Bečej ..... 45 37 N 20 03 E
Belgrade (Beograd) ..... 44 50 N 20 30 E
Bihać ..... 44 49 N 15 52 E
Bijeljina ..... 44 45 N 19 13 E
Bijelo Polje ..... 43 02 N 19 45 E
Bitola ..... 41 02 N 21 20 E
Bjelovar ..... 45 54 N 16 51 E
Bor ..... 44 06 N 22 06 E
Borča ..... 44 52 N 20 28 E
Bosanska Gradiška ..... 45 09 N 17 15 E
Bosanski Šamac .. 45 04 N 18 28 E
Brčko ..... 44 52 N 18 49 E
Bugojno ..... 44 03 N 17 27 E
Bujanovac ..... 42 28 N 21 47 E
Čačak ..... 43 54 N 20 21 E
Celje ..... 46 14 N 15 16 E
Cetinje ..... 42 24 N 18 55 E
Ćuprija ..... 43 56 N 21 22 E
Đakovica ..... 42 23 N 20 26 E
Đakovo ..... 45 19 N 18 25 E
Deligrad ..... 43 37 N 21 35 E
Derventa ..... 44 59 N 17 55 E
Doboj ..... 44 44 N 18 05 E
Dublje ..... 44 48 N 19 30 E
Dubrovnik ..... 42 39 N 18 07 E
Gnjilane ..... 42 28 N 21 29 E
Goražde ..... 43 40 N 18 59 E
Gornji Milanovac .. 44 02 N 20 27 E
Gostivar ..... 41 48 N 20 54 E
Hrastnik ..... 46 09 N 15 06 E
Idrija ..... 46 00 N 14 02 E
Inđija ..... 45 03 N 20 05 E
Ivangrad ..... 42 51 N 19 52 E
Jajce ..... 44 21 N 17 17 E
Javornik ..... 46 14 N 14 18 E
Jesenice ..... 46 27 N 14 04 E
Karlovac ..... 45 29 N 15 33 E
Kavadarci ..... 41 26 N 22 01 E
Kičevo ..... 41 31 N 20 58 E
Kikinda ..... 45 50 N 20 29 E
Ključ ..... 44 32 N 16 47 E
Knjaževac ..... 43 34 N 22 15 E
Kočani ..... 41 55 N 22 25 E
Kolašin ..... 42 49 N 19 32 E
Konjic ..... 43 39 N 17 58 E
Koper ..... 45 33 N 13 44 E
Koprivnica ..... 46 10 N 16 50 E
Kosovo Polje ..... 42 38 N 21 07 E
Kotor ..... 42 25 N 18 46 E
Kragujevac ..... 44 01 N 20 55 E
Kraljevo ..... 43 44 N 20 43 E
Kranj ..... 46 14 N 14 22 E
Krško ..... 45 58 N 15 29 E
Kruševac ..... 43 35 N 21 20 E
Kula ..... 45 37 N 19 32 E
Kumanovo ..... 42 08 N 21 43 E
Kuršumlija ..... 43 09 N 21 16 E
Lazarevac ..... 44 23 N 20 16 E
Leskovac ..... 42 59 N 21 57 E
Ljubić ..... 43 54 N 20 22 E
Ljubljana ..... 46 02 N 14 30 E
Loznica ..... 44 32 N 19 13 E
Majdanpek ..... 44 25 N 21 56 E
Maribor ..... 46 33 N 15 39 E
Mišar ..... 44 44 N 19 44 E
Mladenovac ..... 44 26 N 20 42 E
Mostar ..... 43 21 N 17 49 E
Mratinje ..... 43 16 N 18 49 E
Murska Sobota ... 46 40 N 16 10 E
Negotin ..... 44 13 N 22 32 E
Nikšić ..... 42 46 N 18 58 E
Nin ..... 44 14 N 15 11 E
Niš ..... 43 19 N 21 54 E
Nova Gorica ..... 45 57 N 13 39 E
Nova Pazova ..... 44 56 N 20 14 E
Novi Bečej ..... 45 36 N 20 08 E
Novi Beograd .... 44 49 N 20 27 E
Novi Pazar ..... 43 08 N 20 31 E
Novi Sad ..... 45 15 N 19 50 E
Novo Mesto ..... 45 48 N 15 10 E
Obrenovac ..... 44 39 N 20 12 E
Ohrid ..... 41 07 N 20 48 E
Opatija ..... 45 20 N 14 19 E
Osijek ..... 45 33 N 18 42 E
Pančevo ..... 44 52 N 20 39 E
Paraćin ..... 43 52 N 21 25 E
Peć ..... 42 39 N 20 18 E
Petrinja ..... 45 27 N 16 17 E
Pirot ..... 43 09 N 22 36 E
Pljevlja ..... 43 21 N 19 21 E
Podgorica, see Titograd
Podujevo ..... 42 55 N 21 12 E
Postojna ..... 45 47 N 14 14 E
Požarevac ..... 44 37 N 21 12 E
Požega ..... 43 51 N 20 02 E
Prešević ..... 42 19 N 21 39 E
Priboj ..... 43 35 N 19 32 E
Prijedor ..... 44 59 N 16 42 E
Prijepolje ..... 43 24 N 19 39 E
Prilep ..... 41 21 N 21 34 E
Priština ..... 42 40 N 21 10 E
Prizren ..... 42 13 N 20 45 E
Prokuplje ..... 43 15 N 21 36 E
Ptuj ..... 46 25 N 15 52 E
Pula ..... 44 52 N 13 50 E
Ražanj ..... 43 40 N 21 33 E
Rijeka ..... 45 21 N 14 24 E
Rudnik ..... 44 08 N 20 31 E
Ruma ..... 45 01 N 19 49 E
Šabac ..... 44 45 N 19 43 E
Sanski Most ..... 44 46 N 16 40 E
Sarajevo ..... 43 50 N 18 25 E
Senta ..... 45 56 N 20 05 E
Sesvete ..... 45 50 N 16 07 E
Šibenik ..... 43 44 N 15 53 E
Sisak ..... 45 29 N 16 22 E
Sjenica ..... 43 16 N 20 00 E
Skopje (Skoplje) .. 42 00 N 21 29 E
Slavonska Požega ..... 45 20 N 17 41 E
Slavonski Brod ... 45 09 N 18 02 E
Smederevo ..... 44 39 N 20 56 E
Smederevska Palanka ..... 44 22 N 20 58 E
Sombor ..... 45 46 N 19 07 E
Split ..... 43 31 N 16 26 E
Srebrenica ..... 44 06 N 19 18 E
Sremska Mitrovica ..... 44 58 N 19 37 E
Sremski Karlovci .. 45 12 N 19 56 E
Stara Pazova ..... 44 59 N 20 10 E
Štip ..... 41 44 N 22 12 E
Strumica ..... 41 26 N 22 39 E
Subotica ..... 46 06 N 19 40 E
Surdulica ..... 42 41 N 22 10 E
Svetozarevo ..... 43 59 N 21 15 E
Tetovo ..... 42 01 N 20 59 E
Titograd (Podgorica) ..... 42 26 N 19 16 E
Titov Veles ..... 41 42 N 21 48 E
Titov Vrbas ..... 45 34 N 19 39 E
Titova Mitrovica ... 42 53 N 20 52 E
Titovo Užice ..... 43 52 N 19 51 E
Travnik ..... 44 14 N 17 40 E
Trbovlje ..... 46 10 N 15 03 E
Trogir ..... 43 32 N 16 15 E
Trstenik ..... 43 37 N 21 00 E
Tržič ..... 46 22 N 14 19 E
Tuzla ..... 44 33 N 18 41 E
Uroševac ..... 42 22 N 21 10 E
Valjevo ..... 44 16 N 19 53 E
Varaždin ..... 46 18 N 16 20 E
Vareš ..... 44 10 N 18 20 E
Velenje ..... 46 22 N 15 07 E
Velika Plana ..... 44 20 N 21 05 E
Vinkovci ..... 45 17 N 18 49 E
Virovitica ..... 45 50 N 17 23 E
Vlasotince ..... 42 58 N 22 08 E
Vranje ..... 42 33 N 21 54 E
Vršac ..... 45 07 N 21 18 E
Vučitrn ..... 42 49 N 20 58 E
Vukovar ..... 45 21 N 19 00 E
Zadar ..... 44 07 N 15 15 E
Zagorje ..... 46 08 N 15 00 E
Zagreb ..... 45 48 N 16 00 E
Zaječar ..... 43 54 N 22 17 E
Železnik ..... 44 43 N 20 23 E
Zemun ..... 44 50 N 20 24 E
Zenica ..... 44 13 N 17 55 E
Zrenjanin ..... 45 23 N 20 23 E

Physical features and points of interest

Adriatic Sea ..... 43 00 N 16 00 E
Balkan (Stara) Mountains ..... 43 15 N 25 00 E
Begej, *river* ..... 45 13 N 20 19 E
Bilećko, Lake, *reservoir* ..... 42 50 N 18 26 E
Bjelašnica Mountains ..... 43 09 N 18 23 E
Bled (Bledsko), Lake ..... 46 22 N 14 06 E
Bojana, *river* ..... 41 52 N 19 22 E
Bosna, *river* ..... 45 04 N 18 28 E
Brač Island ..... 43 19 N 16 40 E
Bregalnica, *river* .. 41 36 N 21 56 E
Buško, Lake ..... 43 40 N 17 02 E
Cres Island ..... 44 40 N 14 25 E
Crna, *river* ..... 41 33 N 21 59 E
Dalmatia (Dalmacija), *region* ..... 43 00 N 17 00 E
Danube (Dunav), *river* ..... 45 20 N 19 00 E
Đerdap, see Iron Gate Reservoir
Dinaric (Dinara) Alps ..... 43 50 N 16 35 E
Djerdap National Park ..... 44 30 N 22 10 E
Doiran (Dojransko), Lake ..... 41 13 N 22 44 E
Drava, *river* ..... 45 33 N 18 55 E
Drina, *river* ..... 44 53 N 19 21 E
Dugi Island ..... 44 00 N 15 00 E
Dunav, see Danube
Durmitor Mountains ..... 43 09 N 19 02 E
Durmitor National Park ..... 43 06 N 19 01 E
Fruška Gora National Park .... 45 09 N 19 39 E
Galičica National Park ..... 41 00 N 20 51 E
Gvozdena, see Iron Gate
Hvar Island ..... 43 07 N 16 45 E
Ibar, *river* ..... 43 44 N 20 45 E
Iron (Gvozdena) Gate, *gorge* .... 44 41 N 22 31 E
Iron Gate (Đerdap) Reservoir ..... 44 30 N 22 10 E
Istria (Istra) Peninsula ..... 45 00 N 14 00 E
Julian (Julijske) Alps ..... 46 20 N 13 45 E
Južna Morava, *river* ..... 43 41 N 21 24 E
Kajmakčalan, Mount ..... 40 58 N 21 48 E
Kopaonik Mountains ..... 43 15 N 20 50 E
Korab, Mount .... 41 46 N 20 32 E
Korčula Island .... 42 57 N 16 55 E
Kotor (Kotorska), Gulf of ..... 42 25 N 18 40 E
Kozara Mountains ..... 45 00 N 16 55 E
Kožuf (Kóziouf) Mountains ..... 41 09 N 22 10 E
Kras (Karst) Plateau ..... 45 48 N 14 00 E
Krk Island ..... 45 05 N 14 35 E
Krka, *river* ..... 43 43 N 15 51 E
Krka River National Park .... 43 50 N 15 55 E
Kupa, *river* ..... 45 28 N 16 24 E
Lastovo Island .... 42 45 N 16 50 E
Lim, *river* ..... 43 45 N 19 13 E
Lošinj Island ..... 44 35 N 14 28 E
Lovćen, Mount ... 42 24 N 18 49 E
Majevica Mountains ..... 44 36 N 18 45 E
Maleševske Mountains ..... 41 40 N 23 00 E
Mavrovo National Park ..... 41 45 N 20 40 E
Medvednica Mountains ..... 45 55 N 15 58 E
Mljet Island ..... 42 45 N 17 30 E
Morača, *river* ..... 42 16 N 19 09 E
Morava, see Velika Morava
Mura, *river* ..... 46 18 N 16 55 E
Neretva, *river* .... 43 02 N 17 27 E
Nidže Mountain ... 41 00 N 21 50 E
Nišava, *river* ..... 43 22 N 21 46 E
Ohrid (Ohridsko), Lake ..... 41 00 N 20 45 E
Osogovske Mountains ..... 42 10 N 22 30 E
Pag Island ..... 44 30 N 15 00 E
Pannonian Basin, *region* ..... 46 45 N 19 00 E
Papuk Mountains .45 31 N 17 39 E
Pašman Island .... 43 57 N 15 21 E
Pelister, Mount ... 41 00 N 21 11 E
Pelister National Park ..... 41 03 N 21 12 E
Piva, *river* ..... 43 21 N 18 51 E
Plitvička (Plitvice) Lakes ..... 44 53 N 15 36 E
Plitvička Lakes National Park .... 44 51 N 15 37 E
Plješevica Mountains ..... 44 45 N 15 45 E
Prespa (Prespansko), Lake ..... 40 55 N 21 00 E
Psunj Mountains .. 45 24 N 17 20 E
Rab Island ..... 44 46 N 14 46 E
Raša, *river* ..... 45 03 N 14 03 E
Šar Mountains .... 42 05 N 20 50 E
Sava, *river* ..... 44 50 N 20 28 E
Scutari (Skadarsko), Lake ..... 42 10 N 19 20 E
Soča, *river* ..... 45 43 N 13 33 E
Šolta Island ..... 43 23 N 16 17 E
Stara, see Balkan Mountains
Sutjeska National Park ..... 43 18 N 18 45 E
Tamiš, *river* ..... 44 51 N 20 39 E
Tara, *river* ..... 43 21 N 18 51 E
Tisa, *river* ..... 45 15 N 20 17 E
Titov Vrh, Mount .. 41 59 N 20 48 E
Triglav, Mount .... 46 23 N 13 50 E
Una, *river* ..... 45 16 N 16 55 E
Vardar, *river* ..... 40 35 N 22 50 E
Velebit Mountains ..... 44 38 N 15 03 E
Velika Kapela Mountains ..... 45 13 N 15 02 E
Velika Morava (Morava), *river* ... 44 43 N 21 03 E
Venice, Gulf of ... 45 15 N 13 00 E
Vis Island ..... 43 02 N 16 10 E
Vlašić Mountains .. 44 19 N 17 40 E
Vrana (Vransko), Lake ..... 43 53 N 15 34 E
Vrbas, *river* ..... 45 07 N 17 31 E
Zeta, *river* ..... 42 28 N 19 16 E
Zlatibor Mountains ..... 43 40 N 19 43 E

Rural migration and urbanization

number of people living in towns and cities. About a third of the population lives in towns of more than 20,000 people. Migration from rural to urban areas has been one of the most significant features of postwar Yugoslavia's social and economic life. Major towns such as Skopje have drawn populations from agricultural regions extending outside their own spheres of influence, and much internal migration has been directly to Belgrade. Industrial development based on the exploitation of minerals, including metal ores, has led to the rapid urbanization of Sarajevo and Zenica in the Bosna River valley and of Tuzla and Banja Luka in the north. In Hercegovina, Mostar has grown rapidly. The greatest concentrations of urban population continue to be around the edges of the mountain heartland, in the richer agricultural valleys and lowlands, and especially along the Sava valley from Ljubljana in

the northwest to Zagreb and Belgrade farther east. In the postwar period this historically important axis has also been extended southward to Niš and Skopje.

The importance of agriculture is reflected in rural settlement. In the prosperous north are located the large farmhouses and modern mechanized cooperative farms of the Vojvodina, which contrast with the dry-stone or wooden houses scattered around meagre water supplies on the unproductive soils of Dalmatia, Montenegro, and western Bosnia. The historic ports and fishing villages of the Adriatic have experienced a rapid development of apartment buildings and tourist hotels, testimony to the increasing importance of tourism in Yugoslavia's economy.

THE PEOPLE

Ethnicity, language, and religion. The complex history of the Balkan Peninsula has left Yugoslavia with a population of great ethnic diversity. Official statistics record 25 specific ethnic groups, and Yugoslavia's federal structure is largely designed to accommodate this diversity. About 80 percent of the population is descended from the various Slav tribes that migrated into the region between the 6th and 8th centuries AD. The Albanians, generally regarded as descendants of the ancient Illyrians, are the largest non-Slav ethnic group. They represent earlier settlement, as do the Vlachs. Other minority groups are drawn from more recent migrations and include Magyars, Turks, Gypsies, Slovaks, Romanians, and Bulgarians.

The Slovenes in the northwest have their own language. Inhabiting a region that was formerly part of the Austrian lands of the Dual Monarchy, they defended their national identity against pressure to Germanize their culture and have continued to participate in a broader stream of Alpine culture. The Croats inhabit a broad belt across the northern part of the country, extending from the Istria Peninsula to the plains between the Drava, Danube, and Sava rivers. In the south they inhabit the southeastern side of the Dinaric Alps and Dalmatia. There are also substantial areas of Croatian settlement in central Bosnia. Like the Slovenes, Croats use the Roman alphabet, they have a dominant Roman Catholic religious tradition, and their sense of nationality has been shaped in part by the experience of Austro-Hungarian domination.

The Serbs

The largest ethnic group is the Serbs. They predominate in the Republic of Serbia, in western Bosnia, and in southern Hercegovina. Converted to Christianity under the Greek Orthodox tradition, Serbs write their language in the Cyrillic alphabet, and after centuries of subordination to the Ottoman Empire they have adopted many Turkish loanwords. Serbs and Croats are generally recognized as speaking the same language, Serbo-Croatian, in spite of the differences in orthography, pronunciation, and vocabulary that have resulted from the separate historical experiences of the two peoples.

In the north the province of Vojvodina includes a large Magyar-speaking minority among its many ethnic groups. In the south the province of Kosovo has become almost entirely Albanophone. Although historically some Albanians have adopted each of the major religious traditions of the area, those in Yugoslavia are mainly Muslim. The large size and rapid rate of growth of this group has become the focus of political conflict. Albanians also make up significant minorities in Montenegro and Macedonia.

The Montenegrins are culturally very close to the Serbs but owe their separate status to their success in having retained a large measure of independence when the rest of the peninsula was under Turkish rule. Like the Serbs, Macedonians follow the Orthodox tradition and use the Cyrillic alphabet, but their language, which has only been standardized in the postwar years, is most closely related to Bulgarian.

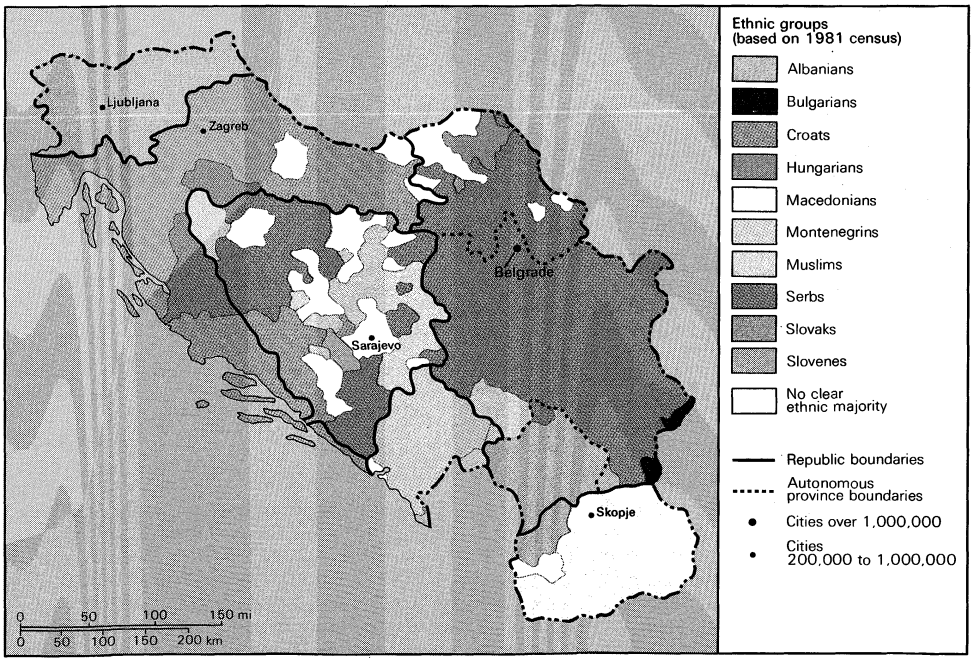
In the Republic of Bosnia and Hercegovina, Muslims make up 40 percent of the population. They are Serbo-Croatian-speaking descendants of the indigenous Slav population who converted to Islām at the time of the Ottoman invasions of the 16th century.

Since World War II, largely in Serbo-Croatian-speaking areas, there has been the gradual emergence of a sizable section of the population who prefer to describe themselves as "Yugoslavs." Although these people make up only a small percentage of the population, their numbers are growing steadily, more as a result of ethnically mixed marriages than because of high rates of natural increase.

Within former Habsburg parts of Yugoslavia, language is the most significant focus of national identity, with the prevailing Roman Catholic religion taking second place. In the former Ottoman regions, where the ruling stratum was distinguished by adherence to Islām, religion tends to be the more salient factor.

Religion

Rates of religious practice are low throughout the country. Only about 10 percent of the population has an active commitment to religion. The percentage is higher among the older rural inhabitants, who are oriented toward a more traditional set of values, and in Slovenia. After a period of open official hostility immediately after World War II, religious toleration became the rule. There have been signs of a new interest in religion among the young, although this often is linked to the revival of nationalism.



Ethnic composition of Yugoslavia.

This development has fostered fears of a return to the intercommunal violence of the wartime period and has led to a souring of relations between church and state.

Demographic trends. The population of Yugoslavia has a rate of natural increase that is among the highest in Europe, with relatively high birth and low death rates in rural areas. The overall birth rate has, however, declined in recent years, although the infant mortality rate has remained high. Both birth and death rates vary widely among the republics and autonomous provinces. Kosovo, with a younger population and also a large Albanian population, has a significantly higher birth rate but a low death rate. Vojvodina, Serbia, Croatia, and Slovenia all have birth rates that are below the national average.

Several million Yugoslavs and their descendants live permanently abroad, and more than 750,000 short-term migrant workers, including their families, live in other European countries. Another 110,000 Yugoslavs live in Australia, the Unites States, and Canada. The remittances of these migrant workers have been a valuable source of convertible currency for the national economy and have helped to raise the standard of living for those families left behind. At the same time, migration has absorbed surplus domestic labour and has helped reduce unemployment. The position of migrant Yugoslav workers has been precarious in countries that have experienced reduced levels of economic growth, and large numbers of workers returned during the 1970s and '80s.

THE ECONOMY

The national economy of Yugoslavia has undergone a remarkable transformation since 1945, within a distinctive socialist ideological framework and under the special difficulties resulting from the country's diverse cultures, history of imperial dominance, contrasting regional levels of economic development, and losses in World War II. Economic recovery was rapid in the postwar period, assisted by loans from both East and West. It was particularly high between 1953 and 1965 but slowed dramatically in the 1980s. The social and cooperative sector of the economy has provided an increasing proportion of national income, and industrial growth has been a key to economic development, especially in less developed areas such as Kosovo and Montenegro. Overall, however, productivity remains low and unemployment high.

Since 1945 the Yugoslavs have developed a number of distinctive features in their economic system. In the immediate postwar period the basic means of production, the financial institutions, and the means of communication were all nationalized. The system, which has come to be known as "self-management," is generally believed to have been the creation of Yugoslavia's leading ideologist, Edvard Kardelj. It reached its most developed form in the Law on Associated Labour of 1976, under which the means of production and other major resources are not regarded as state property (as in the Soviet Union) but as social property.

Self-management

Individuals participate through membership in the work organizations in which they are employed. Each organization is governed by a workers' council, which in turn elects a board of management. If the organization is very small, these functions are exercised by the whole work force. The director and other senior management officials of an organization are ex officio members of the board, though they have no vote and can be removed by the initiative of the council. In practice, highly trained experts with access to information, time, and other resources not easily available to ordinary workers are at a definite advantage when it comes to making business decisions. Consequently, the conduct of many Yugoslav enterprises remains as securely in the hands of their management as in any other system.

Although self-management is nominally a uniform and rather rigidly prescribed system, the extent to which it actually realizes the aspirations of its designers to institute genuine "workers' control" and to overcome the "alienation of labour power" varies enormously among enterprises, depending on such factors as the size, experience, and level of education of the work force and the character of the technical and commercial operations in question.

Work organizations may be either "basic organizations of associated labour" (constituted on the lines described above) or "complex organizations of associated labour," which are typical of larger enterprises. The latter involve the federation of more than one basic organization, each of which is concerned with a segment of overall activity. (Manufacture and distribution, for example, may be assigned to different basic organizations.) Shared functions, such as accounting and research and development, may be allocated to a collective services division.

Self-management is often assumed to be a feature specific to industrial or business life, but it is a form of sociopolitical organization that functions throughout administration, the social services, and culture. Thus banks, hospitals, schools, and municipal services are all run along self-managing lines.

Associated with the constitutional reform of 1963 was a series of economic measures that has had an equally important impact on the economic structure of Yugoslavia. These changes have been described in terms of market socialism. They mirror at the economic level the constitutional move toward greater regional and local autonomy, accelerating the shift away from centralized planning, giving greater autonomy to enterprises, and giving banks a larger say in the allocation of investment. Changes in the foreign trade system removed protective barriers and exposed Yugoslav firms more fully to international competition. In 1967 for the first time foreign capital could be invested in Yugoslav firms, with a maximum holding of 49 percent by foreign investors. These changes were intended to give a greater role to the market in determining the direction and pattern of development, while retaining the economy's socialist institutional framework. Efficiency was encouraged by promoting competition between enterprises and by using the criterion of profitability to direct administrative and political decisions about investment.

Not all of Yugoslavia's economy lies within the social sector. The present framework for the operation of private firms was laid down in 1965. In addition to the significant proportion of agriculture that continued in private hands, small private businesses which employed up to five people other than the members of the proprietor's family were permitted. These operations are particularly important in tourism and in transport and construction.

The private sector

In spite of the achievements of self-management, immense problems have been encountered. The attempt to ensure thoroughgoing participatory democracy means that long hours are consumed in meetings. Because work organizations are the principal agencies through which the social security and taxation systems are implemented, they must support a cumbersome and expensive bureaucracy. The attempt to introduce market mechanisms has been only partly successful, and the excessive decentralization of the system has resulted in the creation of a range of regional and even local monopolies rather than genuine competition. In the absence of real stimulus to efficiency, workers' councils have often raised wage levels above the true earning capacity of their organizations. In this they have been aided by local political officials, anxious to ensure their own positions, and by banks, over whose affairs large firms have had considerable influence.

The bureaucratic burden necessarily carried by self-managing organizations has provided a considerable disincentive to the creation of smaller enterprises, and the Yugoslav economy suffers from a general lack of specialist skills, particularly in the areas of high or specialized technology, business services, and agriculture. In 1988 pressures to remedy this situation led to the reform of laws relating both to the maximum level of employment and the maximum landholding permitted in the private sector.

Extreme decentralization in the economy and the consequent lack of adequate instruments for central economic management exacerbated the effects of the international oil crisis of 1973 and the world recession of the late 1970s and early '80s, resulting in relatively high levels of inflation and a growing foreign debt. Yugoslavia's inflation has been the highest in Europe. When the economic crisis reached such severe proportions that the aid of the International Monetary Fund (IMF) had to be sought in 1983,

the extreme fragmentation of the institutional structure meant that many months passed before the precise scale of the country's foreign debt could be established.

A policy of "stabilization" introduced under IMF advice proved difficult to implement. The package of reforms fought through in the late 1980s centred on the strengthening of the hand of the federal government in the control of the economy, particularly that of the National Bank of Yugoslavia, and included measures to open further the economy to foreign participation, to eliminate the remaining price controls, and to make it easier to close uneconomic plants. These reforms encountered fierce resistance from sections of the Communist Party leadership and from organized labour.

Strikes

Strikes have been of special interest to observers of the country since the first of these took place in the coal mines at Trbovlje (Slovenia) in 1958. Since then the annual rate of work stoppages has risen steadily. For the most part these have been short-lived occurrences intended to express dissatisfaction rather than to force the hand of management. By the late 1980s, however, there were more than 1,500 such strikes each year, indicating the seriousness of the conflict over government economic strategy and questioning the claims of self-management to have ensured economic democracy.

The main sources of revenue to the federal government are turnover taxes and payroll taxes, which are actually collected by the republics and provinces, and customs duties, which the federal government collects directly, along with a stamp duty on legal transactions. Personal income taxes are relatively unimportant, being levied only on those with very high incomes and the self-employed. The main expenditures are for defense and grants to pension funds. Fiscal devices have featured rarely as instruments of economic regulation. The inefficiency of the taxation system remains politically controversial.

Resources. Yugoslavia is a leading producer of antimony, copper, lead, bauxite, and magnesite. Antimony is mined at Ložnica in Serbia, and significant deposits have been found near Srebrenica in Bosnia. Copper is mined near Zaječar, in eastern Serbia. Kosovo has rich deposits of lead and zinc, with mines at Trepča, near Titova Mitrovica. Iron ore is mined at Omarska, which is part of the Ljubija mining complex near Prijedor in Bosnia. Iron ore deposits also occur in the Jastrebac, Pasjača, and Kukavica mountains of Serbia. Copper is mined at Majdanpek in eastern Serbia. Bauxite is mined at Posušje, Ličtica, Stolac, and Čitluk, near Mostar in Hercegovina. Coal, mainly in the form of lignite and bituminous coal, is mined at Ugljevik in Bosnia, Raša in the Istria region in Croatia, and Baljevac in Serbia. Fields in Vojvodina produce natural gas and petroleum. Oil has been discovered in the Stig region, near Požarevac in Serbia, and offshore drilling has continued along the Adriatic Sea. There is a large uranium deposit near Leskovac in eastern Serbia.

The third of Yugoslavia that is forested provides timber resources, giving rise to furniture, paper, and associated industry. Timber is an important export commodity. One of Yugoslavia's most important natural resources is the hydroelectric power offered by its rugged terrain, a potential that has not been fully exploited.

Agriculture. The variety of soils and climate is generally favourable to agriculture, especially in the northeastern plains. About a third of the area of the country is under cultivation. Although overall there has been a threefold increase in production since 1945, the rise has not kept pace with demand for food and industrial crops. Thus, there have been price increases, shortages, and a constant need to import agricultural products. The country has a mixture of private and socially owned farms. Private farms account for more than 80 percent of the area under cultivation. Holdings are, however, small (less than 7 acres [2.8 hectares]) and are split into a number of separate plots. About 60 percent of private farms are operated within voluntary cooperative schemes. Farm workers make up more than a quarter of Yugoslavia's work force.

Problems in agriculture

The small size of most private farms, their remoteness from markets, the inadequate use of fertilizer, and the backwardness of farming techniques help to explain why productivity remains low. Mechanization is a problem on very small farms. These difficulties are compounded by the aging of the farming population. The government has taken measures to stimulate the more efficient farms and to keep young workers in agriculture through means such as increasing the maximum permitted size of private holdings from 25 to 75 acres. Socially owned farms are considerably more productive than private ones. In the mid-1980s the social sector held about 16 percent each of the cultivable land and of livestock but provided about 40 percent of the total agricultural product.

Major food crops are corn (maize), wheat, potatoes, sugar beets, and rye. Fodder grasses have almost doubled in area in the postwar period, while cereals have declined or remained constant, testimony to the changing diet as Yugoslavs have become richer and to the growing importance of fresh meat in the export market. The numbers of pigs and poultry especially have increased. Other important crops are sunflower seeds, tobacco, and grapes (for wine).

Forestry is located where timber resources are most readily exploited, mainly in Bosnia and Hercegovina, Croatia, Serbia, and Slovenia. The major tree types are beech and oak in natural forests, but in areas of afforestation the principal types are spruce, black pine, and Scotch pine. During the 1980s Yugoslavia was an important export earner in wood, furniture, and cork products.

Industry. The process of industrialization had hardly begun in Yugoslavia by 1939; the few industries were located mainly in the northern half of the country, with some extractive industries in the southwest. World War II reduced Yugoslavia's industrial base, but since 1945 there has been considerable development. Industrial output increased some 10-fold between 1955 and 1985.

Attention has been paid to the role of industrial development in the improvement of the standard of living of the country and its use in promoting equitable distribution of national wealth. The initial phase of development emphasized production of primary capital goods such as iron and steel. In the 1960s more attention was paid to consumer goods production, and some traditional industries, such as timber, have declined relative to consumer goods. Primary industry has developed in areas with a good supply of natural resources and energy, while consumer industry has tended to be located nearer the northern urban markets. Secondary activity such as engineering is also concentrated in the north, with particular emphasis on heavy electrical engineering, a useful export base. Automotive engineering was initially imported in the form of complete plants or the licensed assembly of Western manufacturers' vehicles. There is now an important agricultural and commercial vehicle sector. The shipbuilding industry provides considerable employment and substantial foreign earnings. In the south the textile industry has a long history and is partly based upon local wool and cotton. A growing chemical industry provides the basis for modern plastics manufacture, synthetic-fibre plants, and the production of artificial fertilizers. In general, however, the growth of the primary and intermediate sectors has not kept pace with the demand for consumer-oriented industry. In terms of value the most important classes of industry are metal fabrication, electrical generation, transmission and distribution, finished textiles industry, furniture and other wood products, food products, machinery, and electrical appliances. Persons employed in mining and manufacturing constitute about a quarter of the total labour force.

Energy

More than one-third of Yugoslavia's electricity is produced by hydroelectric power stations; the rest comes from thermal stations and the nuclear power station at Krško in Slovenia. The principal fuels for the thermal power stations are domestic lignite and brown coal, but a considerable quantity of natural gas and oil derivatives is used, some of which is imported. The largest of the hydroelectric power stations is that at Đerdap on the Danube (the Iron Gate scheme). A joint venture with neighbouring Romania, it has a capacity of 2,300 megawatts. Other major schemes are at Bajina Bašta and Mratinje in the Drina valley and along the Cetina River in Dalmatia.

Tourism is of special importance to the Yugoslav economy. It provides a major source of convertible currency

and helps the balance of payments. Tourism offers opportunities for employment in many areas that would otherwise have very limited potential, and it has been a stimulus to the building and construction industry. In addition, it has been necessary to develop transportation facilities for tourist needs.

The most favoured locations for tourism are the seaside resorts of the Adriatic coast, where development is heaviest. The major cities of Yugoslavia, such as Zagreb and Ljubljana, are also attractive tourist destinations, followed by the mountain regions, where particular attention has been paid to the development of winter holiday resorts. Important cultural sites are also favoured, and following their opening to foreign visitors there has been a development of some of Yugoslavia's 150 spas. The largest of these, Stubičke Toplice, is located north of Zagreb, and others are near Sarajevo and south of Belgrade.

Major developments did not get under way until the 1960s, but since 1965 the numbers of both domestic tourists and foreign visitors have tripled. More than a quarter of foreign visitors are from West Germany, followed by Italy, Austria, and the United Kingdom, but most European countries are represented, including considerable numbers from eastern Europe.

Trade. Yugoslavia is not a member of either of the major European trading groups, the European Economic Community and the Council for Mutual Economic Assistance, but it cooperates with both and has contractual relations with more than 80 countries. The volume of foreign trade has risen throughout the postwar period. More than 70 percent of the exports are manufactured goods, especially machinery and transport equipment, electrical equipment, and clothing and footwear. Corn, meat and meat preparations, and fruits and vegetables are also important export items. About half of the export trade is with countries having centrally planned economies, especially the Soviet Union, Czechoslovakia, and Poland. About one-third of the foreign trade is with countries having developed market economies, including Italy, West Germany, and the United States.

Yugoslavia's major import is fuel, especially petroleum, accounting for about a quarter of the total by value. Other major imports are chemicals, sheet and plate steel, machinery and electrical equipment, and automobiles and buses. The principal food import is coffee. Some raw materials, such as rubber, crude cotton, and cellulose, are also imported. Yugoslavia's major import partner is the Soviet Union, followed by West Germany, Italy, Iraq, the United States, Czechoslovakia, France, and Austria.

Foreign investment in Yugoslavia is limited, but recent changes have allowed foreign partners to own up to 98 percent of the capital of joint ventures. The principal areas of investment are motor vehicle manufacture, chemicals, pulp and paper, and tire manufacturing. In former years

The foreign trade deficit

the foreign trade deficit was balanced by revenues from services and invisibles, especially shipping, tourism, and the remittances of Yugoslavs employed abroad. However, the servicing of an increasing foreign debt, a reduction in remittances from abroad, the increased cost of imports, and domestic inflation have resulted in the imposition of import controls in order to manage the deficit.

Transportation. Yugoslavia is located at the intersection of major European routes, along the Danube, the Sava, and the Morava–Vardar–Nišava valleys, but in modern times the difficult terrain has led to inadequate development of the road system and to a poorly unified rail system. By the 1970s, however, a much-expanded rail system, with new trans-Balkan routes, standardization of gauges, and some electrification, was responsible for 80 percent of public transport. This rationalization process has continued with the closure of all narrow-gauge lines and some redundant track and with the improvement of track to allow faster and heavier trains. More than a third of the rail network is electrified. Important new lines have also been opened, linking Belgrade with Bar on the Adriatic coast and completing the first Albanian-Yugoslav rail link.

Road traffic has developed rapidly and includes local, national, and international movement. During the 1980s more freight was carried by road than rail. About half of Yugoslavia's motorways were built between 1981 and 1985. They link the principal cities of the north and the developing routes for holiday and freight traffic from central Europe to the Adriatic and the south. After the accession of Greece to the European Economic Community, funds were made available to develop the road link to Greece through Yugoslavia.

Belgrade and Zagreb are the principal air transport centres, linked to airports in each of the republic's capitals. Major developments have taken place to serve European tour operators, and new airports have been built, often in difficult coastal, island, or mountainous locations, at Pula, Rijeka, Ohrid, Maribor, Tivat, Zadar, Dubrovnik, and Split.

Mediterranean transport and trade was formerly the lifeblood of the Adriatic ports of Yugoslavia, but it is now limited to fishing and holiday cruising, although there are important ship repair and bulk ore export facilities.

ADMINISTRATION AND SOCIAL CONDITIONS

Government. *The constitutional framework.* The framework of government in Yugoslavia is regulated by the constitution of 1974. It reflects the resolution of the political leaders to steer a middle way between the Western and Soviet traditions of government, through the Yugoslav system of "delegation."

There are three levels of government: the commune, the republic, and the federation. The 530 communes range in size from about 5,000 people in rural areas to 50,000 people in urban areas. Each is governed by an assembly. The communes are vital institutions in Yugoslav political life, as they are the direct agents for the collection of most government revenue and provide necessary social services.

The communes are incorporated in the six republics and the two autonomous provinces (which exercise many of the functions of republics). The republics and provinces are governed by assemblies of delegates selected at the communal level. The federation has its own assembly, which sits in Belgrade. Elections are for terms of four years; no delegate may serve in the same capacity for more than two terms, making for a system of "rotation." Any delegate's mandate may be revoked by the appropriate constituency at any time.

Plural voting

Yugoslavia has a system of plural voting, according to residence, employment, and sociopolitical participation. The assemblies of the communes, republics, and autonomous provinces each consist of three chambers. The Chamber of Associated Labour is formed from delegations representing the self-managing work organizations. The Chamber of Local Communities consists of local citizens drawn from territorial constituencies. The Sociopolitical Chamber is elected from members of the Socialist Alliance of the Working People of Yugoslavia, the League of Communists, the trade unions, and organizations of war veterans, youth, and women. The Federal Assembly has two chambers: the Chamber of Republics and Provinces and the Federal Chamber, which has delegations from work organizations and sociopolitical organizations.

The executive functions of government are carried out by the Federal Executive Council (FEC), which introduces proposals for legislation to the Federal Assembly and appoints the secretaries of federal committees (*i.e.,* ministers). The FEC consists of a president, members representing the republics and provinces, and federal secretaries and other officials representing various administrative agencies. Members of the FEC are normally elected for four years.

The constitution of 1974 vested the presidency of the Federation in Josip Broz Tito for life. Following his death in 1980 it was transferred to a collective presidency composed of one representative from each of the republics and autonomous provinces, with the intention of preserving a balance within the leadership of all the regional and national interests. The president of the League of Communists is an ex officio member. Members are elected for a term of five years. Each year they elect a president and a vice president. The president acts as the head of state. The collective presidency has not found it easy to impart a strong shared sense of direction to the country.

More than 1.6 million individuals occupy positions

within the system of delegation, about 53,000 of them as members of the assemblies. In addition there are nearly half a million members of workers' councils.

The judiciary. There are municipal, republican, and federal courts. Judges at each level are appointed by the various assemblies. At the summit of the system is the Constitutional Court, composed of 14 members (a president and 13 judges), two drawn from each of the republics and one from each of the autonomous provinces. They are appointed by the Federal Assembly for nonrenewable eight-year terms.

Sociopolitical organizations. The constitution recognizes an important role in the structure of delegations for the sociopolitical organizations. The most important of these is the League of Communists of Yugoslavia. This is organized on a republican basis and governed by a presidium of representatives from each of the republics. It has a membership of more than two million. The league's program affirms that it is the leading ideological force in Yugoslav society. Its rule is implemented largely by the active commitment of its members and their occupancy of elective positions in the system of government and leadership posts in other bodies. Important in this respect are the Confederation of Trade Unions, the League of Socialist Youth, and the Veterans' Organization. The trade unions do not see themselves as serving principally the function of collective bargaining over wages in a self-managing system but emphasize more the tasks of providing political leadership and welfare for their members.

The League of Communists

The Socialist Alliance of the Working People of Yugoslavia unites all those committed to socialism. Its most important function, in addition to the political motivation and education of its members, is the management of the electoral process.

The armed forces. The distinctive Yugoslav philosophy of defense—usually referred to as "Total National Defense and Social Self-Protection"—rests on the assumption that all citizens are potentially involved in the defense of their country. The Yugoslav People's Army, including the navy and air force, is made up of both professionals and conscripts. There is universal male conscription, with a term of 15 months for the army and 18 for the navy and air force. Conscientious objection is forbidden by law. These units are supplemented by reservists of the Territorial Defense Force. The Civilian Defense Force comprises virtually all other adult citizens, particularly women.

International relations. Since the early 1960s the backbone of Yugoslavia's policy in international relations has been its nonaligned stance with respect to the great-power blocs. President Tito was one of the founders of the nonaligned movement, but since his death Yugoslavia's attention abroad has been directed to its position as a European, and especially a Mediterranean, country and to the development of relations with its Balkan neighbours.

Education. There are three levels of education in Yugoslavia. Elementary education, which is compulsory from the age of seven, lasts for eight years. Secondary education lasts for four years in either a vocational or general secondary school. Higher education is provided in high schools, academies, and universities. Nursery education is an important supplement to the system, but provision is uneven and generally lacking in rural areas.

During the 1970s the Yugoslav secondary education system underwent reform, designed to strengthen the vocational element. The strengths of the secondary level of education vary with the economic and cultural needs of the locality. On the Dalmatian coast are found several schools specializing in tourism, for example, and Zagreb has an important program for musical education.

University education has expanded rapidly since the war. There are 18 universities and a large number of research institutes. The University of Belgrade (founded 1863) is the largest institution of higher education in Yugoslavia. Two very old universities are those of Ljubljana, founded in 1595, and Zagreb, founded in 1669.

Universities

The financing of education includes the provision of a limited number of bursaries, loans, and credits. Instruction is provided in minority languages, even up to the university level, where there is adequate demand.

Health and welfare. The general characteristics of social conditions in contemporary Yugoslavia have been shaped by the exigencies of transforming an economically undeveloped and largely traditionalist society in the direction of modernity. Enormous strides have been made in this respect: in 1953 illiteracy rates stood at 45 percent, as opposed to 10 percent in 1986; infant mortality rates declined from 116.5 to 27.1 per thousand in the same period.

The institutional framework for the provision of social services is provided by two devices unique to Yugoslavia—the self-managing community of interest and the social compact. Institutions such as clinics or universities are run on a self-managing basis by their staffs, but they are responsible for their general policy and financial accountability to the self-managing communities, on whose governing bodies sit delegates from the various parties.

Yugoslavia has insurance and benefit schemes covering health protection, pensions, disability payments, family allowances, and unemployment relief. These are organized principally on a local basis, and the level of provision varies enormously among different communes. Access to benefits is principally through the place of employment, which has meant that serious gaps in provision have developed in unemployment benefits, coverage for the families of the unemployed, and coverage for those working in the private sector. From the mid-1970s there has been a steady movement to extend coverage.

Housing has presented a difficult problem for postwar Yugoslavia. Needs have been met generally by the provision of small apartments, constructed in large blocks. Housing is largely a communal concern, although larger enterprises may take responsibility for housing their employees. There remain serious shortages, however, of which high prices and long journeys to work are evidence.

CULTURAL LIFE

The country's historical division between areas of Austro-Hungarian and of Ottoman suzerainty is still visible in a number of respects. Relations of authority, at all levels from parental to governmental, tend to be both more formal and more patriarchal in character in the former Ottoman areas. There are clear regional differences of dialect, and the change from the use of the Roman to the Cyrillic alphabet is one that obviously strikes the visitor.

Yugoslavs have a strong sense of the importance of family, and patterns of kinship and patronage provide an important framework for a good deal of social life, exemplified in the institution of *kumstvo* (sponsorship, or godparenthood).

The contrast between rural and urban areas is one of the most important lines of cultural cleavage in Yugoslavia. Although these differences have been attenuated since the war, folk costume is still to be seen in many rural areas, particularly at weddings, on saint's days, and at markets. Even among urbanized Yugoslavs elements of peasant culture remain important. Traditional music is widely popular and has become blended with imported pop styles. Traditional dances are performed at many festivities. Indigenous traditions of painting thrive in Croatia, especially in the Hlebine school of painters (named after the village around which several exponents live). This naive art is represented by the work of the Generalić family, Ivan Večenaj, and others. Urban areas too have long and vigorous cultural traditions. Even in folk song there are forms of distinctively urban music, the *starogradske pjesme.*

Importance of folk culture

Serbo-Croatian and Slovene became standardized as literary languages during the 19th century, and writers began to use Western literary forms, such as the novel (for example, August Šenoa in Croatia and Branislav Nušić in Serbia). Dalmatia has an even older urban literary tradition. The auditorium of the theatre on the island of Hvar, opened in 1612, is still preserved. The present National Theatre building in Belgrade (probably the most important centre of theatre in the country) dates from 1868, and that in Zagreb from 1895. There is a rich dramatic repertoire in both Slovene and Serbo-Croatian. Macedonian literature has experienced remarkable growth following the official recognition of the language in 1945, and literary work is encouraged in several minority languages.

Painting and the graphic arts also have a long history in Yugoslavia. The country's many ecclesiastical frescoes of the medieval period are counted among the great treasures of world art. The European tradition of academic painting was imported during the 19th century, and the subsequent movements in pictorial and plastic art have had their adherents in Yugoslavia. Opera, ballet, and orchestral music also began to develop, particularly in the Habsburg lands, during the 19th century. There are specialized academies to support training in all aspects of artistic endeavour.

Yugoslav writers and artists have rarely achieved international acclaim, with some important exceptions. Ivo Andrić was awarded the Nobel Prize for Literature in 1961, and his citation noted especially his novel *Na Drini ćuprija* (Bridge on the Drina; 1945). Printmakers, especially from the Academy for Fine Art in Ljubljana (such as Vladimir Velicković and Miroslav Šutej) have exhibited widely abroad. Yugoslav sculpture has a strong reputation dating back to the work of Ivan Meštrović in the interwar years. Sculptures are a frequent attraction in parks and other public spaces throughout the country.

Another field in which Yugoslavs have secured international recognition is that of the cinema. Filmmaking began in Yugoslavia in the last months of World War II. Documentary films appeared first, and later animated and feature films were made. After a period of socialist realism, the cinema has emerged as one of the most powerful media of social and political criticism. Important filmmakers include Dušan Makavejev; Slobodan Šijan, who was awarded the Georges Sadoul prize in 1981 for his inventive and humorous treatment of the period of Nazi invasion; and Emir Kusturica, who won the Golden Lion at the 1981 Venice Film Festival for his *Sječaš li se Dolly Bell?* ("Do You Remember Dolly Bell?").

Film festivals

Film festivals are held in Pula, Belgrade, and Zagreb. The Dubrovnik summer festival is a showcase principally for the performing arts; Struga in Macedonia is the venue for an annual international festival of poetry.

Each of the regions of Yugoslavia has its own television and radio service, and the country has become a significant producer of program material, although a substantial amount of television programming is imported. Domestic television production is generally considered to be more innovative than elsewhere in eastern Europe.

There are more than two dozen daily papers: the only one providing a genuinely federation-wide coverage is *Borba,* the organ of the Socialist Alliance. The Belgrade daily *Politika* was established in September 1904. The press in Yugoslavia has a reputation for relative independence, with high-quality coverage of current affairs served in particular by weekly magazines such as *Danas* (Zagreb) and *NIN* (*Nedeljne Informativne Novine;* Belgrade). Conflict with the authorities is avoided as much by the anticipatory self-censorship of journalists as by judicial interference in the press. (J.B.A./N.H.C.)

For statistical data on the land and people of Yugoslavia, see the *Britannica World Data* section in the BRITANNICA WORLD DATA ANNUAL.

History

Yugoslavia is in many respects a very young country. It was formed from parts of the Austro-Hungarian and Ottoman empires, when these were broken up by the Treaty of Versailles in 1919, together with the former independent states of Serbia and Montenegro. Known as the Kingdom of Serbs, Croats, and Slovenes (the major national groups from which it was composed), the new state only adopted the name Yugoslavia in 1929, when King Alexander, frustrated by the slowness of progress toward national unity, established a royal dictatorship. The "First Yugoslavia" fell apart in 1941 under the impact of invasion by the Axis powers, but the country emerged after the war in its contemporary form as a federation of socialist republics. In spite of the relative brevity of its existence, an understanding of modern Yugoslavia is not possible without an appreciation of the long and complex historical process by which its constituent regions and peoples have been brought together within a common political framework.

THE PREHISTORIC PERIOD

Traces of human habitation have been found in the Balkan Peninsula dating back more than 200,000 years to the Lower Paleolithic (Old Stone Age), but a clear picture of the lives of the inhabitants only emerges in the Neolithic period, from around 7000 BC. At that time the domestication of plants and animals permitted the existence of larger groups and the creation of more permanent settlements, from the remains of which can be inferred a good deal about the culture. Together with east central Europe, the Balkans formed the heartland of an Old European civilization that flourished between 7000 and 3500 BC. There is evidence of dense settlement particularly in the Pannonian Basin, along the Sava and Danube rivers, and spreading northward into modern Hungary along the Tisa and southward down the Morava–Vardar corridor. Food production had developed to the point that it was possible to support a measure of craft specialization, including pottery making and the smelting of copper, and small towns were formed. A written script had evolved, suggesting that differentiated governmental and religious institutions had also developed. Several important sites in Yugoslavia give insights into Old European culture, particularly those at Starčevo and Vinča, near Belgrade, and at Lepenski Vir, on the Danube above the Iron Gate.

Early settlements

The earliest settlements relied largely on hunting, with the aid of dogs, the only domesticated animal, and fishing. Later emmer (a primitive variety of wheat) was cultivated, and sheep, goats, and pigs were raised. Habitations (at Vinča laid out clearly in streets) were of wattle-and-daub construction, with evidence of the use of split planking and clay plastering for floors. A number of figurines and other objects indicate a developed cultic life, and linear markings on vases, spindle whorls, and figurines, and occasionally on tablets, confirm that the people had a knowledge of writing. The pattern of the fine burnished pottery indicates regular trade with the Maritsa area of central southern Bulgaria, and the use of the Aegean *Spondylus* shell in decorative art suggests even wider links of communication.

After 3500 BC the region was gradually infiltrated by seminomadic pastoral peoples, believed to be speakers of languages of the Indo-European family, who came southward and westward from the Russian steppes. Their way of life differed in many respects from that of the Old Europeans. Ruled by military aristocracies, they constructed hill forts, and they had domesticated the horse and developed horse-drawn vehicles. One of their centres has been excavated at Vučedol, near Osijek. By the beginning of the 3rd millennium they had penetrated as far as Macedonia and the Adriatic coast, and the destruction of the Mycenaean civilization is attributed to their arrival. Their extensive trade routes carried amber, gold, and the bronze that was the basis of their superior military technology.

They were divided into tribal groups, one of which, the Illyrians, become firmly established throughout western Yugoslavia, Albania, and Epirus. By the 7th century BC they had acquired the capacity to work with iron, of which northwestern Bosnia was an important centre, and this skill became the basis both of their extensive trade with the emerging Greek city-states and of the power of the native aristocracies. Magnificent royal burials testify to the importance of these aristocracies. An Illyrian kingdom emerged by the end of the 5th century BC around Shkodër (northern Albania and Montenegro). It survived until its annexation by the Romans in 168 BC.

During the 1st millennium BC, Greek influence expanded throughout the coastal regions, particularly in the conduct of trade. The treasure discovered at Trebenište, near Ohrid, attests to the circulation of Greek craftsmanship, and the Greek origins of several settlements along the Adriatic coast are remembered in their names: Vis (Issa), Havar (Pharos), Korčula (Corcyra Melaina), Solin (Salonae), Zadar (Iader) and Trogir (Tragurion). To Greek cultural influence can be attributed the introduction of the potter's wheel and coinage and the rise of the mercantile city.

East of the Morava–Vardar the land was periodically subordinated to the warrior kingdoms of the Dacians and

Thracians. Much of the life of Illyrians, Dacians, and Thracians is known not only from archaeological finds but also through the writings of contemporary witnesses such as Homer, Herodotus, Thucydides, and Xenophon.

In the middle of the 4th century BC, Philip II succeeded in extending Greek military power beyond his native Macedonia into both Thrace and Illyria, and his son Alexander the Great took the battle against the Dacians beyond the Danube in 335 BC.

None of the early aristocratic lineages managed to ensure the stable development of a state, particularly as the area continued to be subjected to invasions by tribes from the north and east. Beginning in about 300 BC, bands of Celts began to penetrate southward. Their superiority rested in part on their mastery of iron technology, which they used to beat both swords and plowshares. The extent of Celtic expansion is indicated not only by their material remains but also by place-names. The name Singidunum, by which the Romans knew the settlement on the site of Belgrade, is at least partly of Celtic origin.

Celtic expansion

At the time of Roman colonization many of the principal elements of the ecology, economy, and technology of the peoples of the region, elements that were to provide the framework for their daily lives until the 19th century, were in place.

THE ROMAN EMPIRE

Macedonian hegemony in the Balkans was short-lived. At the end of the 3rd century BC the Romans began their expansion into the Balkan Peninsula in search of iron, copper, precious metals, slaves, and agricultural produce. The Roman struggle for domination, against the fierce resistance of the native peoples, lasted three centuries. The Celts were overpowered at Delphi in 280–279 BC. The Illyrians were heavily defeated in a campaign that took place between 35 and 33 BC, but they were not finally subdued until AD 9. Their land became the province of Illyricum. The area that is now eastern Serbia was conquered by Crassus, proconsul of Macedonia, in 29 BC and incorporated into the Roman province of Moesia. The conquest of the Balkans was not completed until the reign of the emperor Trajan, in AD 106. The experience and skill of the warriors of the region was recognized by their conquerors, and Thracian and Illyrian legions later provided a significant component of the Roman armies. Five natives of Illyria became emperors: Claudius, Aurelian, Probus, Diocletian, and Maximian.

Roads, arenas, aqueducts, bridges, fortifications, and other constructions attest to the thoroughness of Roman occupation. Many of these are still in a good state of preservation; the emperor Diocletian's palace at Split, the amphitheatre at Pula, and the Roman townships at Bitola (Heraclea) and Stobi are examples. A network of roads was built for trade and for military purposes, although many of them followed older routes. The names of several modern towns reveal their Roman origins: Sisak (Siscia), Sremska Mitrovica (Sirmium), Niš (Naissus), and Skopje (Scupi).

Diocletian divided the empire in AD 285. Despite a brief attempt at reunification under Constantine, a fundamental and permanent division was imposed after 395. The line ran roughly northward from the modern Yugoslav-Albanian border on the Adriatic to Sirmium, whence it followed the line of the Sava and Danube rivers, creating a cultural boundary that has had profound consequences for the development of the South Slav lands.

THE BARBARIAN INVASIONS AND THE FALL OF ROME

Roman domination in the region was of relatively short duration. Military clashes with the Goths began early in the 2nd century, and their increasing severity compelled the Romans to withdraw from Dacia by 272. Gothic raids probed far southward into Thrace throughout the 3rd century. The Goths were followed by Huns, Bulgars, and Avars over the next 200 years. The collapse of the Western Empire in the face of the advancing Germanic Ostrogoths at the end of the 5th century left the Balkans nominally under the rule of Constantinople. In fact, the disruption of imperial administration had gone so far that effective control was no longer possible.

Along with other seminomadic peoples during this time, there began to move into the area tribes of Slavs, a group of Indo-European-speaking peoples. The Slavs moved southward from central Poland, where they had long been settled. In social structure they differed in many respects from the other barbarian peoples, having a largely settled way of life, with agriculture and animal husbandry providing most of their needs. They lacked the traditions of military aristocracy that were typical of their neighbours. Organized as loose fraternities of clans owing allegiance to a *župan,* or chief, they settled the sparsely populated areas left by the raids of the more warlike peoples, typically paying tribute to them and providing them with auxiliary troops.

Slavs

Early in the 6th century, Slavs crossed the lower Danube, and by the middle of the century their raiding parties were reported as far south as Epirus and on the Adriatic coast. The Byzantine emperor Justinian I attempted to control the situation, reconquering the coastal fringe of Dalmatia and inviting the Slavs to become his *foederati,* or paid border guards. The relative strength of the forces in the area is suggested by the Slavs' effective vassalage to the Avars, a Turkic people of warrior-nomads who continued to lead their Slavic subjects in raids against cities of the Byzantine Empire. After 591 the empire made a sustained attempt to subdue the Balkans, but mutiny among the imperial armies in 602 put an end to this and the Byzantine emperors found themselves paying tribute to the Avars. From this time onward, contemporary sources refer to the area as Sclavinia—the land of the Slavs.

It was not until the defeat of a combined Avar-Persian invasion in 626 that the tide was turned, and both the Avar and Bulgar khanates began to crumble, undermined as much by the rebellion of their subjects as by Byzantine strength. Shortly afterward the emperor Heraclius formed an alliance with two of the stronger Slavic tribes, the Serbs and the Croats, who at that time were settled north of the Carpathian Mountains. With the aid of the Byzantine navy the Serbs and Croats occupied the hinterland of the Dalmatian coast before pushing the Avars and Bulgars eastward, effectively Slavicizing the whole region. Although they mingled with other Slavic tribes already settled in the area, the Serbs and Croats gave their names to parts of modern Yugoslavia: Serbia and Croatia. The remaining non-Slav peoples of the region, Illyrians (ancestors of the modern Albanians), Italians, Vlachs, and Morlachs, survived in some areas.

BYZANTIUM AND THE EARLY SLAV STATES

The extremely complex history of the South Slavs during the early Middle Ages can be explained by reference to two main factors. In the first place, the early Slav kingdoms were located at the intersection of interests of several surrounding larger powers, whose fortunes waxed and waned. From the 8th century onward there were several important forces bearing on the area. To the northwest was the expanding Frankish empire, which reached its height under Charlemagne. To the north and east the Magyars were increasingly the dominant power, supplanting the Avars. (Between the Franks and the Magyars the short-lived Slav kingdom of Moravia, set up in the 840s, was squeezed out at the beginning of the 10th century.) To the east the Bulgars also benefited from the decline of the Avars. The Byzantine Empire to the south, although it never managed to consolidate its hold on the Balkans, kept alive a nominal claim to the region and intermittently succeeded in exercising imperial authority. Within the circle of powerful contenders, a succession of petty principalities, kingdoms, and aspiring empires rose and fell until the Ottoman invasion in the 14th century. Although none of these states attained any great extent or survived for more than a few generations, each has significance as the basis of historical myths that still bind their audiences today. Each nationality grounds its claims to a home territory, to a cultural identity, and to a specific stake in the contemporary political and economic struggle by reference to a baseline in history—to the point at which it had "its own" state.

The second main factor in the history of the South Slavs was the nature of their social and political organization,

Social organization

which did not make for the ready creation of stable and centralized states. The basis of social organization among the South Slavs was the *zadruga,* a large extended family, which was governed by a fairly democratic consensus of its adult members (though with a strict division between the male and female spheres of activity) under the leadership of a patriarch (*domaćin* or *starešina*). The *zadruga*s were typically united on a village basis around a single lineage (*bratstvo*) under a headman, or *čelnik*. Larger political units covering a district might be gathered under a *župan,* who would sometimes have his seat at a particular fortified strong point, called a *grad.* Because this system was based on ties of kinship and locality, together with a strong sense of racial separatism that undermined cooperation with the non-Slav populations of the region, it militated against the sustained collaboration of these larger groups, although several *župani* might on occasion be constrained by an external threat to unite under a prince (*knez*). Developed during the migratory period, this system survived in some remoter districts even after the creation of a unified South Slav state had rendered it obsolete.

Because of the relevance of this period for the subsequent creation of distinct national identities, it is useful to discuss the history of the early states separately and in detail.

The Slovenes. The Slovenes were among the first of the Slavic tribes to settle in the Balkans, appearing in the upper Sava valley during the latter half of the 6th century. At that time they owed allegiance to the Avar khans. After the defeat of the Avars by Heraclius, a Slavic kingdom emerged under Samo (reigned 627–658). His kingdom extended from the Sava valley northward as far as Leipzig. It came under Frankish rule in 748, and the processes of Germanizing the aristocracy and converting the population to Christianity began. These policies were prosecuted with particular vigour under Charlemagne. From this time onward the Slovenes were included in the Western, or Roman Catholic, church.

During the 10th century Styria and Carniola were set up as march counties to defend the Holy Roman Empire from Magyar incursions. In the 13th century they fell under the rule of Otakar II of Bohemia, who like Samo tried to establish a Slav empire. Following the defeat of Otakar in 1278, Styria was acquired by the Habsburg family, providing the foundation for their territorial expansion. Carinthia and Carniola were added in 1335. These three political units by no means corresponded to Slovene linguistic boundaries, which tended to contract southward over the centuries. Istria was added to the Habsburg domains in 1374 and the city of Trieste in 1382.

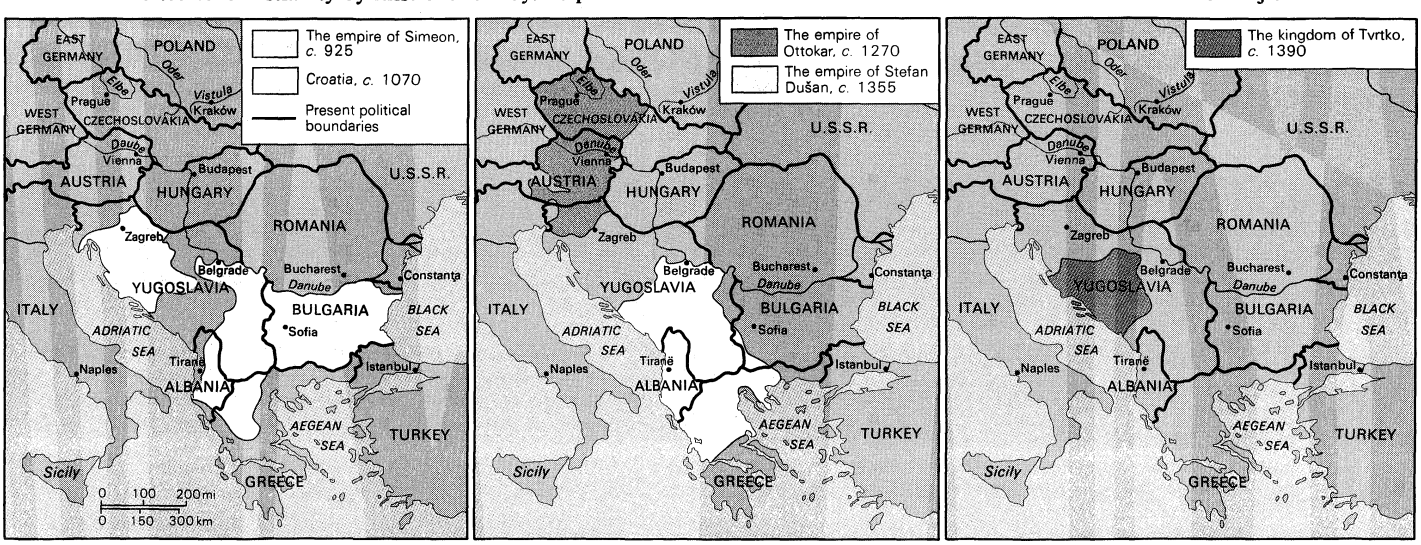
The Croats. Although the Croats arrived in the Balkans under the patronage of Byzantium, their allegiance was temporary. In 803 the loose federation of Croatian tribes, spreading from northern Dalmatia into Pannonia, accepted the suzerainty of Charlemagne, and they were converted to Christianity by missions sent by the patriarch of Aquileia. This antedated the mission to the Slavs of Cyril and Methodius; hence they remained oriented toward the Western church.

In 887 Emperor Leo VI reasserted Byzantine authority in the region, but about 924 a *župan* of Nin named Tomislav led an uprising against imperial control and assumed the title of king. The Croatian capital was set up at Biograd on the coast. Little is known about the kingdom of Tomislav and his successors. On the death of King Zvonimir in 1089 the succession was disputed, and appeal was made to his brother-in-law Ladislas I of Hungary, who took control of Croatia in 1091. In 1094 the bishopric of Zagreb was established as the principal ecclesiastical centre of Croatia. Ladislas' successor, Koloman, was crowned king of Dalmatia and Croatia in 1102. In spite of periodic challenges to the legitimacy and terms of the constitutional link with Hungary, it remained in force until the collapse of the Austro-Hungarian monarchy in 1918.

Union with Hungary

Dalmatia. Although the Slav immigrants of the early 7th century were able to expel or assimilate the majority of the existing population throughout the greater part of the Balkan Peninsula, they were less easily able to establish themselves on the Adriatic coast. There, partly because of its naval superiority, Byzantium was able to sustain its authority more readily than elsewhere, and several wealthy maritime cities—especially Cattaro (Kotor), Zara (Zadar), and Ragusa (Dubrovnik)—managed to preserve their distinctive cultural identity and political autonomy. After the northern part of the coast was annexed to Charlemagne's crown in 806, an uneasy dual sovereignty shared by the Carolingians and the Byzantines was sustained over the next two centuries. (In view of the strength of Dalmatian piracy in this period, it seems to have been a rather ineffectual sovereignty.)

During the 10th century Venice rose in importance as a maritime power. In 1000 the doge Pietro Orseolo II cleaned out the Neretva pirates and, with the sanction of the Byzantine emperor, assumed the title of duke of Dalmatia. Venetian trading interests were established in several coastal cities. Croatian power revived briefly on the coast under the Croatian kingdom in the middle of the 11th century, but it was almost immediately absorbed under the Hungarian crown. The coronation of Koloman as king of Croatia and Dalmatia in 1102 signaled the start of a period of sustained and intense conflict between Hungarian and Croatian interests on the one hand, and Venetian on the other, for power along the Adriatic. Control of the coastal cities was particularly important for Venice as a maritime trading power. Between 1115 and 1420 there were 21 wars between Venice and Hungary; the Dalmatian cities changed hands repeatedly, but their burghers retained a measure of autonomy by playing off one side against the other. During the 14th century the balance of power on the Adriatic was upset by the rise of the Bosnian state under Tvrtko and of the Nemanjid



The legacy of the medieval empires.

empire in Serbia. Venice was able to capitalize on the disorder, so that by 1420 only Almissa (Omiš) and Ragusa remained outside Venetian control.

In spite of Venetian ascendancy (which was only effective within a narrow coastal strip), the Slavicization of Dalmatia continued during the 14th century as more and more Slavs sought refuge in the region from the advancing Turks.

Bosnia and Hercegovina. In the early Middle Ages Bosnia never developed the sense of political cohesion that was displayed by other regions in the Balkans, possibly due in large measure to the characteristics of its physical geography. The early history of the region is obscure, and even the origins of the name Bosnia are uncertain. The emperor Constantine VII Porphyrogenitus (reigned 913–957) referred to it as a part of the land of the Serbs. From the arrival of the Slavs in the 7th century it seems to have preserved to a considerable extent the local autonomy of the tribal *župe* (districts), although different areas from time to time recognized the overlordship of Croats, Serbs, Hungarians, and Venetians, as well as that of the Byzantine emperors.

In 1180 a Bosnian despot named Kulin took for himself the title of ban (governor) and established a state covering the centre and north of the modern republic of Bosnia and Hercegovina. While his state proved to be ephemeral, his religious loyalty was of lasting significance. Although at first he was an adherent of the Roman Catholic church, Kulin declared his conversion to the sect of Bogomils. This movement appeared in the Balkans during the 10th century, taking root especially in Macedonia and Bulgaria. By the end of the 12th century, Bogomilism was well established in Bosnia. The name refers to its supposed founder, a Bulgarian monk named Bogomil, but it is known to be only one version of a group of related heretical sects that flourished across Asia Minor and southern Europe during the Middle Ages under a variety of names, the best-known being the Patarenes, Cathars, and Albigensians. The theology of these movements was a syncretism of Christian and Eastern beliefs, and it was condemned as heretical principally for its Manichaean dualism. (According to this belief the world is governed by two principles, good and evil, and human affairs are shaped by the conflict between them; the whole visible world is given over to Satan.) Roman Catholic and Orthodox theologians regarded this view as a denial of the ultimate sovereignty of God; in addition, the Bogomils challenged the authority of the established church. The Bogomil sect became known as the "Bosnian church," and its members elected their own bishop and elders. Many members of the aristocracy were converted to Bogomilism (their acceptance of Christianity seems to have been rather incomplete anyway), which was widely popular because of its opposition to tithing.

The Bogomils

Not surprisingly these developments attracted the hostility of the papacy, and in 1221 a papal legate was sent to Bosnia to institute a purge of the local clergy. A mission of the Dominican order with a record of successful combat against the Cathars in France was sent to establish religious conformity, and these ecclesiastical measures were backed up by the military interference of Hungary. After the death of Kulin in 1204 the power of the Bosnian ban was reduced, and at times during the 13th century Kulin's successors were little more than puppets of the Croatian or Hungarian (and occasionally also the Serbian) state. The attempt to secure religious orthodoxy was never fully successful, however; the Bogomils also resisted the sustained efforts of Franciscans, who were introduced under the patronage of the Croatian princely family of Šubić (also known as Zrinski) and who became particularly active in the persecution of heretics during the latter part of the century.

The Šubić family was ousted in 1322, and another native Bosnian, Stjepan Kotromanić, was elected ban. He in turn was succeeded in 1353 by his nephew Tvrtko, who embarked upon a career of territorial expansion by annexing the neighbouring principality of Hum and parts of Dalmatia. The Serbian empire to the east was by that time already weakening under the impact of Turkish advance, and Prince Lazar ceded to Tvrtko a substantial tract of what is now Montenegro in return for military aid. At the height of his power Tvrtko styled himself "King of Serbia, Bosnia, Dalmatia, Croatia, and the Primorje."

Tvrtko died in 1391, leaving the inevitable problem of succession. His kingdom disintegrated rapidly into petty chiefdoms patronized by the surrounding powers. The ruler of one of these, accepting the protection of the Holy Roman Empire, assumed the title duke (*herceg*) of St. Sava, which is remembered today in the name Hercegovina.

Serbia. In addressing the history of medieval Serbia it is important to bear in mind that to speak of "nations" in the modern sense is anachronistic. (It needs to be emphasized also that the same point holds true of Bulgaria.) Following the arrival of the Serbs in the Balkan Peninsula early in the 7th century, they seem to have been initially settled in the area between the Tara and Ibar rivers. They rapidly underwent assimilation with the other Slav tribes in the region, constant population movement being necessitated by the military instability of the area and, in particular, by the succession of barbarian raids that regularly depopulated large tracts of land.

Originally the Bulgars were probably a seminomadic tribe of the Ural-Altaic group. Establishing themselves at first as an aristocratic lineage extracting tribute both from the sedentary Slavic tribes and from the other indigenous peoples of the region, they too became assimilated and adopted a Slavonic language. (The term Bulgarian, as opposed to Bulgar, thus correctly refers to this later stage of their development.)

The Bulgars

Throughout the period from the 7th to the 13th century the Byzantine emperors periodically attempted to establish their authority across the Balkan Peninsula through the creation of a system of military marches (themes). They were most successful during the 11th and 12th centuries. The emperors contended for the heartland of the peninsula with a succession of confraternities of the Slavic tribes settled there, frequently using one group as allies in order to secure domination over others. When Byzantine power was weak, these local powers, gathered under the uneasy leadership of a *veliki župan* ("grand *župan*"), managed to establish control over a substantial part of the territory and the *veliki župan* declared himself king or even emperor. It is easy to gain a false impression of continuity between these states simply because several of them have been identified as "Serbian" or "Bulgarian."

The first Serb state emerged about 850 when a *župan* called Vlastimir led a union of southern Serbs in resistance to Bulgarian expansion. His acknowledgment of the suzerainty of the Byzantine emperor was significant in that the Serbian court then became an important channel for the spread of the Eastern tradition of Christianity among the Slavs. In 862 the Moravian prince Rastislav, preferring to avoid entanglement with the Frankish court, sent emissaries to the emperor Michael III requesting that he furnish missionaries. Michael commissioned two brothers from Thessalonica, Cyril (Constantine) and Methodius, to undertake the task and encouraged them to preach in the vernacular. To facilitate this task, Cyril invented a script, based upon Greek but adapted to suit the phonetic peculiarities of the Slavonic tongue, and proceeded to translate the Bible. The script, known as the Glagolitic, was based on the cursive form of Greek characters. It was later replaced by the Cyrillic alphabet, named after Cyril, which was based on the uncial form of Greek letters. Cyril used as his standard the dialect spoken by the Slav tribes of Macedonia, which thus was preserved as Old Church Slavonic.

The dissemination of Christianity to the Slavs was not begun by the brothers, but it received an enormous stimulus from the translation of the scriptures and liturgy. Although the mission to the Moravians was only partly successful, the wider significance of the work of these "apostles of the Slavs" was considerable. Not only was the influence of the Eastern church permanently assured over the greater part of the Balkans, but the Cyrillic alphabet also became one of the most visible cultural badges separating the Serbs and Bulgarians (together with the Montenegrins and Macedonians) from the Croats and Slovenes.

Following the death of Vlastimir his successors lost ground, at first to the Bulgarian empire under Simeon, becoming simply vassals to the Bulgarian tsars by the end of the 10th century, and then to the Macedonian empire of Samuel, until this in turn was overthrown by Byzantium in 1018. A protracted period of struggle between two contending proto-states ensued. A short-lived kingdom emerged at the end of the 10th century, centred on modern Hercegovina and Montenegro and known variously as Zeta or Dioclea (Duklija). This kingdom grew rapidly to take in the Adriatic coast from around Lake Skadar (Scutari) to the island of Brač and extended inland almost to the modern cities of Belgrade and Niš. Its ruler Michael received the title of king from Pope Gregory VII in 1077, and in 1089 the bishopric of Bar (Antivari) was raised to an archbishopric. The last king, Constantine Bodin, was elected emperor in Prizren.

The second Serb state

The second, and ultimately the more successful, state arose some time toward the end of the 11th century. This state was known as Raška (Rascia) and at first was based in the region of modern Novi Pazar. In 1169 Stefan Nemanja became *veliki župan* of Raška, and, seizing the opportunity offered by a disputed succession in Constantinople, he began to extend his territory. By the time of his retirement to a monastery in 1196 he had consolidated control over the former realm of Zeta. His son, Stefan Prvovenčani (the "First-crowned"), was so called because in 1217 the pope granted him the title of king.

In this period, the decline of the Byzantine Empire, which was all but destroyed by the soldiers of the Fourth Crusade (1204), presented an opportunity at first for Bulgarian expansion. For a time the Bulgarians were able to occupy Niš, Prizren, Skopje, and even Ohrid, so that under the emperor Kalojan, during the early years of the 13th century, the second Bulgarian empire achieved the greatest extension of Bulgarian power. Boril, the successor to Kalojan in 1207, was not able to contain a series of attempts to set up splinter states, and the Bulgarian state was steadily displaced by the Serbian Nemanjids. Uroš II (reigned 1282–1321) occupied Skopje and made it his capital.

The youngest son of Stefan Nemanja became a monk at Mount Athos, under the name Sava, and in 1219 was created archbishop of Žiča, near modern Kraljevo, where an autocephalous Serbian church was separated from the Bulgarian-influenced archbishopric of Ohrid. He was later canonized as St. Sava. To escape the constant harassment of raiding parties of Tatars, however, the seat of Nemanjid ecclesiastical order was moved to Peć. In 1375 it was elevated to a patriarchate.

Under Stefan Dušan (reigned 1331–55), the ninth ruler in the dynasty, the Nemanjid empire attained its greatest extent, incorporating Thessaly, Epirus, Macedonia, all of modern Albania and Montenegro, a substantial part of eastern Bosnia, and modern Serbia as far north as the Danube.

The achievements of the Nemanjid dynasty

Ironically, it is conceivable that the greatest achievement of the Nemanjid dynasty was not its territorial expansion but its success in developing for the first time a unified "high culture" for all Serbs, based largely around religious cohesion. The court was committed to the Eastern church, acting to suppress Bogomilism and ending attempts at the Latinization of the western areas. Many churches and monasteries were built that have remained among the architectural glories of the Eastern church; Mileševo (*c.* 1235), Peć (1250), Morača (1252), Sopoćani (*c.* 1260), Dečani (1327), and Gračanica (1321) are the most renowned. The frescoes of the Raška school are known for their capacity to blend a reverential sense of the awe in which secular authority is held with a deep sense of religious devotion. Literary work extended beyond the copying of a considerable number of manuscripts to include pieces of independent creative merit, such as the manuscript biography of Stefan Nemanja prepared by St. Sava and his brother Stefan. Courtly culture became a religious culture, and both church and state benefited from their close partnership. The ecclesiastical authorities acquired prestige and influence, while the court was given powerful symbolic support and was "civilized" in every sense.

During the 13th and 14th centuries the level of economic development rose, although during times of armed strife considerable damage was suffered by the population. Settled agriculture became typical of the entire Balkan Peninsula, with the remaining nomadic and seminomadic peoples becoming marginal. Crops such as hemp, flax, grapes, and oil-yielding plants became more widespread, and the plains of Kosovo and Metohija in particular became areas of dense population and fairly intensive cultivation, probably supporting more people than today.

Mining also grew considerably in importance. Copper, tin, silver, and gold had all been exploited in Roman times, but production intensified (often with the help of German-speaking colonists known as "Saxons," though they did not necessarily come from Saxony) as the demand for coin and luxury goods expanded in the new imperial courts and the centres of ecclesiastical authority. Stone castles and great fortifications, however, were relatively rare, and Western travelers and emissaries from the Byzantine court remarked on the modesty of the wooden palaces in which even the greatest Slav potentates lived at the time.

Expansion of trade

Trade expanded, particularly in the hands of Ragusan and Italian merchants, who led caravans along the old Roman routes. Administration improved; the high-sounding titles adopted by officials ("despot," "caesar," or "sebastocrat") were more than mere mimicry of Byzantium. An important step in the direction of separating administration from the personal whim of the ruler was taken by Tsar Dušan, who in 1349 promulgated his Zakonik, or code of laws.

Macedonia. The status of Macedonia has remained one of the most intractable issues in the Balkans throughout the 19th and 20th centuries. Bulgarians have always claimed that "Macedonians" are culturally, and especially linguistically, indistinguishable from their own people. Since 1945, however, the recognition of a Macedonian language and the granting of republican status to Macedonia within contemporary Yugoslavia has been founded on claims about the specificity of Macedonian culture. The assertions of both sides are grounded ultimately in interpretations of events in the early medieval period.

The Bulgarian view dwells on the lack of clear distinctions between early variants of the old Slavonic languages, explaining those developments that have been regarded as peculiar to Macedonian language as reflecting a process of Serbianization (dating, indeed, to the Nemanjid period), as well as more recent and politically motivated interference. The state of Tsar Samuel is regarded by Bulgarian historians as having been essentially continuous with earlier and later Bulgarian empires. The links between the ecclesiastical centre of Ohrid and successive Bulgarian secular powers are also stressed.

The Yugoslav view challenges these assumptions. The antiquity of the name Macedonia as designating a separate regional entity is underlined, and indications of the continuous development of linguistic specificity are central to the argument. The empire of Samuel, with its capital in Ohrid (not Preslav) is interpreted as having been a result of rebellion against Bulgarian domination.

These points can in no way do justice to the extent and complexity of the historical arguments on both sides, and advocates of each position can be found not only among Bulgarian and Yugoslav historians but also among independent scholars. In all probability the arguments are unresolvable because of the difficulty in making solid historical constructions on the basis of material that is inevitably relatively scarce and intrinsically ambiguous. In addition, to question the national identity of the people of the region during this period is anachronistic. "Nationality," in anything like its modern sense, is a cultural artifact of the French Revolution and after. It is of vital importance, however, to an understanding of the more recent history of the Balkans, to realize that the historical origin myths of the region's peoples may be no less significant because they are disputed by historians.

THE OTTOMAN CONQUEST

The Ottoman Empire had its origins in a small Turkish emirate established in the second half of the 13th century

in northwestern Anatolia. The name Ottoman is derived from Osman I, the chieftain who ruled the emirate from 1281 to 1324. His policy of aggressive expansion soon brought under his control large tracts of Asia Minor, and in 1354 his successors gained a foothold on the European mainland, on the Gallipoli Peninsula. By 1363 Adrianople (modern Edirne) had fallen, and it soon became the Ottoman capital, whence they moved up the Vardar and Morava valleys. The great military advantages possessed by the Turks were their light and highly mobile cavalry and their superb organization. Although these advantages brought them rapid gains in open country, the immensely strong fortifications of Constantinople itself resisted the Turks until they acquired artillery. The city, and with it the Byzantine Empire, finally fell after a prolonged siege in 1453, by which time half of the Balkans had succumbed to the invaders.

In Skopje in 1346, Dušan was crowned Emperor and Autocrat of the Serbs and Greeks, the Bulgarians and Albanians, but his empire was short-lived. By the time of his death in 1355 the Turkish march northward had already begun. His successors were unable to sustain his achievements, and almost immediately the state began to disintegrate under rival clan leaders. The fall of Adrianople shocked the several factions into momentary unity under Vukašin, the king of the southern Serbian lands, and his brother John Uglješa, the despot of Sérrai, but their forces were defeated in 1371 at Chernomen (Chirmen, Černomen), on the Maritsa River, where both were killed.

The Ottoman conquest of the Balkan Peninsula was not a smooth progression. Slav leaders were not infrequently willing to ally themselves with the Ottomans in the hope of securing aid against rivals. In this way they were able to retain a nominal independence for some years in return for a variety of forms of vassalage. (One of the most celebrated of these leaders was Marko Kraljević, the son of Vukašin and a chieftain of Prilep, who is immortalized in many of the heroic folk ballads of Serbia and Macedonia.) Many localities changed their allegiance several times before the indigenous nobility was finally ousted by the Ottomans. For this reason it is not easy to document the progress of Turkish domination with any exactness, although Sofia was captured in about 1385 and Niš in 1386.

Ottoman expansion was not without serious setbacks. In 1387 or 1388 a combined force of Serbs, Bosnians, and Bulgarians inflicted a heavy defeat on the Ottoman army at Pločnik. Turkish advance through Albania was relatively rapid, as many of the local inhabitants enlisted Ottoman aid against their Slav overlords, but the attempt to push farther north into Bosnia was at first unsuccessful. The turning point came when the Bulgarian tsar Ivan Shishman broke with the alliance of Slavonic powers and accepted Ottoman suzerainty. No longer threatened from the east, the armies of Sultan Murad I were able to concentrate their weight against Serb resistance. Led by the Serb knez Lazar Hrebeljanović (he did not claim Dušan's imperial title), a combined army of Serbs, Bosnians, Croats, Albanians, Bulgarians, and Walachians met Murad's forces in battle. On St. Vitus' Day (Vidovdan), June 28 (June 15, old style), 1389, on the Kosovo Polje, the Serbs and their allies suffered a defeat that has become hallowed in several great heroic ballads. The vision of Lazar on the eve of the battle, the alleged betrayal by the Bosnian Vuk Branković, and the killing of Murad by Miloš Obilić have been given assured immortality in Serbian folk literature. This material still provides the emotional focus for contemporary Serb-Albanian struggles over the political status of Kosovo.

The Battle of Kosovo

Although the defeat at Kosovo has generally been regarded as setting the seal on Ottoman occupation of the Balkans, its outcome was by no means certain. There had been severe Turkish losses in addition to that of the sultan; and while they were forced to accept the position of vassals to the Turks, the Serb despots continued to rule a diminished state of Raška, at first from Belgrade and then from Smederevo. At the end of the 14th century the Ottoman Empire was threatened again from the east, this time by the invading Mongolian forces of Timur (Tamerlane), and Sultan Bayezid I was captured at the Battle of Ankara in 1402. The despot George (Đurađ) Branković and the Hungarian János Hunyadi together waged a vigorous campaign during the middle years of the century (the "Crusade of Varna"). They reached the Dardanelles, but the plan to reinforce them with the Venetian fleet failed, and effective resistance was ended by their defeat at Varna in 1444. Serbian resistance cannot be considered to have ended until the fall of Smederevo in 1459.

Turkish penetration into Bosnia dates from the confusion that followed the death of Tvrtko in 1391. At first it took the form of raids for plunder and slaves, and for a time the Turks were bought off by the local potentates with tribute. The fall of Constantinople in 1453, the final extinction of the Serbian state, and the defeat of the rebellion of Skanderbeg in Albania (1444–68) removed other impediments to Turkish advance. In 1460 a campaign was launched against Stefan, the last duke of St. Sava. The fundamental weakness of Bosnia lay in its religious structure. Stefan had renounced the Bosnian church in favour of Roman Catholicism, but Bogomilism remained strong, particularly among the peasantry. Both Roman Catholic and Orthodox powers had conducted sustained campaigns of persecution against the Bogomils, and Turkish promises of freedom found a responsive hearing among them. The Ottoman expansion in the Balkans was not motivated by the desire to seek converts to Islām; it is generally agreed that in this early period their principal concern was plunder. Large numbers of Bogomils accepted Islām, however, and they were followed by a significant proportion of the aristocracy, who saw in conversion the opportunity to retain their lands and titles. Consequently, during the 15th century Bosnia developed an unusual social structure in the Balkan Peninsula, having a high proportion of its indigenous Slavic population at all levels in society who were Muslims. The Christian religion (generally Orthodox) was mainly confined to the peasants. Thus Ottoman conquest in this region put down roots among the South Slavs that never developed elsewhere.

Ottoman progress in the mountainous areas of the west was slow and incomplete. In Montenegro, Hercegovina, and the areas around Jajce and Srebrenik local chieftains retained independence for a long time. (Srebrenik submitted in 1520, and Jajce in 1528.) Intermittent uprisings by *baši-knezovi* ("warlords") and bandits remained a problem for the Ottoman overlords. It has been the proud boast of Montenegrins that the tribal inhabitants of their remote and rugged land never surrendered to the Ottomans. Although their chieftains occasionally paid tribute, it is true that Turkish control of this highland region was never secure.

In the latter part of the 15th century Ottoman expansionist aims were directed more toward Syria and Egypt. This was made possible partly by the fall of Constantinople and the demise of the Byzantine Empire as a sea power. The Turks founded their empire initially upon military superiority on land, but they rapidly adapted themselves to emerge as a major naval force, initially through the skills of Arab and Greek seamen. During this period there was a pause in their advance northward into Europe. Maritime ambition brought them into conflict, however, with Venice.

In 1463 a war began with Venice; it was the first of four major conflicts between the two powers that lasted to 1573. The course of these conflicts was one of intermittent attrition rather than spectacular victories, but by 1479 the Venetians were forced to abandon most of their holdings along the Albanian coast. Several of the Adriatic cities managed to retain their independence much longer: Alessio and Durazzo until 1501, and Dulcigno and Antivari until 1571. Budua remained in Venetian hands until 1797, and the Ragusan republic survived until the arrival of Napoleon's armies on the Adriatic coast. Nevertheless, by the middle of the 16th century the Ottoman Empire had emerged as the most powerful naval force in the eastern Mediterranean.

An equally important factor in the decline of Italian influence along the Adriatic coast in this period was the gradual Slavicization of the population, as refugees from the Ottoman advance in the interior moved to the coastal

The decline of Italian influence

cities. The Islāmic impact on the culture of the region was superficial, even in those areas that were subjected to Ottoman rule, and there continued to develop throughout the Middle Ages a chain of centres of distinctive mercantile civic culture along the Adriatic littoral. In large measure the success of the cities depended on their trading function, as commerce was of little interest to the Ottomans anywhere in the empire. Whereas in the interior of the Balkans merchants were typically Jewish, Vlach, or Greek in origin, along the coast it was possible also for Slav and Italian families to acquire wealth and power through trade and finance. They were able to advance their own independence by astutely playing off Venetian against Turkish interests.

The final victory of the Turks in the Balkans came about as the result of important cultural and political shifts in western Europe. In spite of their many differences, until the middle of the 16th century the various European rulers continued to defer to the notion of a united "Christendom": a Christian civilization that stood in necessary opposition to the enemies of the faith. The growing strength of dynastic interests, crystallized by the new religious divisions between Protestantism and Roman Catholicism, seriously undermined any reality that may have lain behind that ideology. This came to a head in the conflict between France, under Francis I, and the Germanic Holy Roman Empire under the Habsburg Charles V. In 1525 Francis appealed to Sultan Süleyman I to enter an alliance and to attack the Habsburg lands. His invitation was accepted, inaugurating an alliance that lasted more than three centuries. The first fruit of this "sacrilegious union of the Lily and the Crescent" was an attack launched against Hungary in 1526. The Hungarians and their Croatian allies were completely routed at the battle of Mohács, and with little other resistance the Turks were able to overrun the greater part of the Pannonian Basin and Slavonia (the lands between the Sava and Drava rivers). A truncated kingdom of Croatia remained, attached to the crown of Ferdinand of Austria, that consisted of little more than the districts around Zagreb, Križevci, and Varaždin.

By the middle of the 16th century Ottoman rule was established securely over the area that later became Yugoslavia. The wave of Ottoman advance passed on and was only turned back short of Vienna in 1529. Ottoman victories had previously been largely assured by the supremacy of their light cavalry. The rapid advances in the development of European artillery, together with the development of military tactics involving the use of disciplined bodies of infantry armed with muskets and pikes, changed the balance of arms.

The process of rolling back the frontiers of Ottoman expansion began in 1593, when a raid by the Turkish governor of Bosnia into Croatia resulted in his sound defeat and a widespread outbreak of fighting. With the support of the Habsburg crown, the prince of Transylvania, Sigismund Bárthory, in 1594 joined with the Walachian prince Michael the Brave to create an anti-Ottoman alliance. The campaign secured no significant revision of borders and Walachia was confirmed as an Ottoman dependency, but the parts of Hungary formerly under Ottoman control were no longer required to pay tribute.

Decline of the Ottoman Empire

The first important reverses in Turkish fortunes in this region were not to come for another 90 years, during which time the Ottoman Empire began to suffer from the internal processes that led to its decline. In 1683 a rebellion in the Habsburg portion of Hungary was seized upon by the grand vizier Merzifonlu Kara Mustafa Paşa as a pretext for further territorial expansion. The rebels appealed to the sultan Mehmed IV for aid against the Habsburg emperor. The French king Louis XIV, hoping to turn the disorder in the Habsburg lands to his own advantage, urged acceptance of the appeal, and Kara Mustafa crossed the Danube with a large army. There was little Austrian resistance barring his way to Vienna, and in July the Turks laid siege to the city with vastly superior numbers, the imperial court having fled. By September the emperor had managed to secure an alliance with the Poles, and the intervention of a Polish army under King John III Sobieski resulted in Kara Mustafa's defeat. On his return to Belgrade he was beheaded by order of the sultan.

Sobieski made limited progress in liberating the Danubian provinces, but Austrian troops, aided by Serbian and Albanian irregulars, captured Belgrade and pursued the retreating Turks deep into Serbia and even into Macedonia. The new grand vizier, Köprülü Fazıl Mustafa Paşa, launched a successful counteroffensive in 1690, driving the invaders back across the Danube.

The war, together with the wider European conflagration, dragged on, and in 1697 the sultan Mustafa II launched a fresh offensive northward from Belgrade. He was met at Zenta (near Subotica) by an Austrian army under the extremely able generalship of Prince Eugene of Savoy. The mutiny of Mustafa's Janissaries during the battle contributed to his disastrous defeat. So great were the Ottoman losses, especially in artillery and treasure, that Sultan Mustafa decided to sue for peace. The progress of the settlement was delayed by the complexity of the wider European interests involved, but the outcome was the Treaty of Carlowitz (modern Sremski Karlovci), signed in January 1699. Within its extensive provisions, the Habsburgs gained Transylvania, Croatia, Slavonia, and all of Hungary except the Banat of Temesvár. Venice made extensive territorial gains along the Dalmatian coast. The recession of the Turkish tide in the Balkans had begun.

SOCIETY AND ECONOMY: THE 14TH TO THE 18TH CENTURY

The Ottoman lands. When the Balkan peoples fell under Ottoman control they became a part of one of the great empires of world history. The negative opinion often held of Ottoman civilization, reflected in such phrases as "the sick man of Europe," is based principally on judgments made of conditions in the 18th and 19th centuries, by which time the state was in obvious decline.

Although the empire is usually referred to as either the "Ottoman" or the "Turkish" empire, it is important to remember that it was not a national state. Political divisions among the populations within the empire, and its political organization, were drawn along religious, and not national, lines. All Muslims were regarded as belonging to a single community of the faithful, the *ummah,* and any person could join the ruling group by converting to Islām. Peoples who were outside the *ummah* were regarded as excluded by religious faith rather than by language or any other characteristic, and each of these non-Muslim religious communities was called a *millet.* Ottoman administration recognized five such groups: Orthodox, Gregorian Armenian, Roman Catholic, Jewish, and Protestant. Each group was under the direction of its religious head. Thus for the majority of the Balkan peoples, being Orthodox, their titular head was the patriarch of Constantinople. With the passage of time, however, the increasing importance of national consciousness was recognized by the Ottoman authorities, and Constantinople became a specifically Greek centre. The Serbs had their own patriarchate at Peć, and the Bulgarians a metropolitanate at Ohrid. These ecclesiastical authorities were expected to assume many civil functions, including the administration of justice, the collection of taxes, and later also education.

The *millets*

At the centre of the system was the sultan and his court—often referred to as "the Sublime Porte" (or simply "the Porte")—based in Constantinople. The origins of the empire in conquest were reflected in its administrative structure, which revolved around the extraction of revenues principally in order to support a military caste. All authority and the right to enjoy possessions were regarded as deriving from the sultan, who "leased" them to subordinates at his own will and for his benefit. There were three types of unit (*mukataa*) that could be bestowed in this way (until the *Tanzimat* reforms of the 19th century): the *tımar,* the *iltizam,* and the *emanet.*

The *tımarlı* held the right to raise revenue from taxes in their area. The taxes were used primarily for their own support, very little of the proceeds returning to the central treasury. Typically, the holder of such a position was a spahi (*sipahi*), or mounted warrior, and from his territory he was expected to support and arm himself in a state of readiness for the service of the sultan. The larger estates were operated on a similar basis but were expected

to support larger numbers of cavalry. The spahi was not a landowner, in that he held his rights at the pleasure of the sultan, and certainly not a farmer, but he was rather a professional soldier supported by the yield of agriculture, industry, or trade in his territory. This was the principal system in operation throughout the Balkans.

The *iltizam* was a kind of "tax farm," the income from which was partly the salary of the holder and partly revenue to be passed on to the Porte. Unlike the spahi, the tax farmer (*mültezim*) was not expected to provide military service and was more typical of the eastern regions of the empire. The *emin,* or steward, of an *emanet* was simply a salaried revenue collector and was found mainly in the larger cities.

The empire was divided into regions and subregions. The Balkan Peninsula became the *beylerbeylik* of Rumeli, with its seat in Sofia. Below the *beylerbeylik* was the *paşalık* (pashalic) and below that the *vilayet.* At the lowest level, a group of spahis were assigned to a *sancak beyi,* "standard-bearer." (A *bey* was a governor or administrator.) In addition to this structure of landholding, the country was divided into judicial districts, or *nahiye.* In the 17th century there were about 40 *nahiye* in the part of the empire that is now in Yugoslavia.

The status of Christians

Initially the granting of *tımar* holdings was extended to former Christian nobles who accepted Ottoman suzerainty. This was a very unstable situation, however, and early in the 15th century it became necessary to be a Muslim to enter the imperial service. In the middle of the 16th century, as the empire expanded to include Arab areas, their staunchly Islāmic chieftains (*ulema*) began to insist on even more rigorous religious segregation, and the entire Muslim *millet* was accorded a higher status.

The situation of the Christian population (the *reaya;* literally "flock") was not one of unmitigated oppression. Christians were exempted from military service, and in some regions the tax burden was lighter than it had previously been, although they were taxed more heavily than the Muslim population. Christians were bound by certain sumptuary laws. Although churches could be repaired, new ones could not be built, and they could not have belfries or towers, which might "offend the eyes of the faithful."

It was possible for subject peoples to rise, on condition of their conversion, to the highest positions in the system. Several grand viziers were born among the subject peoples of the Balkans: for example, Mehmed Paşa Sokollu (Sokolović; served 1565–79) was of Serbian origin, Ibrahim (served 1523–36) was a Greek, Rustem (served 1544–53) was Bulgarian, and Koçu Bey, who was vizier to Murad IV in the 17th century, was Macedonian. There were three routes by which such individuals could enter the sultan's service: by capture as a slave in war, by voluntary conversion to Islām (as with the Bogomils), and by recruitment through the system known as *devşirme.* This last, by far the most typical, involved the conscription of Christian boys between the ages of 10 and 20; conscription took place approximately every five years. The boys were taken to Constantinople, forcibly converted to Islām, and employed in a variety of posts throughout the imperial service. In parts of the empire and at times of particular need this could involve up to a quarter of the boys in the appropriate age group. The most able would be trained for administrative positions, while the others joined the corps of *yeniçeri* (Janissaries). The Janissary corps was an elite, celibate order of infantrymen that, as firearms became a more significant factor in warfare, came to be the most effective part of the Ottoman military.

Ottoman society was principally rural in character, the majority of the population living on small, mixed farms that produced little marketable surplus or in small pastoral communities. There were few large towns, although some administrative, military, and mercantile centres, such as Sarajevo, Monastir (Bitola), and Usküb (Skopje), did grow up. Trade and manufacture were not particularly encouraged by the Ottomans. Their principal concerns were with the extraction of revenue through taxation and the maintenance of order. Commerce was regarded only as a possible source of excise duty. The greater part of trade was carried on by ethnic minorities—itinerant strangers often called indiscriminately *cincari*—who were treated with suspicion. The guilds were seen as necessary, principally as a means of controlling rather than developing craft production. Finance was inherently a dubious affair in view of the Islāmic prohibition of usury.

Cultural life

Levels of literacy remained low for the indigenous peoples. A few had a little Greek—the lingua franca of trade—and knowledge of Old Church Slavonic was mostly confined to the clergy. Culturally, therefore, the population remained highly differentiated, living most of their lives within the confines of local peasant communities, with their own dialects—the vehicle for folk song and poetry—dress, and customs. Some, such as the traders who moved throughout the Balkans with their caravans and the itinerant Macedonian masons and woodcarvers, traveled widely, but the majority inhabited a very small world. Even the Muslim population did not develop a great metropolitan culture in the Balkans outside Constantinople. The bridges, mosques, baths, fortifications, and even the houses of the aristocratic families were relatively humble in comparison with the great achievements of Islāmic art elsewhere.

A feature of Ottoman society that had important consequences for the subsequent development of Yugoslavia was the ethnic dispersion of the population. The wars of conquest resulted in several areas of depopulation. Large numbers moved, hoping at first to escape from Turkish rule. In regions that were thinly peopled the Turks settled immigrants in order to ensure the support of their spahis. Some ethnic groups acquired a reputation for association with specific occupations and came to be widely dispersed; for example, Albanians came to be associated with the police and with silverwork and Greeks and Vlachs with trade.

Even as early as the reign of Sultan Süleyman I (1520–66), the empire began to be afflicted by several internal problems that were to undermine this system. As firearms became more sophisticated and more effective, the relative military significance of light cavalry began to decline, and the rationale behind a system designed to support mounted warriors weakened. All types of *mukataa* had initially been held entirely at the pleasure of the sultan and had reverted to him on the death of the holder. As the holders of office became established as settled aristocratic families, however, they began to demand hereditary rights for their children. The territorial expansion of the empire was brought to a halt during the 17th century, which had particularly important consequences for a structure initially founded upon booty. The need for a large, completely dedicated, and highly mobile corps of Janissaries was greatly reduced, and with the loss of their specifically military function they acquired the right to marry. (The extent of change is suggested by the abandonment of the *devşirme* system around the middle of the 17th century.) The Janissaries began to look for opportunities to obtain land or office. The declining flow of booty shifted the burden of the revenue needs of the empire onto the system of taxation. The result was a growing number of potential functionaries who competed for a limited number of offices, which in turn were expected to yield a higher rate of income. Both state revenues and the income of the officials were adversely affected during a long period of inflation. These developments led to both a steady rise in the level of exactions from the Christian population, through a spread of tax farming, and a growth in the number of holders of former *tımarlı* who tried to turn their holdings into agricultural estates. This latter development, which came to be known as the *çiftlik* system, led to two types of estates, one operating by a variety of sharecropping and the other as commercial latifundia. The gradual expansion of commerce and the introduction of new crops (such as cotton) saw a consistent movement toward the latter type.

Changes in landholding

The disintegration of the old system, therefore, brought with it growing dissatisfaction on the part of the Christian population and an increasing number of protests against economic injustice. These protests sometimes took the form of armed uprisings by the peasantry; uprisings were particularly common in the northern areas, where imperial control was weakest and the Janissaries least disciplined. The greatest of these uprisings took place in 1690, when Serbs rose in support of the Austrian invasion. The re-

treat of the Austrians left the native population seriously exposed to Turkish reprisals, and in 1691 Archbishop Arsenije III Crnojević of Peć led a migration of 30,000–40,000 families from Old Serbia and southern Bosnia across the Danube. The spread of Albanian Muslims into the vacant lands at that time was to provide a continuing source of communal tension. The second common response to oppression was to resort to banditry, and the figure of the *hajduk,* preying on Turks and merchants from his mountain lair, has yielded a rich vein of material for folk literature throughout the region.

The Habsburg lands. *The military frontier.* No single pattern of economic and social relations emerged in the Habsburg lands before the 19th century. Carinthia, Styria, and Carniola (inhabited in part by the Slovenes) were directly attached to the Austrian crown. The Croatian lands, on the other hand, belonged to the Hungarian half of the Dual Monarchy. The status of many areas of Dalmatia was complicated by the legacy of Venetian control in both civic organization and landholding. The area today known as the Vojvodina was an integral part of Hungary.

A prominent feature of Austrian and Hungarian lands was the "Military Frontier" (*Militärgrenze*), which dated from the first attempts to reclaim Croatia and Slavonia from Turkish rule. In 1578 the Austrians began to organize frontier marchlands against the Turks with a system of forts, watchtowers, and warning beacons staffed by a force of *graničari* (frontiersmen). The *graničari* were granted land in frontier regions in return for military service. The system was reorganized and consolidated throughout the next two centuries, especially after the treaties of Carlowitz (1699) and Passarowitz (1718). At its height it extended in a belt of varying widths across the entire northern border of the Ottoman Empire from the Adriatic coast to Walachia and was organized into three "generalates."

Migrations

Having been badly depopulated by Turkish raids, the Military Frontier was resettled largely by refugees of a variety of ethnic origins, but Serbs and Bosnians contributed a large proportion. Particularly important was the great migration of 1691, as a consequence of which the region came to contain some of the major centres of Serbian culture. In 1766, when the patriarchate of Peć was abolished in response to a campaign to Hellenize the church within the Ottoman Empire, control of the Serbian church passed to the Greek patriarch. The ethnic map of Yugoslavia today bears the marks of these migrations.

The Military Frontier was unpopular with both the Croatian and the Hungarian diets, who saw in the direct allegiance of the *graničari* to the emperor a threat to their own authority. The regiments defended their role both vigorously and effectively, however, and the Military Frontier was not absorbed into civil administration until 1873–81.

Mainly because of the prevailing insecurity caused by the region's closeness to Turkish land, large towns were inhibited in their development. Landholding was organized through regiments, with holdings widely dispersed among the families of the frontiersmen. Consequently, no large estates emerged in the region, and feudalism in the narrower sense was foreign to it. The traditions of the Military Frontier (and in particular a strong sense of loyalty to the emperor in Vienna) long outlasted its formal abolition, in spite of the constitutional link between Croatia and Hungary.

Croatia and Slavonia. In Croatia and Slavonia outside the Military Frontier the key to social organization lay in the relationship of the Croatian to the Hungarian crown. The chief concern of the Croatian nobility in the Pacta Conventa of 1102 (the agreement by which King Koloman of Hungary and his successors were recognized as rulers of Croatia) had been to resist the erosion of their rights in favour of the royal prerogative, which King Zvonimir had been pursuing. Because powers and lands reverted to the Hungarian crown, however, it was possible over time to appoint Magyar nobles to great feudal titles in Croatia itself. While stoutly defending their constitutional independence, therefore, the Croatian aristocrats found themselves gradually merging socially with their Magyar peers. The system of "donational" feudalism that prevailed in Hungary gave enormous personal rights to the aristocratic landholder, and large latifundia developed that were similar to those in Hungary proper. In some areas, known as urbarial lands, the serfs were attached to the estate; while this meant that on their estates the landed magnates exercized all the powers of the state, the peasants could not be evicted and had certain rights that the lord was bound to respect. Alodial lands, on the other hand, were the unlimited property of the landholder, and they were typically worked either by the corvée labour of peasants settled on the urbarial lands or by tenants at will who paid rent in money or kind but had no protected rights. The position of the peasants varied somewhat in different regions and under different types of landholding, but in general they were bound both by taxes and by the obligation to work the land of the feudal lords and they were expected to be available for military service. The peasants also paid tithes. One of the largest owners of land throughout Croatia was the church, and the bishop of Zagreb was one of the greatest magnates in the land.

The position of the peasants

In 1671 the Croatian nobility suffered a severe setback when an attempt to free Croatia from Habsburg rule led by Counts Zrinski and Frankopan ended in their execution. Their considerable estates were confiscated and largely distributed to non-Croatian royal servants.

Relations between lord and peasant periodically deteriorated into open rebellion as the latter protested against the infringement of their rights or the level of exactions. The most famous of these rebellions was led by Matija Gubec in 1573. The bloody repression of the rebellions frequently led to the intervention of the crown, as in the *Robot-patent* (forced-labour patent) of 1680, which sought to regulate corvée obligations, and the *Urbar* ("land register") of Charles VI in 1737, which attempted to systematize the definition of peasant rights. The empress Maria Theresa (reigned 1740–80) and her son Joseph II (reigned 1765–90) made repeated efforts to reform the system. In particular Joseph made vigorous efforts to control church lands, and in 1785 he declared the peasants to be personally free; a decree of 1789 abolished the corvée. Both of these monarchs were enthusiastic Germanizers as well as enlightened reformers, and as a consequence they incurred the hostility of both Croatian and Magyar nobles. Many of their legislative measures remained in effect dead letters.

The Vojvodina. The area of modern Yugoslavia known as the Vojvodina does not have any clearly defined historical existence. The name only came into use after 1849 to designate a province that included part of the former Banat of Temesvár, most of Bačka, and a small part of Baranja. Even so, it did develop specific characteristics before that time which are worthy of comment. With the exception of the short period during which Hungary was overrun by the Turks, the whole of the region was an integral part of the kingdom of Hungary. Even during the time of Turkish occupation, it began to receive Serb migrants, and these increased in importance after the Ottomans were forced back across the Danube. At this time the population was almost entirely Serb or Romanian. Colonization in the southern part took the form of the establishment of a military frontier along the lines discussed above, but in the north large latifundia were created under Magyar nobles and ecclesiastical bodies, who also introduced large numbers of peasant colonists from the Rhineland and Upper Austria, adding further to the complexity of the ethnic mix. Outside the military frontier, the situation of the peasants resembled that in the urbarial lands in Slavonia.

Dalmatia. Between 1684 and 1714 control of Dalmatia gradually fell to Venice (especially through the treaties of Carlowitz and Passarowitz), and it remained under Venetian control until the Napoleonic Wars. As rewards for military service against the Turks, the Venetians awarded grants of land that ranged in size from large estates to peasant farms, establishing a kind of military frontier. Agrarian affairs were regulated by the Lex Grimani of 1756. The state remained the owner of the land, granting use of it in perpetuity to the landholder in return for the payment of a tithe. The land could be sublet, and in fact the system of tenancy known as the "colonate" soon came to predominate. The colonus (as the name suggests, the arrangement dated from Roman times) was a sharecrop-

Land tenure under the Venetians

per, normally paying a third of the crop in rent. There were variations in the system, depending on the region, with the trading cities often developing their own agrarian law. The republic of Ragusa was known for a particularly burdensome version of the colonate.

The extreme scarcity of arable land along the coast meant that the life of cultivators was hard, and agriculture was possible in many areas only after the laborious construction of stone terraces extending up the steep hillsides to retain precious fertile soil. Many farmers supplemented their diet and income by fishing.

The cities that continued throughout this period to develop their positions as trading centres sometimes became wealthy, and they were the cradle of developed urban civilizations, though during the 18th century they suffered a slow economic decline. Not only did these cities create the civic, military, and ecclesiastical architecture for which they are still known, but even in the 17th century they supported a rich literary culture. Ethnically the littoral became highly diverse, with many Italians and Serb refugees from the Turks mingling with the Croatian majority. Islāmic cultural influence on the coast was negligible.

Slovenia. The Slovenes never developed their own state, and consequently they did not have an indigenous aristocracy at the time of the incorporation of their tribal areas into the sphere of Frankish, Germanic culture in the 8th century. Subsequently they were subjected to steady pressure to Germanize their culture. The Slovenes never lived under Ottoman rule, although for some time they were subject to the threat of their raids.

Slovenia is the only part of contemporary Yugoslavia to have experienced fully both the Reformation and Counter-Reformation, which left the area with a strong commitment to Roman Catholicism. The church has for several centuries played a key role in the preservation and development of a distinct Slovene culture. On the whole the towns succumbed to Germanization and the countryside remained firmly Slovene, creating a clear and permanent political, economic, and social cleavage.

Economically the Slovene lands were soon incorporated fully into the pattern of Germanic feudalism, and no trace of the *zadruga* organization of the Slavs remained for long. The topography of the region militated against the development of large-scale agriculture, and the larger feudal estates typically contained substantial areas of forest. Cultivation was confined in the main to peasant holdings. The typical Slovene village was occupied by a core of relatively wealthy peasants (*kmetje* or *gruntari*), who farmed the best land and the largest holdings, and poorer cottagers, who had to supplement their living by earnings of other kinds, including service to their better-off neighbours. Peasant rights were at times defended only with difficulty, and there are records of several peasant uprisings, including substantial Slovene participation in the revolt of Matija Gubec. Generally speaking, however, the Slovene lands escaped much of the economic and political upheaval that affected life among other South Slav peoples. As a consequence of this, and of their greater proximity to the major urban and economic centres of the Habsburg empire, the Slovenes reached relatively high levels both of literacy and of technical development and an early integration into a market economy.

Peasant uprisings

THE PASSING OF THE OLD ORDER

The French Revolution signaled the beginning of the transformation of the feudal order throughout the Balkans. The wars of this period precipitated changes in international relations and in the balance of power that ultimately led to the formation of the South Slav state in 1918. The spread of western European social and political thought after 1789 had profound effects on Balkan culture.

The creation of South Slav states. From the end of the 17th century the Dalmatian coast had been under Venetian control, but its territories were divided between France and Austria by the Treaty of Campo Formio (1797) and Venice ceased to exist as an independent state. Dalmatia came first under Habsburg rule, but in 1805 Napoleon acquired Dalmatia, and in 1809 he gained control of Slovenia, Istria, Trieste, and parts of Croatia, creating the Illyrian Provinces. The Ragusan republic was abolished in 1808. Austria-Hungary regained effective control of the Illyrian Provinces in 1813, and they were formally ceded by France at the Congress of Vienna in 1815. The brief period of French occupation had a profound impact on the politics and culture of the area. Serfdom was ended and the peasants were given their land, the Code Napoléon replaced local medieval laws, administration was centralized, finance was controlled, roads were built and land reclaimed and reafforested, and the use of the Croatian language was encouraged. Many of these changes did not survive the return to Habsburg rule, but the Illyrian period contributed significantly to the development of Croatian national awareness.

The Napoleonic period also had a great effect in Montenegro. The surrounding regions were overrun by the Turks at the end of the 15th century, leaving a tiny embattled kingdom centred in Cetinje. The Crnojević dynasty came to an end in 1516, and rule passed to the bishops of Cetinje, elected by local assemblies and consecrated by the patriarch of Peć. Although Cetinje itself was captured by Ottoman forces in 1623 and again in 1687, the occupiers soon learned the truth of the adage that "in Montenegro a small army is beaten, a large one dies of starvation," and they contented themselves with the occasional collection of tribute. After 1696 the office of bishop (*vladika*) became hereditary in the Njegoš dynasty, from which came the rulers of the realm until 1918.

The Russians courted the *vladikas*, seeking to extend their own territory at the expense of the Ottoman Empire and looking for every potential ally who could harass the Turks. On two occasions at the end of the 18th century, Montenegrin forces rose loyally to attack the Turks. During the Napoleonic Wars the *vladika* Peter I also sided with the Russians, resisted the French occupation of Ragusa, and for a time occupied the port of Kotor. The Congress of Vienna returned Kotor to Austria, but with Russian backing Montenegro emerged from the Congress with its territory virtually doubled. Turkish attempts to recoup their losses were repulsed in 1819–21 and 1828–29. More significantly, the country was confirmed as an independent state in international treaties, and the alliance with Russia was firmly cemented.

Influence of Russia

Growing Russian interest was also a factor in the creation of an independent Serb state. The disintegration of Ottoman feudalism produced a highly unstable situation in northern Serbia by the middle of the 18th century. The local Janissaries were virtually beyond the control of the Porte, and their exactions had passed from the collection of taxes to open plunder. The Austrian wars against Turkey in the late 17th century had already successfully exploited local feeling against Ottoman rule. The Habsburg emperor Joseph II reached a secret agreement with Catherine the Great in 1782 for the eventual partition of the Balkan Peninsula. When war broke out between Turkey and an Austro-Russian alliance in 1787, Joseph called on the discontented Serbs to rise against their overlords, and this they did with some success. The treaties of Sistova (1791) and Jassy (1792) that concluded hostilities included a defense of Serb civil rights. The Janissaries were expelled from the pashalic of Belgrade, but they soon returned and a period of endemic political disorder set in.

In 1804 Serb impatience with Ottoman control came to a head, and an uprising broke out in the Šumadija region, south of Belgrade. It was led by George Petrović, called Karageorge (Black George), a successful pig trader, who had served with the Austrians in the war with Turkey in 1787–88. Although the uprising was initiated in the name of the sultan Selim III with the aim of reestablishing his authority, attempts were made to involve both Austria and Russia, and Russia did respond with financial aid and diplomatic backing. In 1805 a Skupština (Assembly) was summoned by Karageorge, and it submitted a list of proposals to the sultan. The proposals included a number of concessions to local autonomy that were unacceptable to the sultan, and a large force was sent to quell the rebellion. The Serbs continued to hold out, however, and they were strengthened by the arrival of Russian reinforcements in 1808. Threatened by Napoleonic invasion in 1812, the tsar

Alexander I concluded a treaty with the Turks. The withdrawal of Russia left the Serbs open to Ottoman reprisals, and by the end of 1813 Karageorge and the remainder of his followers were compelled to retreat across the Danube.

The return of the Turks was accompanied by a widespread reign of terror. Preoccupied with the business of the Congress of Vienna, the major powers showed little interest in the fate of the Christian population, which rose again in self-defense in April 1815, led by Miloš Obrenović. The revolt met with rapid success; the Turks were driven from a wide area of northern Serbia, and they were soon forced to negotiate. The fall of Napoleon meant that Russian interest was rekindled, and under threat of Russian intervention several important concessions were made to the rebels, including the retention of their arms, considerable powers of local administration, and the right to hold their own assembly. The region remained a Turkish principality, with a resident pasha and Turkish garrisons in the principal towns. In effect an independent Serbian state dates from this time.

The 1815 uprising

In June 1817 Karageorge returned from exile. He and Miloš had never enjoyed an easy relationship, and when Karageorge was murdered in mysterious circumstances Obrenović's complicity was suspected. Although Miloš was elected hereditary knez by the Assembly, the feud between the Karageorgević and the Obrenović families continued throughout the century.

The awakening of national consciousness. Moved by the idea expressed in the Illyrian Provinces of the common political and cultural interest of the South Slavs, the poet and philologist Jernej Kopitar published the first grammar of the Slovene language in 1808. He tried to apply the ideas of the great Serb linguistic reformer Vuk Karadžić, concerning the standardization of Slavonic orthography, to his own language and strove to eliminate its many Germanic accretions. Kopitar's ideas bore fruit in 1843 with the publication of the first Slovene-language newspaper in Ljubljana.

In Croatia also the Napoleonic era left a legacy of heightened national awareness. The return of the Illyrian Provinces to Austro-Hungarian control in 1815 brought with it a wave of Magyarization. Magyar replaced Latin as the language of state affairs and in the process threw into the foreground for the first time the lack of a common linguistic heritage among the various peoples attached to the Hungarian crown. Participation in the great multinational community of the Habsburg empire, however, meant that new ideas circulated relatively quickly among the intellectual elites of the various constituent nations. Polish and Czech ideas in particular were taken up by Ljudevit Gaj (1809–72). The seriousness with which he took the Illyrian ideal is evident in the title of the newspaper he founded in the 1830s, *Ilirske narodne novine* ("Illyrian National News"), and in his proposal that Serbian, Croatian, and Slovene should be abandoned in favour of a single Illyrian language. In both the Slovene and Croatian lands a flowering of literary creativity in the vernacular dates from the post-Napoleonic years.

One of the principal consequences of the wars for the Serbs was the extension and deepening of the channels of communication between the Serbs living in Serbia itself and those living in a diaspora across the Danube and throughout the Habsburg lands. The latter had prospered as traders, members of the free professions, and soldiers and in several cases had been accepted into the ranks of the nobility. There was therefore a substantial Serbian middle class in these areas that was lacking in the lands which had long remained under Ottoman tutelage, and this middle class played a crucial role in the growth of national consciousness.

Dositej Obradović (1743–1811) came from this group. Born in the Banat, he was trained in a monastery in the Fruška Gora but made a career as a philosopher and linguist. Attempting to introduce philosophical ideas to his countrymen in their own tongue, Obradović wrestled with the problems of standardizing a Serbian literary language. He was followed in this endeavour by Vuk Karadžić, who had participated in the uprising of 1804 and fled across the Danube with Karageorge in 1813. Karadžić went to Vienna, where he participated in the intellectual life of the South Slav community. His articles came to the attention of Kopitar, who was the imperial censor, and the two men became close friends. Karadžić then conceived a grand project for the creation of a Serb literary language, which included the revision of its orthography, the collection of songs, poems, folk sayings, and stories in the living language of the people, the compilation of a grammar and dictionary, and a demonstration that this language could be used as the vehicle for great literature. In the course of his life all the elements of this scheme were accomplished.

Vuk Karadžić

Karadžić's revised orthography abandoned 18 letters in the Old Church Slavonic alphabet that had no function in the living language. He borrowed one letter, *j*, from the Roman alphabet and devised five new signs to represent sounds of the Serbian language for which there were no existing letters. These proposals met with bitter resistance in ecclesiastical circles, but they were sympathetically received by influential secular intellectuals such as Kopitar, Obradović, and Gaj and have become the standard orthography of modern Serbo-Croatian and Slovene and, with minor revisions, of Macedonian. The alphabet's only drawback is that, because it is strictly phonetic, dialectal variants of the same word are spelled differently, lending fuel to the claim that these dialects should be recognized as distinct languages.

Karadžić's contacts with the other great figures in the development of the literary languages of the South Slavs helped to create a sense of cultural cohesion throughout the region that contributed significantly to the emergence of political unity. His work was not widely accepted in Serbia in his lifetime because of the strength of clerical support for the old literary forms, but the process of political unification that Miloš initiated was to have consequences which could not so easily be controlled. With the growth of political and economic cooperation between Serbs on both sides of the Danube and the Sava came the inevitable triumph of Karadžić's reforms.

CLASSES AND STATES: 1830–1912

Following the disruption of the Napoleonic period and its immediate aftermath, entirely new social and political processes began to shape the lives of the South Slav peoples. They remained overwhelmingly peasant societies, but the old chiefly and aristocratic dynasties were increasingly challenged by the rising middle classes, who saw "national interest" in different terms.

From this period the development of Dalmatia became bound up with that of Croatia-Slavonia. The economy of the region prospered under the new order, especially during the Crimean War, as demand for shipping boomed. The division of its elite into Italian, Slav, and Hungarian factions obstructed their united action to advance the cause of national independence. Although in 1860–61 the Austrians promised the eventual unification of Dalmatia with the other Croatian region, this was never accomplished. In the absence of a unified political voice, Dalmatia was unable to defend adequately its economic interests.

In Slovenia the national movement turned away from party conflict toward the development of new institutions expressing rural opposition to the Germanized towns. The most important of these developments was the cooperative movement. The first cooperative was founded in 1855, and a federation was formed in 1883. A second federation was created as the economic arm of the Slovene People's Party in 1895. These cooperatives provided credit, marketing, and other facilities, enabling both rural and urban Slovenes to break free from German institutions. The Slovene People's Party had close links with the Roman Catholic church, and together parish and cooperative formed a powerful force organizing and raising the national awareness of the people.

The cooperative movement

The revolutionary upheavals that swept many parts of Europe in 1848 had their counterpart in Croatia. The demand for national independence naturally brought the Croats and Hungarians into acute conflict, and the infringement of Croatian autonomy was only prevented when the ban of Croatia, Josip Jelačić, led a Croat–Serb army on behalf of the Habsburg emperor to quell the Mag-

yar revolution. The price was a deepening of the mutual hatred of Slav and Magyar, and there was no extension of real freedom.

The railways were extended into the region during the middle years of the century as tourist resorts, such as those at Opatija (Abbazia) and Bled (Veldes), were developed and as forest and agricultural products, and later mineral resources, were exploited to serve the growing industrial centres of the north.

Political opposition took the form of the vigorous nationalism of Ante Starčević's Party of Right, with its demands for the full restoration of what the party saw as the historical rights of Croatia. Hostility was crystallized in particular by the events of 1867–68 when, following their defeat by Prussia, the Habsburgs were compelled to reorganize the state and the Austrians and Hungarians made a compact, the Ausgleich (Compromise), that established the Dual Monarchy of Austria-Hungary. Under the terms of the Ausgleich the Slav peoples were divided, confirming the union of Croatia-Slavonia with Hungary and linking the Slovenes and Dalmatians to Austria, thus freezing any progress toward an Illyrian state. In 1868 the Hungarians and the Croats concluded the Nagodba ("Agreement"), which recognized the existence of the Croatian nation and the official status of its language, under the government of the Sabor (parliament) in Zagreb and a Croatian ban, although both were firmly subordinated to Budapest. This was made possible by a restricted franchise that gave only 2 percent of Croats a vote while enfranchizing Magyar officials and landowners.

Stalled on the political front, Croatian nationalism grew in strength and expression in the field of culture. In 1866 the great Croatian historian Franjo Rački set up the Yugoslav Academy of Sciences and Arts. Under the leadership of Bishop Josip Juraj Strossmayer of Đakovo, the University of Zagreb was reorganized in 1874.

Slav nationalism was not helped by its internal divisions. The abolition of the Military Frontier in 1881 brought large numbers of Serbs into civil Croatia. Extreme Croatian nationalists saw them as a threat rather than as potential allies against the Magyars, who had no difficulty in playing the Slav parties off against one another.

Creation of the Vojvodina

In the east the Serbs had been rewarded for their support of the emperor in 1848–49 by the creation of a semi-autonomous Vojvodina ("Duchy"). The Ausgleich eradicated this arrangement and exposed them also to the full force of Magyar attempts at assimilation. Extensive land reclamation was coupled with colonization by Magyar speakers; railway construction strengthened the economic ties with Budapest, and industrialization brought with it Hungarian entrepreneurs, technicians, and officials.

The Serbian state expanded steadily through its first half century, almost in spite of its rulers. In 1830 the Ottoman government granted the Serbian principality full autonomy, Miloš was recognized as hereditary prince, and the Serbian church was given independent status. In 1833 Miloš used the pretext of restoring order across the southern border to annex further territory. He attempted a program of domestic reform, but his tendency to behave like a pasha aroused great opposition. He abdicated in 1839, but neither of his sons (Milan and Michael) managed to control the dissenting chieftainly factions and gangs of bandits. A coup d'état in 1842 brought the Karageorgević family to power. The Skupština elected Alexander, the third son of Karageorge, as prince. Alexander's studied neutrality between Austria and Russia made him unpopular, and he was deposed in 1859. The aged Miloš was recalled from retirement, and in 1860 he was succeeded by his son Michael, who continued the work of consolidating the state and modernizing its administration. Michael was assassinated in 1868, probably by supporters of the Karageorgević dynasty. They did not reap the reward for their efforts, however, as the Skupština called his cousin Milan to the throne. Still a minor, and a highly Westernized young man, Milan took little interest in his task and was very unpopular. It may be said that he was saved by the Bosnian insurrection in 1875. Had he not declared war on Turkey in support of the Bosnian insurrection, he might well have been forced to abdicate.

In Bosnia the local Muslim nobility were more repressive of their Christian *reaya* than were Turks elsewhere, but they were also prone to challenge the authority of the Porte. There was a succession of minor peasant revolts and of armed risings of the aristocracy against the sultan. In 1850 the army of Ömer Paşa finally disciplined the nobles, but in 1858 and 1861–62 there were serious violent outbursts, and in 1875, after a particularly bad harvest the previous year, the whole province burst into revolt.

Hoping for an opportunity for expansion in the area, Serbia had been encouraging dissent, and, on the pretext of defending the Orthodox church, Serbia and Montenegro declared war on Turkey in July 1876. Germany, Russia, and Austria-Hungary dispatched a note to the sultan, demanding reform and the effective defense of the rights of his Christian subjects, and in April 1877, when it became evident that its implementation was not forthcoming, Russia joined the war. Following the defeat of the Turks, the Treaty of San Stefano (March 1878) proposed a radical redrawing of the frontiers of the Balkan states, as well as reforms for Bosnia in particular, and the creation of a large Bulgarian state extending westward to include Ohrid. For a variety of reasons this solution was unacceptable to all of the great powers, and a revision was undertaken in the Treaty of Berlin (July 1878). The new treaty reduced the territory of the Bulgarian state, allowed additional territory to Serbia and Montenegro, placed Bosnia and Hercegovina under Austrian administration, and allowed Austrian garrisons in the Sanjak of Novi Pazar, ensuring the separation of Serbia and Montenegro and keeping alive Austrian hopes for the development of a strategically and economically important railway to Constantinople.

The Treaty of Berlin

The Berlin settlement was vital for the subsequent political development of the region. It sowed the seeds of acute Serb-Bulgarian conflict; the two became rivals for the remainder of Turkey-in-Europe. Bulgaria has ever since been haunted by the vision of its San Stefano frontiers. Croatian aspirations for the incorporation of Bosnia were challenged. Serbian opinion previously had tended to favour Vienna; henceforth the two were to be bitter rivals.

The South Slav lands emerged after the Bosnian revolt as the focus for two of the principal international problems of Europe. For the first time the "Eastern Question" had been posed in its full clarity: what was to be the future of the lands held by the now rapidly disintegrating Ottoman Empire? Also, the encouragement of Serbia in its antipathy to Austria by both Germany and Russia indicated quite plainly that each nation saw its own future as enhanced by the undermining of Austro-Hungarian dominance over the heartland of Europe.

In the last two decades of the century the speed of political and economic change accelerated. The earlier period of prosperity in Dalmatia was ended by the arrival of steam shipping (which its conservative shipowners were slow to adopt) and rail transport. The principal Balkan rail links were constructed in the interior. The commercialization of agriculture had produced some prosperous vineyards, but when grape phylloxera struck in the 1890s the wine trade was almost destroyed. An economic decline set in that was to afflict the littoral until after 1945, and huge numbers of impoverished peasants emigrated.

Following the Ausgleich, Dalmatians became closely involved in Croatian attempts to resist Magyarization. In 1905 the Resolution of Fiume, drafted by the deputies to the Vienna parliament from Dalmatia, Croatia, and Istria, included the demand for union of the area with Croatia. Although these demands were ignored, universal manhood suffrage, introduced in 1907, facilitated the raising of Serbo-Croatian to the status of an official language and ended attempts at Italian cultural domination.

A consequence of the Bosnian revolt in Croatia-Slavonia was the violent upsurge of anti-Serb feeling. Although partly engineered by Count Károly Khuen-Héderváry, the Hungarian ban, this was also made much of by Starčević's Party of Right. Croatian nationalism was particularly incensed by the Railway Regulations Act of 1907, which required all officials to speak Magyar.

The new professional and business middle class took a different view of the situation than had the older patrician

and clerical groups, and in the 1890s a more liberal nationalism began to replace the conservative defense of the old rights. The peasantry also became politically engaged, especially through the Croatian Peasant Party, founded in 1905 by the brothers Ante and Stjepan Radić. Croat–Serb cooperation was also stimulated by the infamous Agram (Zagreb) treason trials of 1908–09, in which a group of Serb deputies were charged, on the basis of forged evidence, with revolutionary activity. A Croat–Serb coalition emerged within the Vienna Diet. During this period the old Illyrian ideal was challenged by another vision. "Trialism" envisaged a reorganization of the Austro-Hungarian monarchy, in which the Slav nations were combined in a third, independent state allied to the Habsburg crown.

Trialism

During the last quarter of the 19th century the Vojvodina was transformed into the "breadbasket of the empire." Its large estates underwent rapid commercialization, stimulated by improved communications. Agricultural wage labour replaced the traditional peasantry, so that socially and economically the region acquired much of its modern character. After the restoration of the Karageorgević dynasty in 1903 the Serbian population of the Vojvodina began to turn increasingly to Serbia for their political future, rather than trying to defend their identity within a Hungarian state. The Serbian Radical Party began to play a large role in local political life.

In Serbia itself political life went through a period of acute disorder following the Bosnian uprising. In 1881 King Milan entered into a secret agreement with Austria by which Serbia gained valuable export conditions for agricultural goods on the understanding that, if Serbia refrained from fomenting further disorder in Bosnia, Austrian support would be forthcoming for Serbian expansion into Macedonia. Encouraged by this, Milan undertook a disastrous expedition against Bulgaria in 1885. Its failure, together with the scandals of his personal life, led to Milan's abdication in 1889. After a confused regency, his son Alexander assumed control of the government in 1893, but the factionalism and corruption of the court did not abate. In the face of massive popular and offical hostility Alexander married his mistress Draga Mašin in 1900. The royal couple were brutally assassinated by officers in the palace in Belgrade in 1903, bringing to an end the Obrenović dynasty. The Skupština invited Peter Karageorgević to return, and a period of reform and economic development was instituted.

In 1906 Austrian objections to tariff negotiations with Bulgaria, which they claimed countervened the treaty of 1881, led to the so-called Pig War, in which Austria imposed massive tariffs on Serbian livestock exports. The Serbs were driven to find new outlets for their goods through Salonika.

In the spring of 1908 it became known that the British and Russians were corresponding about the possibility of setting up an independent Macedonia. In an attempt to forestall the division of the empire, a group of Young Turks, junior military officers, staged a coup d'état, overthrowing Sultan Abdülhamid II and declaring a new constitution. Taking advantage of the situation, the Austrian foreign minister, Aloys, Count Lexa von Aehrenthal, entered a secret agreement with his Russian counterpart, Aleksandr Petrovich, Count Izvolsky, whereby he would back a Russian attempt to gain free naval access through the Dardanelles; the price was support for the Austrian annexation of Bosnia. The annexation produced an intense international outcry that nearly precipitated a general European war, and at home Izvolsky was castigated for his connivance at the loss of Slav lands to the Austrians. The Serbs were enraged and threatened war, but when it became clear that the embarrassed Russians were not willing to support them, they were forced to resign themselves to the annexation. Serb anxieties were heightened in September when Prince Ferdinand seized the opportunity presented by the confusion in Constantinople to declare Bulgarian independence, with himself as tsar. Taken together, these developments reinforced Serbian determination to expand into Macedonia.

Opposition to the Obrenovićs in Serbia was in part economic. The state had become heavily paternalistic toward the peasantry. A combination of population growth and the steady commercialization of agriculture left many peasants in debt. A succession of legislative measures attempted to conserve the traditional household and small plot, preventing the growth of larger farms and the free movement of agricultural labour. The failure to address the problems of agriculture led to the rapid emergence of the Serbian Radical Party and the Agrarian Socialists, both expressing widespread rural dissatisfaction.

Opposition to the Obrenovićs

In Bosnia itself the protectorate and subsequent annexation also had dramatic consequences. Administration of the province fell under the Common Ministry of Finance of the Dual Monarchy, and from 1882 to 1903 routine administration was in the hands of Benjamin Kállay, an energetic and determined, if somewhat insensitive, modernizer. Railway and road construction, linked to the rapid expansion of mineral extraction, advanced. There were improvements in administration, communications, health, and public order. None of this made for social peace, however, for conflict over land reform was closely linked to lines of religious conflict, which were exploited by vigorous Serbian agitation, particularly through the organization Young Bosnia.

The closing decades of the 19th century saw deepening conflict and confusion in Macedonia, as the Turkish capacity to keep order decayed and the ambitions of the great powers and the surrounding states sharpened. Serbia, Bulgaria, and Greece all had expectations of territorial expansion at Turkish expense. The abolition of the Serb patriarchate at Peć placed the Serbs at a relative disadvantage, for both of their rivals used ecclesiastical institutions as tools of propaganda. All parties used education as a weapon, opening numerous schools in the attempt to inculcate an "appropriate" sense of ethnic and linguistic identity.

Terrorism was used where more peaceful means of persuasion failed, and the several secret societies—the Greek National Society, the pro-Bulgarian Supreme Macedo-Adrianopolitan Committee, and several groups of Serb *comitadji* (guerrilla fighters)—were directed as much at one another's adherents as at the Turk. The Internal Macedonian Revolutionary Organization (IMRO) demanded the creation of an autonomous Macedonia. On Aug. 2, 1903 (St. Elijah's Day; Ilinden), an uprising broke out centred on the town of Kruševo, and Macedonian independence was declared. The uprising was brutally crushed, but it aroused enormous international concern. The great powers made several attempts to impose reform (notably the so-called Mürzsteg Program) on the Porte, and they sent their own officers to supervise the local gendarmerie—the first international peacekeeping force.

THE BALKAN WARS

In spite of their conflicting interests, a series of secret bilateral treaties involving Serbia, Montenegro, Greece, and Bulgaria were concluded in 1912, creating a Balkan League, the explicit intention of which was to eject the Turks from Europe.

In 1903 there were outbreaks of anti-Turkish violence in Kosovo and Metohija, and they recurred more seriously in 1909. Although these outbreaks were suppressed as savagely as the Ilinden uprising, the Albanian population rose again in 1911 and captured Skopje. In Macedonia the activities of the guerrilla bands intensified. The Balkan and principal European powers pressed the Young Turk government to implement the promised reforms, but to no effect. On Oct. 8, 1912, Montenegro declared war on Turkey, precipitating the First Balkan War. The Turkish army was defeated by the Balkan League with a rapidity that surprised most observers. By the Treaty of London (May 1913) Turkish possessions in Europe were confined to a small area of eastern Thrace. The situation was unstable, however, for several unresolved issues were left for arbitration by the great powers and Bulgaria was greatly dissatisfied by the partition of Macedonia. The Bulgarians opened hostilities again in June, hoping for rectification of their claims, but they were forced to an armistice by the end of July.

Anti-Turkish uprisings

By the Treaty of Bucharest (August 1913) Montenegro

expanded to a common frontier with Serbia, doubling its population. Serbia was awarded substantial territories to the south, including central and northern Macedonia. On Austrian insistence, however, Serbia and Montenegro were forced to yield part of the territory they had occupied to form a newly independent Albanian state. Because Greece obtained Salonika, Kavála, and coastal Macedonia, the Serbs were denied the direct outlet to the sea for which they had hoped.

The international situation was if anything more dangerous at the end of 1913 than in 1911. Although the Serbs, Greeks, and Montenegrins regarded their gains with some satisfaction, the Bulgarians had not relinquished their claims to Macedonia, and they began to look to Austria for redress, their Russian alliance having failed them. The Austrians saw in the emergence of a strong Serbia an end to their own *Drang nach Osten* ("drive to the east"), while Serbian animosity against Austria was intensified. This last factor was especially important, and during a visit to Sarajevo on June 28 (Vidovdan; Serbia's national day), 1914, the Austrian archduke Francis Ferdinand and his wife were assassinated by Gavrilo Princip, an adherent of Young Bosnia. Seeing in the event official Serbian complicity if not direct guilt, the Austrians issued a precipitate and ill-considered ultimatum that went far beyond insisting on effective action against the assassins, including demands for the suppression of anti-Austrian newspapers and the dismissal of anti-Austrian teachers and military officers. The Serbian reply, though conciliatory, was considered unsatisfactory, and in July the two countries went to war, bringing with them a general European catastrophe.

WORLD WAR I

Unification was not part of the war aims of any of the principal units that were to make up the first Yugoslav state at the end of World War I. The Serbs were concerned principally for the defense of their territorial gains of 1912–13, and if they thought of expansion it was in terms of a "Greater Serbia" that might encompass parts of Bosnia. The national ideal of the South Slav peoples of the Habsburg empire was "Trialism": the creation of an autonomous Slav state within the ambit of the Dual Monarchy.

The Austrian offensive against Serbia of Aug. 14, 1914, was forced back within two weeks; after desperate fighting a second attack in November was also repelled. Although heavily outnumbered, the Serbs hoped to hold out until the intervention of their allies. In the winter of 1914–15 a terrible outbreak of typhus struck the country, devastating both the civilian population and the military. By the autumn of 1915 Austrian victories in Galicia enabled them to turn their attention again to the Balkans. An Allied plan to reinforce the Serbs by landing a small force in Salonika developed too slowly, and when the German field marshal August von Mackensen opened a third offensive in October, assisted by the Bulgarians, the Serbian position deteriorated rapidly. The Serbs, deprived of reinforcements and supplies and weakened by disease, were forced to retreat through Albania. Throughout the winter of 1915–16 the remains of their army endured appalling conditions crossing the mountains to the Adriatic coast, whence they were shipped to safety by the French, mostly to Corfu. No assistance was forthcoming from the Montenegrins. When the Austrians attacked in January the Montenegrin king, Nicholas I, fled to Italy, leaving his son to arrange a surrender.

The Yugoslav Committee

During the early period of the war a number of prominent political figures left the Dual Monarchy, and in London they set up a Yugoslav Committee with the aim of conducting propaganda on behalf of their compatriots. These people included Ante Trumbić (a former mayor of Split), Franjo Supilo (a journalist), and Ivan Meštrović (a sculptor). They were assisted greatly by several well-placed sympathizers, especially Wickham Steed, the editor of *The Times,* and the historian R.W. Seton-Watson. One of the committee's most important achievements was the discovery by Supilo of the Treaty of London, a secret document drawn up in April 1915 by which the Italians were promised Istria and large areas of Slovenia and Dalmatia (apparently with the connivance of the Serbian government-in-exile) in return for their participation on the Allied side.

The stagnation of the war during 1916 and early 1917 and the general indifference of the major Allied powers to the fate of the national minorities within the Austro-Hungarian Empire slowly compelled the Yugoslav Committee and the Serbian government-in-exile to come together in common defense. In July 1917 representatives of the two groups met in Corfu and signed the Corfu Declaration, in which a number of common aims were set out for the first time. The declaration called for a single state governed by a democratic and constitutional monarchy, in which there would be equality for the two alphabets, three national names and flags, and religious toleration. The details were left to a future constituent assembly, and in particular no mention was made as to whether its structure was to be federal or unitary.

In spite of the 1915 Treaty of London, the Yugoslav Committee was able to reach some agreement with the Italians, resulting in a Congress of the Oppressed Nationalities of Austria-Hungary in April 1918, from which was issued the Pact of Rome, affirming the right of subject peoples to national unity and self-determination. The growing unity of voice among the national minorities greatly strengthened the interest of the U.S. president Woodrow Wilson in these matters, contributing to the formulation of his Fourteen Points program on national self-determination.

The stagnation on the Salonika front was in part due to the neutrality of the Greek government. The rise to power of the Greek prime minister Eleuthérios Venizélos in November 1916 changed this situation, and he brought the Greeks into the war on the Allied side. It became possible to open a new front against the Bulgarian-German forces in Macedonia, with the Serbian army playing a key part alongside British, French, and Greek units. After two weeks of hard fighting the Bulgarians surrendered. The collapse of the Macedonian front was one of the most important factors precipitating the end for the Central Powers. Following the recapture of Belgrade on Nov. 1, 1918, the Austro-Hungarian forces agreed to an armistice.

Even before the military defeat was secure, the internal divisions within the empire were becoming clear. The Slovene leader Anton Korošec had organized a Yugoslav National Council (Narodno Vijeće) in 1917, openly advocating South Slav unity. There had been a serious mutiny in the Austrian fleet, which was heavily dependent upon Dalmatian seamen, at Kotor in February 1918, and armed bands of deserters harassed the authorities throughout Croatia. Croatian and Slovene deputies to the diets in Vienna and Budapest had been preparing the ground through the National Council, and on Oct. 29, 1918, the Sabor in Zagreb declared the union with Hungary to be severed.

The South Slav state

From this date there was a state that united within itself Serbs, Croats, and Slovenes, but the state was not yet Yugoslavia. Serbia and Montenegro had made no commitment to it. Indeed, in spite of the Corfu Declaration the Serbian leader Nikola Pašić regarded the new state with some dismay. Pragmatic cooperation with the Yugoslav Committee was one thing; a de facto government was another matter. In November continuing Italian action made it apparent that the Italians were not content with the territories allocated to them by the 1915 Treaty of London. This placed considerable pressure on the "Yugoslavs" to seek the effective support of the advancing Serbian army. All sides were constrained by the major Allied powers to reach an accommodation, and a conference held in Geneva on November 6–9 concluded with a declaration of union by representatives of the Yugoslav Committee, the National Council, and the Serb political parties. In September the Montenegrins rose against Austrian occupation, and on November 26 a national assembly in Podgorica declared for union with Serbia under the Karageorgević dynasty, formally deposing the absent king Nicholas.

No treaties had been concluded determining the boundaries of the new state, and there was no agreement among the parties as to the constitutional form that it was to adopt; the state had no army, no judiciary, and no cur-

rency. Nevertheless, on Dec. 1, 1918, a delegation from the National Council invited the prince regent Alexander to proclaim the new union. On December 4 the Kingdom of Serbs, Croats, and Slovenes was announced to the world. Alexander's father, Peter Karageorgević (reigned 1918–21), was proclaimed king.

It is commonly stated that Yugoslavia was a "creation of the Treaty of Versailles." External factors were important in the formation of the Yugoslav state, but if this phrase is intended as more than a concise statement of the idea that a united South Slav state was a consequence of World War I, it is seriously misleading.

Memorials in Yugoslavia frequently commemorate the dead of the wars of 1912–19. Hostilities placed a great burden of suffering on the peoples of the area and devastated large parts of the country. Reconstruction would have been difficult in such an economically backward country even under favourable circumstances. The matter was complicated by the need to construct almost from scratch both a unified administrative apparatus and a constitutional government to guide the process.

Border disputes

Of the kingdom's seven international borders, only one (with Greece) was not contested at the peace conferences. The settlement with Italy was complicated by the occupation of Fiume by the nationalist forces of Gabriele D'Annunzio and by Italian endeavours to go beyond the boundaries envisaged by the Treaty of London. The issue was not settled until the Treaty of Rapallo in November 1920, which awarded to Italy the city of Zara (Zadar), several Dalmatian islands, Istria, Gorizia, and a sizable part of the Julian Alps. The border with Austria was resolved only after a plebiscite in the Klagenfurt basin (October 1920). Peace could not be concluded with Hungary until the withdrawal of Yugoslav troops from the city of Pécs in August 1921, and rectifications of the borders had to be negotiated with both Romania and Bulgaria. Most protracted of all was the dispute with the newly independent state of Albania, which was only concluded in July 1926.

THE KINGDOM OF SERBS, CROATS, AND SLOVENES

The new state faced at its birth problems arising from its national composition, economic structure, and international location. The boundaries separating the several states constructed on the ruins of the Ottoman and Habsburg empires inevitably failed to reflect the complexity of centuries-old settlement patterns. The "South Slav state" was by no means populated only by South Slavs. Of the nearly 12 million inhabitants recorded in the census of 1921, more than 12 percent were not speakers of Slavonic languages, including more than half a million German speakers and more than 400,000 each of Hungarian and Albanian speakers. The Christian population was mainly divided between adherents of the Eastern Orthodox and Roman Catholic churches, but more than a tenth of the total population were Muslims. In the absence of any common traditions or political institutions, the task of welding such disparate raw material into a nation was difficult.

Although the new state received substantial sums in war reparations from the Central Powers, the problems of physical and organizational reconstruction were immense. These were particularly pressing problems in Macedonia, where two decades of guerrilla strife had been followed by the Balkan Wars as well as the 1914–18 struggle. Parts of the kingdom had already begun to industrialize and to commercialize, but the majority of the population were still living in primitive and isolated communities dependent upon subsistence agriculture, the techniques of which had hardly changed for a thousand years. Even where modern means of communication had been inherited these had been constructed with entirely different needs and interests in mind. The rudimentary Serbian rail system pointed toward the port of Salonika, whereas that of the northern regions was integrated with the Austrian and Hungarian systems. Communications in Bosnia had been built as much with the strategic requirements of Austria as the economic needs of the locality in view. No modern rail or road link connected Belgrade and Zagreb.

At the international level, the shattering of the two former empires had led to a realignment of interests and powers in Europe. The rise of Germany and the Soviet Union during the interwar period was to impose an entirely new balance of considerations on the development of diplomatic relations.

The failure of nation building: 1918–41. Elections in November 1920 produced a constituent assembly whose acrimonious deliberations foretold its own failure. The 419 members of the assembly represented no fewer than 15 parties, most with specifically ethnic constituencies. The fundamental divergence of opinion between the parties concerned the choice between a unitary or a federal state. The defeat in principle of the federal idea led to the withdrawal of the Croatian Peasant Party—with 50 seats the largest of the federalist groups—from the assembly. Following the assassination of a minister by a young Communist in 1921, the Communist Party (with 58 seats) was declared illegal. An alliance of the principal Serb parties (Radical and Democrat), together with the Muslim Organization, was able to press through a highly centralized constitution, modeled on that of prewar Serbia. It was promulgated on the Serb national day (Vidovdan), June 28, 1921, indicating a lack of awareness on the part of Serb politicians of the depth of difference between the political traditions of the South Slav peoples. Serb experience had revolved around the creation of a strong state, that of the Croats and Slovenes around the struggle to defend the nation against too strong a state.

The constituent assembly

The belief of the Croats and Slovenes in the lack of Serb good faith was validated by the intransigence with which successive governments imposed centralist solutions on the country's problems. Serb mistrust of the Croats was confirmed by the flirtation between Stjepan Radić, the leader of the Croatian Peasant Party, and the Moscow-sponsored Green International and by the bewildering series of political about-faces that Radić was prepared to accomplish. These ranged from his rigid abstentionism in the constituent assembly in 1920–21 to his willingness to enter a coalition government with Pašić in 1925–26. With few exceptions, therefore, the decade 1919–29 was characterized by growing bitterness on the part of non-Serb groups, accentuated by the discovery of deepening corruption on the part of politicians who were beyond any effective democratic control.

When in June 1928 a Montenegrin deputy shot two of the Croatian deputies dead in the Skupština and mortally wounded Radić, the days of the Vidovdan constitution were numbered. The Croatian deputies withdrew to Zagreb and set up an alternative assembly. For some months the Slovene leader Korošec tried to hold things together, but without success. King Alexander opened discussions with Radić's successor, Vladimir Maček, to see if a solution to the constitutional crisis could be found. It became evident that the Serbs were unwilling to contemplate a federal state at any price, while the Croats were unprepared to consider anything else. Frustrated by the inability of the politicians to make progress, on Jan. 6, 1929, the king dissolved the Skupština and declared a personal dictatorship.

The king set out to break the old ties of region and ethnicity and to develop in their place a new sense of common national identity. The name of the state was changed to Yugoslavia, and the former regional names were abolished in a reorganization of local government into nine banovine (governorships), together with the prefecture of Belgrade. Their boundaries were drawn with the intention of weakening traditional regional loyalties. Parties that appealed to specific regional or religious constituencies were banned. The royal couple made a courageous and successful visit to Zagreb. In spite of the popular appeal of some of Alexander's measures, others only exacerbated hostility to the regime, including the suppression of the patriotic gymnastic societies, interference with the judiciary, the suppression of the free press, and the arrest and even torture of many critics of royal centralism. The imprisonment of several leading Communists at this time created an esprit de corps that was to stand them in good stead in later years. Many Croatian extremists fled to Italy and Hungary, where they set up the terrorist Ustaša organizations.

Yugoslavia

A new constitution was promulgated in 1931. It nomi-

nally returned the country to representative government, but its provisions were so heavily centralist that it failed to secure the support of the Croats and of many liberal groups, along with some Serbians and the increasingly articulate Macedonians. Under its provisions, only parties that had registered supporters in every electoral district could compete in elections, making political life extremely difficult for virtually every opposition group. Real democracy was further hampered by the gerrymandering of constituency boundaries, open voting, terrorism, and electoral corruption. Although the official Yugoslav National Party (later the Yugoslav National Union) was challenged in the elections of 1935 and 1938 by a United Opposition bloc led by Maček, the diversity of aims of opposition groups and the disabilities under which they were organized undermined their effectiveness.

A decisive change was brought about by the assassination of the king by an Ustaša member during a state visit to France in 1934. A regency was established, headed by Prince Paul, the uncle of the heir to the throne, Peter II. In the elections of 1935, the government list secured 303 seats with 1,746,982 votes, and the opposition won 67 seats with 1,076,345 votes. As a result of accusations of gross corruption and the resignation of several ministers, the regent invited Maček for consultations. The minister of finance, Milan Stojadinović, was asked to form a government, which was generally believed to signal acceptance of the need for accommodation and conciliation. The new government began propitiously with the release of many political prisoners.

Stojadinović

Some commentators have portrayed Stojadinović as cautious, seeing the failure of his attempt to establish a concordat with the Vatican in 1937 as a sign of weakness and lack of direction. Others have focused on his willingness to enter dialogue with Hitler and Mussolini, and the emergence of apparently fascistic tendencies within his own party, as indications of his own hostility to democratic politics. Under his administration there was a continuing disintegration of Yugoslav political life, marked by a reduction in the effectiveness of central government and a growth of political violence.

In spite of his electoral victory in December 1938, Stojadinović was compelled to step down when five of his ministers resigned in protest against a provocatively nationalist speech by one of their colleagues. Stojadinović was succeeded by Dragiša Cvetković, who had been negotiating secretly on the regent's behalf with Maček. These discussions resulted in August 1939 in an agreement between the Croatian leader and the government that granted a new and semi-independent status to Croatia under its own ban and Sabor. A coalition government was created, and the electoral law was revised in preparation for new elections. Despite the continuation of terrorist outrages in Zagreb, there was a revival of hope that a solution to Yugoslavia's constitutional problems might be found. This hope remained untested at the onset of war in 1941.

Economic change between the wars. Notwithstanding its tempestuous internal politics, the period immediately following World War I was a prosperous one for the Yugoslav kingdom. The growing demand for food both at home and abroad gave a strong stimulus to agriculture. The merchant marine benefited enormously from the expansion of world trade. Adriatic tourism boomed as resorts such as Brioni and Dubrovnik enjoyed a fashionable reputation. Inheriting from Serbian experience a disposition for active intervention, the new government used substantial loans (especially from France and the United States) to stabilize the dinar and improve transport. The Zagreb–Belgrade rail link was completed in 1925. Aided by war reparations from the Central Powers, industry was built up behind protective tariff walls.

One of the earliest measures announced in 1918, even before the constituent assembly had met, was a program of land reform that abolished serfdom, announced the expropriation of large estates, and promoted the colonization of abandoned and undeveloped land. The haste with which this measure was promulgated reflected the fear of bolshevism in the wake of the Russian Revolution. Its implementation in specific areas was left to separate legislative acts introduced over the next two decades. These acts became bargaining counters in a succession of deals to ensure government advantages, so that the program's effects were less radical than its intentions. Colonization was typically used as a means of manipulating local demography to suit the interests of Belgrade. The redistribution of land was not coupled effectively either with investment or with the rationalization of holdings. Nevertheless, the reform ensured that Yugoslavia would remain a country of peasant farmers even after World War II.

For the most part there was no sustained acceptance of the need for the government to promote agricultural development, and the peasantry were left to help themselves through the various cooperative movements, especially in Slovenia and Croatia. During the economic crisis of the 1930s, however, measures were taken to control the supply of rural credit, a moratorium was imposed on peasant debt, and central purchasing agencies were established for some crops and forest products.

Industrialization

In contrast, industrialization was a consistently enunciated policy of all postwar governments. Extractive industries, forestry, power generation, and metallurgical concerns were built up with foreign capital. Some manufacturing (notably textiles) developed with the aid of tariff protection, and machinery was acquired under the war reparations scheme.

On the whole the first decade of peace was one of prosperity, and even the Western economic crisis of 1929 left Yugoslavia relatively untouched. It was not until 1931 that real economic difficulties set in, as the cushion of war reparation was removed, the German banking collapse reverberated in the Balkans, French economic support was withdrawn, and Britain departed from the gold standard. The "price scissors"—the widening gap between agricultural and industrial prices—began to affect the agrarian population drastically. Yugoslavia was hit especially hard by the economic measures taken by the League of Nations against Italy in 1934–35 following the annexation of Abyssinia. Italy had been one of Yugoslavia's principal trade partners.

A wholesale reorganization of the country's trading relations was brought about by the recovery of Germany under the Nazis. Hermann Göring visited Belgrade in 1934, and a commercial treaty was signed. Yugoslavia gradually was drawn into a more binding relationship with the German Reich, as favourable terms were extended to exports, and Yugoslav companies were incorporated into German cartels in line with plans for economic expansion. By 1938 trade with Germany accounted for 53 percent of exports and 65 percent of imports.

Yugoslavia's international relations. Threatened on several sides by irredentist regimes still entertaining active aspirations for its territories, Yugoslavia clung to the habit of seeking great-power sponsorship. In this respect the 1920s were characterized by a strongly pro-French orientation. In 1920–21 the Little Entente was created, involving Yugoslavia with Czechoslovakia and Romania. In the attempt to build up the stability of the region and to counter Soviet influence, the French in turn entered into alliances with these countries.

When a left-of-centre government was returned in the French elections of 1932, the French premier Édouard Herriot reversed this antisocialist stance, and with it the desire to bolster explicitly antisocialist regimes such as that of Yugoslavia. There was no possibility of turning to Italy for support, and relations with Hungary remained strained; both countries were involved in the training of terrorist groups active on Yugoslav soil. From 1933 onward the king took the initiative in building closer ties with Yugoslavia's Balkan neighbours—Bulgaria, Greece, and Turkey, together with the existing commitment to Romania—a policy that bore fruit in the Balkan Entente.

In the late 1930s it became clear that this modest measure of collective security was no match for the real threat to the independence of the state, German expansion. Following the 1938 Anschluss the Yugoslavs found themselves drawn steadily into closer economic ties with those states that they regarded as their chief diplomatic rivals. Successive governments worked hard to maintain a po-

sition of independence. The replacement of Stojadinović by Cvetković in February 1939 was motivated as much by the regent's concern to stiffen resistance to German diplomatic demands as by his desire to promote a solution to the Croatian problem. As war drew closer German pressure to associate Yugoslavia with the Tripartite Pact (Germany, Italy, and Japan) grew.

The possibilities for the maintenance of independence were running out, however. The fall of Czechoslovakia cut Yugoslavia off from its principal armaments supplier, the Molotov–Ribbentrop Pact (August 1939) made it clear that the Soviet Union was unwilling to play its traditional role of "Protector of the Slavs" vis-à-vis Nazi Germany, and Romania was partitioned in 1940. Meanwhile, Albania was invaded by Benito Mussolini's forces in April 1939 and by Greece the following year. In March 1941 Prince Paul and his ministers finally agreed to accept Adolf Hitler's demands and signed the Tripartite Pact.

The Tripartite Pact

The journey of Cvetković and Alexander Cincar-Marković, the foreign minister, to Vienna had been kept secret, so that the capitulation only became public following their return on March 26. The response was one of public outrage at what was regarded as a national humiliation. In a bloodless coup d'état led by several air force officers, the regents and their senior ministers were arrested and sent into exile, King Peter's majority was proclaimed prematurely, and amid massive and emotional demonstrations of popular support a government of national unity was formed.

The active compliance of the Balkan states had been important for the Germans, securing their southern flank in preparation for an attack against the Soviet Union. Yugoslav bravado spoiled this, and Hitler determined to finish both Yugoslav and Greek resistance. On April 6, 1941, German troops invaded Yugoslavia. As British prime minister Winston Churchill described it to the House of Commons: "Early this morning the Yugoslav nation found its soul." It was, however, already too late to save its territory.

WORLD WAR II

The partition of Yugoslavia. Within two weeks Yugoslav resistance was crushed, and King Peter and his ministers flew to Athens, later setting up a government-in-exile in London. Most of the Slovene lands were directly annexed to the Third Reich. The southern part was annexed to Italy, which also appropriated large areas of Dalmatia. The Italians created client states in Montenegro and Albania, to which were joined the mainly Albanophone areas of Kosovo, Metohija, southern Montenegro, and western Macedonia. Hungary incorporated Prekmurje and Međumurje and most of the Vojvodina. The Banat was administered directly by Germany, and a puppet regime was installed in a greatly diminished Serbia under the former minister of war Milan Nedić. Bulgaria took over the rest of Serbia and Macedonia. An enlarged Croatia, which included Bosnia and Hercegovina, was turned into the Independent State of Croatia (Nezavisna Država Hrvatska; NDH), headed by the leader of the Ustaša, Ante Pavelić. The country was divided for military purposes between German and Italian zones of control.

In Croatia the indigenous fascist regime set about a policy of "racial purification" that went beyond even Nazi practices. Minority groups such as Jews and Gypsies were to be eliminated, as were the Serbs: it was declared that one-third of the Serbian population would be deported, one-third converted to Roman Catholicism, and one-third liquidated. The Muslims were officially adopted as Croats, and an attempt was made to turn them against the Serbs. Ustaša bands terrorized the countryside. The partial collaboration of the Catholic clergy in these practices compromised church–state relations seriously after the war, and it continues to be a component of Serb–Croat suspicion.

Rival resistance movements. Although the Yugoslav Royal Army disintegrated rapidly in the face of the Axis attack, groups of its personnel did not surrender but went into hiding with their weapons. They adopted the name Chetnik (Četnik), which means a member of a band or company. (The term had its origins in the groups of armed irregulars who harassed the Turks during the 19th century.) The soldiers remained an organization of voluntary reservists until 1940, when they were reconstituted as units trained for guerrilla fighting. In May 1941 an organization of these bands emerged under the leadership of Colonel Dragoljub (Draža) Mihailović, an experienced officer who had fought in the Balkan Wars and World War I and who established himself in the Ravna Gora district, near Valjevo.

The Chetniks

In July a second armed resistance movement was created by the Communist Party; it came to be known as the Partisans (Partizani). Communism only seems to have found a firm footing in Yugoslavia around the time of World War I, its principal points of growth being in the new industrial areas around the major cities and among displaced agricultural workers in Macedonia. Very much on the fringes of the international movement, the Yugoslav Communist Party had a reputation for factionalism. It was repeatedly infiltrated by government agents and throughout the interwar period was under constant attack, many of its active members spending terms in prison.

In 1934 a young metalworker from Zagreb, Josip Broz, was elected to the Central Committee. He was sent to Moscow to work in the Balkan secretariat of the Comintern. In 1937 Broz was appointed secretary general of the Yugoslav Party, and he returned to Yugoslavia to work underground, reorganizing the battered and demoralized movement. This he did with great success, more than doubling its strength in the years before the outbreak of war and extending its support among students and intellectuals.

After the fall of Yugoslavia in April 1941 the Communist Party cautioned its supporters in the armed services to conceal their weapons and hold themselves in readiness for action. The position was altered radically by the German invasion of the Soviet Union in June. On July 4 an armed uprising was proclaimed (still celebrated in Yugoslavia as a public holiday, Dan Borca ["Fighters' Day"]). Broz, now known by the code name Tito, emerged as the head of a force of Communist insurgents. In September 1941 they occupied the Serbian town of Užice, with its important armaments factory, and took control of part of western Serbia and eastern Bosnia, an area that became known as the Užice Republic. Another rising took place in Montenegro.

In place were the essential elements that were to characterize the complex, bitter, and bloody struggle of the next four years. Yugoslavia was occupied by a variety of alien forces that were supported by several indigenous fascist groups and opposed by armed Yugoslav insurgents fighting for the liberation of their country. The liberation movement, however, was divided between two camps, each of which stood for a very different conception of the future. The Chetnik organization of Mihailović was affiliated to the royal government-in-exile, and it was almost exclusively composed of Serbs whose vision of the future of Yugoslavia was of a strongly unified country in which a "Greater Serbia" would play the leading role. Although the Partisans played down their specifically Communist commitment in the early days of the war, they were firmly led by the Communist Party, which soon showed that it intended to create a federal and socialist republic. Four distinct types of struggle were thus overlaid in Yugoslavia between 1941 and 1945, while it featured as one theatre of the wider international conflict. Within that context, a war of national liberation was being conducted by various groups committed from several standpoints to the independence of their country. For some groups (especially the Croatian nationalists), however, their national independence was seen as guaranteed by the support of the Axis powers, and their fight against both types of insurgent groups made the war also a civil war. Finally, the Communists aimed beyond national liberation to the overthrow of the regime and the creation of a socialist system.

Internal conflicts

The military and political struggle. The depth of disagreement between the rival groups became clear in October and November 1941 when the Germans and their allies in Serbia mounted an offensive against the Užice

Republic. Talks took place between the two liberation movements in an attempt to coordinate their action, but all contact was broken off at the end of November. Mihailović saw his role as sustaining a movement that by itself would never be strong enough to confront the Germans but that could be activated in the event of an Allied landing in the Balkans. Meanwhile, he believed, it was important not to provoke German reprisals against the civilian population. This low-key approach was directly opposed by the Communists, who regarded such passivity as more demoralizing than any reprisals that their own armed action might precipitate. Henceforth the two groups were to regard each other with as much hostility as they did the occupying forces.

Forced to abandon Užice, the Partisans retreated to the Bosnian town of Foča, which they made their headquarters in December 1941. There they established the Proletarian Brigades, corps of picked troops noted for their fighting qualities, their ideological commitment, and their all-Yugoslav character. This measure angered the Soviet government, which had insisted that the national rather than the revolutionary character of the struggle should be emphasized, and it caused the Soviets some embarrassment since they had recognized a royalist Yugoslav legation in Moscow. At the end of March 1942 a second offensive was mounted by Italian, German, Ustaša, and Chetnik units, which compelled the Partisan forces to move to Bihać in the northwest.

In November 1942 Tito staged a political coup by organizing the Anti-Fascist Council for the National Liberation of Yugoslavia (AVNOJ). Its propaganda value lay not only in the fact that his movement was strong enough to organize such an occasion but also in the specific recognition given by its resolutions to the equality of the different Yugoslav peoples.

During the winter of 1942–43 the Germans suffered several reverses in the Russian campaign, and the Allied advance began in North Africa. It appeared important to the German war effort to ensure that, if an Allied landing were to take place in the Balkans, the Allies would not be aided by local resistance forces. In January, therefore, Operation Weiss was launched against both Partisan and Chetnik forces. Tito directed his major effort against the Chetnik formations, inflicted heavy losses on them, and broke out of encirclement into the Durmitor area of Montenegro. An expected Italian move against him did not materialize, and the Communists were given an important breathing space.

Operation Weiss

In May a further German offensive, Operation Schwartz, proved to be the turning point of the war. In the face of hugely superior enemy concentrations and in spite of severe losses, the Partisans escaped encirclement by forcing an exit up the Sutjeska gorge and back into central Bosnia. German plans to eliminate Communist resistance were thwarted. In September the Italians surrendered to the Allies, and the Liberation Army was able to acquire huge quantities of war matériel and to secure control over large areas of the coast. Possibly most important of all, the battle of Sutjeska was witnessed by a British mission that had been sent to establish liaison with Tito.

Tito, Mihailović, and the Allies. Allied contact with the various resistance groups dated from the autumn of 1941, when Captain D.T. (Bill) Hudson led a mission to establish contact with Mihailović. Intelligence about the situation in Yugoslavia remained extremely poor, so that even by the spring of 1943 the leadership of the Communist movement remained unknown. The Special Operations Executive (SOE) did not establish contact with Tito until the missions of April–May, and their arrival coincided with the start of Operation Schwartz, which gave them an impressive introduction to Partisan military capability. In September Brigadier Fitzroy MacLean arrived with a commission from Churchill to determine whether Tito or Mihailović was providing more effective military resistance to the Germans. The account that MacLean was given of the battle of Sutjeska was of first importance in persuading Churchill that Allied support should be switched from the royalists to the Communists. A divided and disorganized government-in-exile in London was never able to compensate for the failure of their forces to win the military confidence of the Allies, and although the government of King Peter retained the official recognition of the Allies to the end of the war, it was increasingly regarded as an irrelevant embarrassment.

A second session of AVNOJ was held in November 1943 in the Bosnian town of Jajce. The session declared itself to be a "temporary government" in competition with the exiles in London. Although these independent moves continued to anger Joseph Stalin, at the Teheran Conference (November–December 1943) Churchill, Stalin, and the U.S. president Franklin D. Roosevelt agreed to support Tito as the only effective resistance in Yugoslavia. A Soviet mission was sent, though with little aid, and the British mission to Mihailović was withdrawn. Mihailović's attempt in January 1944 to mount a countercoup to AVNOJ, in his St. Sava Congress at Ba, attracted little recognition.

Partisan victory. Although by the beginning of 1944 it was becoming increasingly evident that the Germans were losing the war, in a final offensive in May the Germans almost captured Tito in a parachute raid on the Bosnian town of Drvar. By this time, however, Allied air and sea support was giving growing strength to the Communist armies, which began to push the invaders back into Serbia. Anticipating a Soviet advance through Romania and Bulgaria, Tito flew to Moscow in September to discuss collaboration with the Red Army.

Under British pressure, King Peter withdrew support from Mihailović in June, and in the summer of 1944 negotiations were opened involving Tito and the former Croatian Peasant Party leader Ivan Šubašić, the prime minister of the royal government-in-exile, over the nature of a future regime in Yugoslavia. Although it was agreed that a Communist government would not be imposed in Yugoslavia, the strength of Tito's position was shown by the stipulation that the king would not return to the country before he was specifically invited. The National Liberation Committees that were being established in the areas controlled by the Communist forces were recognized as the effective provisional administration.

National Liberation Committees

Partisan military forces increasingly turned their attention to dealing with the Chetniks during 1944, and the latter were forced into growing accommodation with the Nedić regime. As the struggle degenerated into civil war, the Germans were able to make a relatively organized withdrawal. By September the Soviets were at the Yugoslav borders. Mihailović made a belated attempt to secure support by proclaiming a general uprising in support of the advancing Allies, but his offer met with little interest from the Soviets. Belgrade fell to a combined operation of Yugoslav Communist and Soviet troops on Oct. 20, 1944. Shortly afterward Šubašić arrived there to participate in the setting up of a provisional government.

Mihailović had been strongly supported in Serbia, which was taken by the Partisans only with much difficulty. Compelled to retreat into Bosnia, he still commanded a force of some 50,000 men, around which he attempted to rally an anti-Communist alliance. Cooperation with the Ustaša was out of the question, and the other groups were too fragmented, too disparate in their aims, and in several cases already too involved with the Germans to permit the creation of an effective counterforce to the Communists. Furthermore, the establishment of the Tito–Šubašić government undermined his efforts. Bitter and confused internecine fighting continued alongside the pursuit of the retreating Germans throughout the early months of 1945. Berlin fell to the Allies on May 2, 1945, and the German forces in Yugoslavia surrendered on May 15. The anti-Communist forces split and many units retreated into Austria with the Germans, but Mihailović was unwilling to give up the struggle and attempted to fight his way back into his home territory of Serbia across the mountains of Bosnia. Exhausted, outnumbered, and short of equipment and ammunition, his force was beaten and dispersed in central Bosnia. Mihailović himself evaded capture until March 1946. He was tried for high treason and executed in July. Meanwhile the victorious Partisans were challenging the Allies on their Italian and Austrian frontiers in an

attempt to secure as advantageous as possible a position in the negotiations over the creation of a new state.

THE SECOND YUGOSLAVIA

The problem of Yugoslavia's frontiers. As at the end of World War I, one of the first problems facing the new Yugoslavia was the extent and security of its frontiers. Having occupied the city of Trieste on May 1, 1945, the Partisans hoped that its possession was secured, but the Allies insisted that the status of the city be resolved at the Paris Peace Conference. The difficulty of reconciling competing claims was temporarily resolved by the establishment of the Free Territory of Trieste (FTT) in 1946 under the aegis of the United Nations. The UN boundary commission, anticipating the resolution of the problem, divided the FTT into two zones, A and B, and this arrangement was confirmed in 1952. The Yugoslavs protested fiercely at every stage, even to the point of threatening military action. In October 1954 a memorandum of understanding signed by Great Britain, the United States, Italy, and Yugoslavia ended the military occupation of the FTT. The situation of the small Slovene minority in Italy is still occasionally the focus of bad feeling between the two states. Yugoslav claims in Austria were even less successful, as these were treated as entirely peripheral to the wider problem of the future of the country, which remained under joint occupation.

Trieste

Yugoslav and Albanian Communist insurgents cooperated against the Italians during the war, although in Kosovo the local population repeatedly showed its hostility to both Yugoslav resistance movements. In December 1944 an uprising took place in Kosovo in protest of the agreement between Tito and the Albanian leader Enver Hoxha that the area would revert to Yugoslavia after the war. A treaty of friendship was signed between the two countries in July 1946, and the Yugoslavs contemplated the incorporation of Albania at this time, but collaboration was ended by the dispute with the Cominform (see below).

Among the more successful aspects of the new Yugoslav federal structure was the creation of the Macedonian republic, although it brought with it problems in relations with the bordering states of Bulgaria and Greece. In 1944 the Yugoslav and Bulgarian Communist parties had discussed union in a Balkan federation after the war. Their negotiations ran into trouble over the status of Macedonia and were shelved in the face of the disapproval of the Allied powers and the uncertainties of the Greek civil war. Defeated in the referendum of 1946, the Greek Communists resorted to armed rebellion, aided by the Yugoslavs and Bulgarians. The Macedonians took advantage of the situation to promote vigorously the cause of unification, to the annoyance of their Greek comrades. The conclusion of hostilities left no particular legacy of territorial contention between Yugoslavia and its Romanian and Hungarian neighbours. The peace conference awarded to Yugoslavia the former Italian possessions in Istria and Dalmatia.

The framework of the new state. The Tito–Šubašić government of 1945 was intended as only a transitional administration. Three representatives of the government-in-exile joined a largely Communist administration based on Tito's National Liberation Committee. Elections to the constituent assembly in November returned a massive majority for the Communist-led People's Front, and Tito began the gradual process of easing out of office the remaining leading non-Communist politicians.

The new constitution establishing the Federal People's Republic of Yugoslavia (Federativna Narodna Republika Jugoslavije; FNRJ) was promulgated on Jan. 31, 1946. It was modeled on the constitution of the Soviet Union. The monarchy was replaced by a federation of six republics and two autonomous provinces. There was a bicameral Federal Assembly, and day-to-day affairs were conducted by a Presidium. In spite of the federal form of the constitution, power was held firmly at the centre and exercised by the Communist Party through a close interlocking of state and party functions. The major productive forces and means of communication and exchange were nationalized, and a rigid central planning apparatus was put in place. The new regime was a classic example of the Stalinist "dictatorship of the proletariat," and it was referred to by contemporary Yugoslavs as "administrative socialism."

The 1946 constitution

Although the "loyal opposition" was quickly but relatively gently eased from power, "disloyalty" of any kind was ruthlessly suppressed. Those suspected of collaboration with the former enemy were punished or killed and their property confiscated. The trials of several prominent collaborators (especially Mihailović and the Croatian archbishop Stepinac) were given wide international publicity.

A law on agrarian reform introduced in August 1945 expropriated large estates, together with the lands of collaborators and those who had fled. Although the reform set up a number of large state farms, most of the land was distributed to small peasant landholders, and in particular to those who had served with the Partisan armies.

The break with the Cominform. The Communist International (Comintern) was disbanded in 1943 and replaced in 1947 by the Communist Information Bureau (Cominform), with headquarters in Belgrade. From the outset relations between the Yugoslav Communist Party and the Soviet Union had been uneasy, dating to Tito's independence in the conduct of the "national liberation struggle." Three substantial areas of disagreement were the structure and operation of the joint trading companies that were set up between Yugoslavia and the Soviets, Yugoslav irritation at the failure of the Soviet Union to reciprocate their own supportive role in foreign policy, and Soviet criticism of the organization of the Yugoslav army and the restrictions placed on the role of Soviet advisers.

Indications of the severity of these disagreements emerged in February 1948, when further discussion of the planned Balkan federation was quashed. An acrimonious correspondence developed between the two parties, in which the Yugoslavs were accused of ideological, economic, and political indiscipline, and they in turn protested the misconduct of Soviet advisers (including accusations of espionage). To the astonishment of external observers the Cominform expelled the Yugoslavs on June 28, 1948, and the organization's headquarters were transferred to Bucharest. A political, diplomatic, and economic boycott of Yugoslavia was begun by the socialist countries.

Expulsion from the Cominform

Stalin's expectation that the boycott would result in the downfall of the Tito regime was unfounded. The wartime experience had given the Yugoslav leadership both the internal solidarity and the popular loyalty necessary to survive the resulting enormous strains. A purge was undertaken of "Cominformist" sympathizers. Steps were also taken to secure diplomatic backing and military and financial aid from the West by modifying the intransigence of Yugoslav foreign policy, accepting Allied dispositions in relation to Trieste, and withdrawing support from the Greek Communists.

To cope with the enforced economic isolation a program of industrialization was begun, which included the collectivization of agriculture. The ill-planned, insensitive, and hasty execution of the collectivization plan led to its abandonment by 1953. In spite of this disaster the regime survived, and the experience turned out to be critical in launching the country on its distinctive course of economic and political development.

The Yugoslav road to socialism. Although Yugoslavia's first response to the attack of the Cominform was defensive, it quickly developed confidence in its own superior socialist legitimacy. In June 1950 the Basic Law on the Management of State Economic Enterprises by Working Collectives took the first steps toward what has since come to be known as the system of self-management. A similarly important change initiated at the same time was the reform of local government in the constitutional law of 1953. This began the process of political decentralization in Yugoslavia, giving enormous powers to the *opština* (commune). Political decentralization and the system of self-management came to embody the major strengths and the weaknesses of the Yugoslav system.

Following the shock of 1948 an early Yugoslav response to foreign policy problems was to seek the support of its neighbours in a Balkan Pact, signed with Greece and Turkey in 1953. A more significant possibility began to emerge after Tito's visit to Burma in January 1955, when

he began to discover the similarity between the Yugoslav situation and that of several other (principally former colonial) states. The movement of the nonaligned countries was given greater coherence at the Bandung Conference of Afro-Asian nations in April 1955.

The nonaligned movement

Notwithstanding the dramatic apology for Soviet actions delivered by Nikita Khrushchev during his visit to Belgrade in May 1955, the Yugoslavs discovered in nonalignment a foreign policy stance that legitimated their independence from the Soviet Union, gave them an important role in international affairs, and underlined in their external relations the respect for national identity that had become so central to domestic policy. The Yugoslavs resisted the invitation to draw back into the Soviet fold, and, together with the heads of state of Egypt, India, Ghana, Indonesia, and Ethiopia, Tito worked to create a third force of nonaligned states, which was launched at their summit conference in Belgrade in September 1961.

That ideological conformity was still required was demonstrated by the case of Milovan Djilas, a veteran Communist of prewar days and the liberation struggle who had become vice president after the war. In a series of highly critical articles published in the party paper *Borba* during the winter of 1953–54, Djilas attacked the moral degeneration of the party and demanded greater democratization of Yugoslav society. He was dismissed from public office and soon withdrew from the party. Criminal charges followed in 1955 and 1956 when Djilas gave interviews to foreign journalists that were alleged to contain hostile propaganda. His book *The New Class* (1957) alleged that the Communist bureaucracy had acquired the characteristics of an exploiting class. For this Djilas received an additional jail sentence, and he was the focus of controversy for his dissident views on several occasions thereafter.

During the 1950s the basic elements of the "Yugoslav road to socialism" were put in place: self-management, decentralization, and nonalignment. While a good deal of ideological freedom was permitted, it became clear that there was a difference between criticism within the system and criticism of the system.

Economic development and economic reform. The enormous task of postwar economic reconstruction was tackled with great determination, and with assistance from the United Nations Relief and Rehabilitation Administration great strides were made during the period of the first Five Year Plan, begun in 1947. The central aim of development was the creation of a base of domestic heavy industry. Although investment in communication, health, and education also received a high priority, other specifically economic goals were subordinated to this aim.

The planning assumptions of this period were challenged by the break with the Cominform, and during the following decade a looser system of planning control was developed, with more initiative devolved to the enterprises, the local authorities, and the highly decentralized banking structure. Measured in terms of growth rates, the economic performance of the decade was impressive. It did bring with it important problems, however. By the early 1960s the economy was overheating, and imports were not matched by exports, the deficit being financed by foreign loans. Protectionism had created industry that was often inefficient, and it shielded this industry from the consequences of poor location policy. Tentative efforts at reform in 1961–62, abolishing income controls and liberalizing foreign trade, made the position worse.

The constitution of 1963

In an attempt to escape from this situation, a new constitution was introduced in 1963 that strengthened self-management and extended its provisions beyond industrial organizations into services and the administration. (The constitution also gave greater importance to the republics and autonomous provinces.) An ambitious economic reform introduced in 1965 was designed to move the country toward "market socialism" by abolishing many price controls and requiring enterprises to compete more effectively with each other and in the "international division of labour." To some extent the reform program was a success, in that the late 1960s were years of unparalleled prosperity. Some economic sectors did particularly well: Yugoslavia emerged as a major international tourist destination, and some branches of manufacture, such as metal goods and textiles, became highly profitable on both the domestic and foreign markets.

Industrialization and urbanization had created a country that was radically different in standard of living and style of life from the economically backward peasant economy of the prewar years. By 1961 the proportion of the population dependent upon agriculture had fallen to 49 percent and by 1971 to 36 percent. The rural sector contributed only a fifth to the total social product by the end of the 1960s.

The reform program had only partly tackled the main issues, however, and it had done nothing to correct certain fundamental weaknesses. Unemployment emerged as a serious problem, only partly alleviated by the flow of migrant workers to other European countries. Instead of creating a genuine market, the strength of the republics resulted in a series of local monopolies in many products. Most seriously, the country's prosperity was not shared evenly among the regions, and the northern republics of Slovenia and Croatia began to draw steadily away from the others in the level of their development and their standard of living. Patterns of economic development thus tended to follow the deeply rooted historical cleavages among the regions. Efforts to correct this situation through the creation in 1961 of a federal fund for underdeveloped areas became particularly controversial, for the more developed republics resented the diversion of their resources into projects in the poorer regions, where they had no say in the choice of projects to be supported, and because there was no clear accountability for spending.

By the end of the 1960s inflation was added to other economic problems. A price freeze was imposed in 1970, and a general package of stabilization measures was introduced the following year. In 1973 the rapid rise of oil prices on the world market hit Yugoslavia hard.

Crisis and dissent. The mid-1960s were years of greater openness of cultural and political expression, evidenced by the growth of an independent religious press (most notably the Roman Catholic paper *Glas Concila*) and critical publications such as the journal of political philosophy *Praxis*. The arts experienced a period of experimentation and became significant, if often controversial, vehicles for dissent. Yugoslavia's reputation for liberalism was enhanced in 1966 when Alexander Ranković, one of the inner circle of the Communist leadership from the wartime period and head of the internal security services, was dismissed following allegations of the abuse of his powers and the service purged.

The growing economic crisis contributed to the sharpening of political conflict. There were massive student demonstrations in 1968, and although these met with a favourable response from some leading Communists, demands for increasing regional autonomy on the part of Kosovo's Albanians met with stern repression.

Unrest in Croatia

Matters came to a head in 1970–71 in Croatia when a powerful mass movement emerged demanding greater political autonomy and economic autarchy for the republic. Fearing a return to the interethnic strife of the prewar period, the Communist Party suppressed these manifestations of dissent, abolished the influential Croatian cultural society Matica Hrvatska, and purged the Croatian Party.

The response to the growing difficulties of the country took the characteristic form of a campaign to account for the deficiencies of the system, not by examining its structure but by insisting that self-management had to be made to work properly. In 1971 a series of amendments to the constitution were introduced, and these were consolidated in a new constitution in 1974. This was followed in 1976 by the Law on Associated Labour. These measures gave contemporary form to the Yugoslav system and provided an institutional framework for handling the question of succession to the presidency.

The year 1971 also saw the beginning of a campaign to tighten party discipline and assert its ideological authority. This development was misplaced, however, as the new institutional structure confirmed the republicanization of Yugoslav politics and paved the way for the growing political autonomy of the eight republican and provincial

parties. Repressive measures such as the disciplining of a group of dissident professors at the University of Belgrade in 1975 were therefore purely cosmetic in character and bore no relation to the real problems of the lack of economic and political coordination in Yugoslavia.

The death of Tito and after. The death of President Tito in May 1980 signalized the passing of a generation that had created modern Yugoslavia, a generation whose wartime experiences provided it with a common culture and the basis of collective authority. The new leadership had to make its mark by effectiveness alone, without the Partisan mystique.

Tito's death also pushed to the fore important questions about the mechanisms by which decisions were taken, questions that had been masked by his personal authority. As long as Tito was at the helm, it was impossible to examine properly the extent to which self-management provided the most appropriate answers to questions about economic accountability, or the degree to which the arrangements regarding republican and communal autonomy guaranteed the rights and possibilities for the development of different ethnic groups.

In 1981 disturbances in Kosovo underlined once again the intractability of problems related to Yugoslavia's ethnic diversity. The issue of the proper structure of the Serbian republic provoked major political crises in 1987 and 1988, in both Montenegro and the Vojvodina as well as in Kosovo and Serbia.

In 1983 the country's foreign debt had become such a problem that the International Monetary Fund was asked to intervene, and throughout the 1980s successive governments struggled to formulate and implement programs of economic reform. These events culminated in the unprecedented resignation of the cabinet of Branko Mikulić in December 1988.

While it became obvious that neither the problem of Yugoslavia's ethnic diversity nor that of its economic management could be easily solved, the end of the 1980s at least saw the way clearing toward a more frank appraisal of the institutional means through which they might be addressed. This process was assisted greatly by wider changes throughout the Soviet bloc, stimulated by reforms in the Soviet Union itself. In the attempt to reconcile the demands of economic development, the creation of an effective state, and the cultural autonomy of its constituent nations, Yugoslavia had moved from the earlier administrative socialism to self-management. Neither of these programs had been conspicuously successful, however.

For later developments in the history of Yugoslavia, see the *Britannica Book of the Year* section in the BRITANNICA WORLD DATA ANNUAL.

For coverage of related topics in the *Macropædia* and *Micropædia,* see the *Propædia,* sections 921, 924, 962, 963, and 972, and the *Index.* (J.B.A.)

BIBLIOGRAPHY

Physical and human geography: Overviews of the country may be found in LOVETT F. EDWARDS, *Yugoslavia* (1971); FRED SINGLETON, *Twentieth Century Yugoslavia* (1976); and RICHARD F. NYROP (ed.), *Yugoslavia: A Country Study,* 2nd ed. (1982). Works on the geography include ROY E.H. MELLOR, *Eastern Europe: A Geography of the COMECON Countries* (1975); DEAN S. RUGG, *Eastern Europe* (1985); DAVID TURNOCK, *Eastern Europe: An Historical Geography, 1815–1945* (1989); and FRANCIS W. CARTER (ed.), *An Historical Geography of the Balkans* (1977). Statistical data are provided by *Statistički Godišnjak Jugoslavije* (annual); an English translation of the prefatory material, table of contents, and captions is published in companion issues titled *Statistical Yearbook of the Socialist Federal Republic of Yugoslavia* (annual). A useful summary is *Yugoslavia 1945–1985: Statistical Review* (1986), published by the Federal Statistical Office. JOHN J. HORTON, *Yugoslavia* (1977), is a valuable bibliographical guide.

Sociological studies include VERA ST. ERLICH, *Family in Transition: A Study of 300 Yugoslav Villages* (1966; originally published in Serbo-Croatian, 1964); and JOEL HALPERN and BARBARA KEREWSKY HALPERN, *A Serbian Village in Historical Perspective* (1972, reissued 1986). The relation between religion and government is examined in STELLA ALEXANDER, *Church and State in Yugoslavia Since 1945* (1979). Aspects of economic, political, and social conditions are discussed in JOZO TOMASEVICH, *Peasants, Politics, and Economic Change in Yugoslavia* (1955, reprinted 1975); LARS NORD, *Nonalignment and Socialism: Yugoslav Foreign Policy in Theory and Practice* (1973); *Constitutional System of Yugoslavia,* trans. from Serbo-Croatian (1974, reprinted 1980), a special issue of *Yugoslav Survey* (quarterly); PETER JAMBREK, *Development and Social Change in Yugoslavia: Crises and Perspectives of Building a Nation* (1975); LJUBO SIRC, *The Yugoslav Economy Under Self-Management* (1979); and FRED SINGLETON and BERNARD CARTER, *The Economy of Yugoslavia* (1982).

Socialism as practiced in Yugoslavia is analyzed in SHARON ZUKIN, *Beyond Marx and Tito: Theory and Practice in Yugoslav Socialism* (1975); GERSON S. SHER (ed. and trans.), *Marxist Humanism and Praxis* (1978), a collection of essays by members of the dissident group; APRIL CARTER, *Democratic Reform in Yugoslavia: The Changing Role of the Party* (1982); and HAROLD LYDALL, *Yugoslav Socialism: Theory and Practice* (1984).

History: General coverage is presented by L.S. STAVRIANOS, *The Balkans Since 1453* (1958, reissued 1965); CHARLES JELAVICH and BARBARA JELAVICH (eds.), *The Balkans in Transition: Essays on the Development of Balkan Life and Politics Since the Eighteenth Century* (1963, reprinted 1974); STEVAN K. PAVLOWITCH, *Yugoslavia* (1971); VLADIMIR DEDIJER *et al., History of Yugoslavia* (1974; originally published in Serbo-Croatian, 1972); and FRED SINGLETON, *A Short History of the Yugoslav Peoples* (1985). JOHN ALEXANDER, *Yugoslavia Before the Roman Conquest* (also published as *Jugoslavia Before the Roman Conquest,* 1972), surveys archaeology and prehistory. Byzantium and the early Slav states are described in GEORGE OSTROGORSKY, *History of the Byzantine State,* 2nd ed. (1968; originally published in German, 3rd ed., 1963); DMITRI OBOLENSKY, *The Byzantine Commonwealth: Eastern Europe, 500–1453* (1971), and *The Bogomils: A Study in Balkan Neo-Manichaeism* (1948, reprinted 1978); FRANCIS DVORNIK, *The Slavs in European History and Civilization* (1962); and A.P. VLASTO, *The Entry of the Slavs into Christendom: An Introduction to the Medieval History of the Slavs* (1970). Works on the Ottoman Empire include PETER F. SUGAR, *Southeastern Europe under Ottoman Rule, 1354–1804* (1977); and BRUCE MCGOWAN, *Economic Life in Ottoman Europe: Taxation, Trade, and the Struggle for Land, 1600–1800* (1981).

Yugoslavia from the 14th century to 1914 is studied by STANKO GULDESCU, *History of Medieval Croatia* (1964), and *The Croatian-Slavonian Kingdom, 1526–1792* (1970); GUNTHER E. ROTHENBERG, *The Austrian Military Border in Croatia, 1522–1747* (1960), and *The Military Border in Croatia, 1740–1881: A Study of an Imperial Institution* (1966); FRANCIS W. CARTER, *Dubrovnik (Ragusa): A Classic City-State* (1972); LEOPOLD RANKE, *A History of Servia and the Servian Revolution,* 2nd ed. (1848, reprinted 1973; originally published in German, 1829); HAROLD W.V. TEMPERLEY, *History of Serbia* (1917, reprinted 1970); GEORGES CASTELLAN, *La Vie quotidienne en Serbie au seuil de l'independance, 1815–1839* (1967); PETER F. SUGAR, *Industrialization of Bosnia-Hercegovina, 1878–1918* (1964); FRANCIS SEYMOUR STEVENSON, *A History of Montenegro* (1912, reprinted 1971); ELISABETH BARKER, *Macedonia: Its Place in Balkan Power Politics* (1950, reprinted 1980); and H.R. WILKINSON, *Maps and Politics: A Review of the Ethnographic Cartography of Macedonia* (1951).

Appraisals of Yugoslavia's role in World War I are found in VLADIMIR DEDIJER, *The Road to Sarajevo* (1966); R.W. SETON-WATSON, *Sarajevo: A Study in the Origins of the Great War* (1926, reprinted 1973); and IVO J. LEDERER, *Yugoslavia at the Paris Peace Conference: A Study in Frontiermaking* (1963). The period between the wars is portrayed in HUGH SETON-WATSON, *Eastern Europe between the Wars, 1918–1941,* 3rd ed. rev. (1962, reissued 1986); IVO BANAC, *The National Question in Yugoslavia: Origins, History, Politics* (1984); JACOB B. HOPTNER, *Yugoslavia in Crisis, 1934–1941* (1962); I. AVAKUMOVIC, *History of the Communist Party of Yugoslavia,* vol. 1 (1964); and DRAGIŠA N. RISTIĆ, *Yugoslavia's Revolution of 1941* (1966). The impact of World War II is examined in ILIJA JUKIĆ, *The Fall of Yugoslavia* (1974); PERO MORAČA and VIKTOR KUČAN, *The War and Revolution of the Peoples of Yugoslavia, 1941–1945* (1962?); PHYLLIS AUTY and RICHARD CLOGG (eds.), *British Policy Towards Wartime Resistance in Yugoslavia and Greece* (1975); JOZO TOMASEVICH, *The Chetniks* (1975), on the Yugoslav resistance movement; and WALTER R. ROBERTS, *Tito, Mihailović, and the Allies, 1941–1945* (1973, reprinted 1987).

Discussions of the postwar period can be found in DENNISON RUSINOW, *The Yugoslav Experiment, 1948–1974* (1977); and BRANKO HORVAT, *An Essay on Yugoslav Society* (1969; originally published in Serbo-Croatian, 1969). Analyses of contemporary problems include HAROLD LYDALL, *Yugoslavia in Crisis* (1989); MARKO MILIVOJEVIĆ, JOHN B. ALLCOCK, and PIERRE MAURER (eds.), *Yugoslavia's Security Dilemmas: Armed Forces, National Defence, and Foreign Policy* (1988); and STEVAN K. PAVLOWITCH, *The Improbable Survivor: Yugoslavia and Its Problems, 1918–1988* (1988). (J.B.A.)

Zoroastrianism and Parsiism

Zoroastrianism is the major ancient, pre-Islāmic religion of Iran. It survives there in isolated areas but also, more prosperously, with the Parsees, or Parsis (hence Parsiism), of India, descendants of immigrants who went there from Iran some time after the Muslim conquest. In modern times, a few adherents have transported the religion into the West. For a discussion of the context in which Zoroastrianism arose, see MIDDLE EASTERN RELIGIONS, ANCIENT: *Iranian religions.*

This article is divided into the following sections:

Nature and significance 1116
History 1116
  Pre-Zoroastrian Iranian religion
  The reformation of Zoroaster
  The Arsacid period
  The Sāsānian period
  Post-Islāmic Iranian Zoroastrianism
  The Parsis in India
Beliefs and mythology 1118
  Sources
  God
  Cosmogony
  Cosmology
  Concepts of man
  Eschatology
Practices and institutions 1119
  Cultic places
  Priesthood
  Ceremonies
  Burial rites
  Festivals
  Ethics
Iconography 1121
Relation to other religions 1121
Conclusion 1121
Bibliography 1121

NATURE AND SIGNIFICANCE

The ancient Greeks saw in Zoroastrianism the archetype of the dualistic view of the world and of man's destiny. Zoroaster was supposed to have instructed Pythagoras in Babylon and to have inspired the Chaldean doctrines of astrology and magic. It is likely that Zoroastrianism influenced the development of Judaism and the birth of Christianity. The Christians, following a Jewish tradition, identified Zoroaster with Ezekiel, Nimrod, Seth, Balaam, and Baruch, and even, through the latter, with Christ himself. On the other hand, Zoroaster, as the presumed founder of astrology and magic, could be considered the arch-heretic. In more recent times the study of Zoroastrianism has played a decisive part in reconstructing the religion and social structure of the Indo-European peoples.

Monotheism and dualism

Though Zoroastrianism was never, even in the thinking of its founder, as aggressively monotheistic as, for instance, Judaism or Islām, it does represent an original attempt at unifying under the worship of one supreme god a polytheistic religion comparable to those of the ancient Greeks, Latins, Indians, and other early peoples.

Its other salient feature, namely dualism, was never understood in an absolute, rigorous fashion. Good and Evil fight an unequal battle in which the former is assured of triumph. God's omnipotence is thus only temporarily limited. In this struggle man must enlist because of his capacity of free choice. He does so with his soul and body, *not* against his body, for the opposition between good and evil is not the same as the one between spirit and matter. Contrary to the Christian or Manichaean (from Manichaeism—a Hellenistic, dualistic religion founded by the Iranian prophet Mani) attitude, fasting and celibacy are proscribed, except as part of the purificatory ritual. Man's fight has a negative aspect, nonetheless: he must keep himself pure; *i.e.*, avoid defilement by the forces of death, contact with dead matter, etc. Thus Zoroastrian ethics, although in itself lofty and rational, has a ritual aspect that is all-pervading. On the whole, Zoroastrianism is optimistic and has remained so even through the hardship and oppression of its believers.

HISTORY

Pre-Zoroastrian Iranian religion. The religion of Iran before the time of Zoroaster is not directly accessible, for there are no reliable sources more ancient than the prophet himself. It has to be studied indirectly on the basis of later documents and by a comparative approach. The language of Iran is closely akin to that of northern India, and hence the people of the two lands probably had common ancestors—the Indo-Iranians, or Aryans. The religion of the latter has been reconstructed by means of common elements contained in the sacred books of Iran and India: mainly the Avesta and the Vedas. Both collections exhibit the same kind of polytheism, with many of the same gods, notably the Indian Mitra (the Iranian Mithra), the cult of fire, sacrifice by means of a sacred liquor (*soma* in India, in Iran *haoma*), and other parallels. There is, moreover, a list of Aryan gods in a treaty concluded about 1380 BC between the Hittite emperor and the king of Mitanni. The list includes Mitra and Varuṇa, Indra, and the two Nāsatyas. All of these gods also are found in the Vedas, but only the first one in the Avesta, except that Indra and Nāñhaithya appear in the Avesta as demons; Varuṇa may have survived under another name. Important changes, then, must have taken place on the Iranian side, not all of which can be attributed to the prophet.

Daivas and *asuras*

The Indo-Iranians appear to have distinguished, from among their gods, the *daiva* (Indo-Iranian and Old Persian equivalent of Avestan *daeva* and Sanskrit *deva,* related to the Latin *deus*), meaning "heavenly," and the *asura,* a special class with occult powers. This situation was reflected in Vedic India; later on, *asura* came to signify, in Sanskrit, a kind of demon, because of the baleful aspect of the *asura*'s invisible power. In Iran the evolution must have been different: the *ahuras* were extolled, to the exclusion of the *daevas*, who were reduced to the rank of demons.

The reformation of Zoroaster. Zoroaster (Zarathushtra) was a priest of a certain *ahura* (Avestan equivalent of Sanskrit *asura*) with the epithet *mazdā,* "wise," whom Zoroaster mentions once in his hymns with "the [other] *ahuras.*" Similarly, Darius I (522–486) and his successors worshipped Auramazda (Ahura Mazdā) "and the other gods who exist" or "Ahura Mazdā, the greatest god." The two historically related facts are evidently parallel: on both sides the rudiments of monotheism are present, though in a more elaborate form with the prophet Zoroaster.

It has not yet been possible to place Zoroaster's hymns, the *Gāthās*, in their historical context. Not a single place or person mentioned in them is known from any other source. Vishtāspa, the prophet's protector, can only be the namesake of the father of Darius, the Achaemenid king. All that may safely be said is that Zoroaster lived somewhere in eastern Iran, far from the civilized world of western Asia, before Iran became unified under Cyrus II the Great. If the Achaemenids ever heard of him, they did not see fit to mention his name in their inscriptions nor did they allude to the beings who surrounded the great god and were later to be called the *amesha spentas*, or "bounteous immortals"—an essential feature of Zoroaster's doctrine.

Religion under the Achaemenids was in the hands of the Magi, whom Herodotus describes as a Median tribe with special customs, such as exposing the dead, fighting evil animals, and interpreting dreams. Again, the historical con-

nection with Zoroaster—whom Herodotus also ignores—is a hazy one. It is not known when Zoroaster's doctrine reached western Iran, but it must have been before the time of Aristotle (384–322), who alludes to its dualism.

Darius, when he seized power in 522, had to fight a usurper, Gaumata the Magian, who pretended to be Bardiya, the son of Cyrus the Great and brother of the king Cambyses. This Magian had destroyed cultic shrines, *āyadanas*, which Darius restored. One possible explanation of these events is that Gaumata had adopted Zoroastrianism, a doctrine that relied on the allegiance of the common people, and therefore destroyed temples or altars to deities of the nobility. Darius, who owed his throne to the support of some noblemen, could not help favouring their cult, although he adopted Auramazda as a means of unifying his empire.

Xerxes, successor to Darius, mentioned in one of his inscriptions how at a certain (unnamed) place he substituted the worship of Auramazda for that of the *daivas*, which does not mean he opposed the *daeva* cult as such, as a true Zoroastrian would have done, but only that he eradicated somewhere—probably in Babylon—the cult of deities alien to the religion of the *ahuras*. It points to a change of attitude, compared with Cyrus' tolerance of alien religions, such as the Babylonian or the Jewish religions.

From Artaxerxes II (404–359/358) onward, the inscriptions mention, besides Auramazda, Mithra and the goddess Anahita (Anahit), which proves only a change of emphasis, not the appearance of new deities.

The Arsacid period. In consequence of Alexander's conquest, the Iranian religion was almost totally submerged by the wave of Hellenism. At Susa, for instance, which had been one of the capital cities of the Achaemenids but where the religion of Auramazda was not indigenous, the coinage of the Seleucid and Arsacid periods does not represent a single Iranian deity.

Then the Iranian religion gradually emerged again. In Commagene in the middle of the 1st century BC, gods bear combinations of Greek and Iranian names: Zeus Oromazdes, Apollo Mithra, Helios Hermes, Artagnes Herakles Ares. The first proof of the use of a Zoroastrian calendar, implying the official recognition of Zoroastrianism, is found some 40 years earlier at Nisa (near modern Ashkhabad in Soviet Turkmenistan). By then some form of orthodoxy must have been established in which Auramazda and the entities (powers surrounding him) adjoin other gods such as Mithra, the Sun, and the Moon.

Hellenistic syncretism

In Persis (modern Fars), from the beginning of the Christian Era to the advent of the Sāsānians (early 3rd century AD), any allusion to the fire cult disappears. The coins seem to indicate, in not showing the fire altar, that the prince had lost interest in the Iranian religion.

The Sāsānian period. With Ardashīr, the future founder of the Sāsānian dynasty, the situation was different; and this may suggest that his religious zeal—as a hereditary priest of Staxr (Istaxr)—may have helped him seize power in his native province, even before he started attacking his Arsacid suzerain, Artabanus V.

Two persons are recorded, in different sources, as helping to establish Zoroastrianism under the first Sāsānians: Kartēr and Tansar. Whereas Kartēr is known through contemporary inscriptions, most of which were written by himself, Tansar (or Tosar) is only remembered in later books. The latter tell us that Tansar, an *ehrpat,* or theologian, undertook the task, under Ardashīr's command, of collecting the sacred texts and fixing the canon. Kartēr, who was already active under Ardashīr I but more so under Shāpūr and his successors, recounted his brilliant career, which reflects the birth of a hierarchy. He was still an *ehrpat* under Shāpūr, as he restored the "Mazdean religion . . . in the land of non-Iran reached by the horses and men of the king of kings." Under Hormizd he was made "*magupat* of Ormazd," a term apparently created for him and meaning "chief of the Magians of Auramazda." Under Bahrām I (AD 273–276), Mani, the founder of Manichaeism, who had enjoyed a degree of tolerance under the two preceding kings, was sacrificed to the interests of Zoroastrianism and died in prison. Bahrām II named Kartēr "Saviour of the Soul of Bahrām," elevated him to the rank of the "grandees of the realm," and gave him the additional titles of "judge of the empire," "master of rites," and "ruler of the fire of Anahit-Ardashīr at Staxr and of Anahit the Dame." Promoted to the apex of his career, Kartēr persecuted "Jews, Buddhists, Brahmins, Nasoreans [Judeo-Christians?], Christians, Maktaks [Mandeans, Manichaeans?], and Zandīks [Mazdean heretics]." Narses (293–302), who began his struggle for power when Bahrām II was still on the throne, seems to have recovered the title of chief of the Staxr temple that his predecessor and adversary had surrendered to Kartēr. Under Shāpūr II, the high priest Aturpāt, at a council summoned to fix the text of the Avesta, proved the truth of his doctrine by submitting to the ordeal of molten metal poured on his breast and was victorious over all kinds of sectarians and heretics.

Under Bahrām V (420–438), presumably, the title *magupatān magupat* (chief magus of the chief magi) was created. Under Qobād (or Kavādh; 488–496 and 498/499–531), Iran traversed its gravest social and religious crisis under the impact of Mazdak. This reformer, whose doctrines were partly inspired by those of Mani, was granted an interview by Qobād—as Shāpūr I had received Mani a long time before, but with a more decisive success. Perhaps the King hoped that by abolishing property and the family he would reign over a docile mass. The Mazdakites favoured the abolition of all social inequalities, chiefly of private property, the main cause of all hatred. Everything was to be held in common, including women. These views directly threatened the rich as well as the Mazdean clergy, who soon understood this. Qobād was dethroned and replaced by his brother Jāmāsp. After two years in exile, Qobād recovered his throne, but he had been cured of his egalitarian views and decided to liquidate the Mazdakites.

Mazdak and Mazdakism

Khosrow I continued the work of his father, Qobād, and thus the Mazdakite upheaval made way for a strong state and an established Mazdean Church. The religious books give Khosrow the unique title of *Anōsharvan,* "with the immortal soul," probably for having crushed Mazdakism and for enabling the "good religion" to triumph.

Khosrow II (590/591–628) married a Christian woman and may have been a Christian himself. He was superstitious and dabbled in astrology.

Post-Islāmic Iranian Zoroastrianism. Islām won a decisive victory at al-Qādisīyah in 635 over the armies of Yazdegerd III, the last Sāsānid. Islām, in principle, tolerated the ancient religion, but conversions by persuasion or force were massive in many provinces. Zoroastrianism fomented rebellion and brought persecutions upon itself. There were pockets of survival, notably in Persis, the ancient centre of the Achaemenian and Sāsānian empires. Books were produced to save the essentials of the religion from a threatened disaster. The disaster did occur but exactly why and how is not known. Zoroastrians, called Gabars by the Muslims, survived in Iran as a persecuted minority in small enclaves at Yazd and Kerman.

The Parsis in India. From the 10th century onward, groups of Zoroastrians emigrated to India, where they found asylum in Gujarāt. Their connection with their coreligionists in Iran seems to have been almost totally broken until the end of the 15th century. Reestablished in 1477, the connection was kept up chiefly in the form of an exchange of letters until 1768. Under British rule, the Parsis, who previously had been humble agriculturists, started to enrich themselves through commerce, then through industry. They became a most prosperous and "modern" community, centred in Bombay. Formerly they had adopted the language (Gujarati) and the dress of their Hindu milieu. Later they adopted British customs, British dress, the education of girls, and the abolition of child marriage. In their enterprises as well as in their charities they followed the example of the West. From the 19th century on, they were able to help their less favoured brethren in Iran, either through gifts or through intervention with the government.

They also adapted themselves to their Indian culture by minimizing what was repugnant to the Hindus, namely, blood sacrifice; and they surrendered to some extent to the vogue of astrology and to theosophy. On the other

hand, ever since they were attacked by Christian missionaries for their dualism, they have been emphasizing the monotheistic aspect of their doctrine.

BELIEFS AND MYTHOLOGY

Sources. Only the hymns, or *Gāthās*, are attributable to Zoroaster. They are written in various metres and in a dialect different from the rest of the Avesta, except for seven chapters, chiefly in prose, that appear to have been composed shortly after the prophet's demise. All these texts are embedded in the *Yasna,* which is one of the main divisions of the Avesta and is recited by the priests during the ceremony of the same name, meaning "sacrifice." The *Visp-rat* ("All the Judges") is a *Yasna* augmented here and there by additional invocations and offerings to the *ratu*s (lords) of the different classes of beings. The *Vidēvdāt,* or *Vendidad* ("Law Rejecting the Daevas"), consists of two introductory sections recounting how the law was given to man, followed by 18 sections of rules. The *Siroza* enumerates the deities presiding over the 30 days of the month. The *Yasht*s (hymns) are each addressed to one of 21 deities such as Mithra, Anahita, or Verethraghna. The *Hadhoxt Nask* ("Section Containing Sayings") describes the fate of the soul after death. The *Khūrda Avesta,* or Small Avesta, is made up of minor texts.

The Avesta

The Avesta is, therefore, a collection of texts compiled in successive stages until it was completed under the Sāsānians. It was then about four times larger than what has survived. A summary of its 21 books, or *Nask*s (of which only one is preserved as such in the *Vidēvdāt*), is given in one of the main treatises written during the brief Zoroastrian renascence under Islām in the 9th century; the *Dēnkart,* the "Acts of the Religion." It is written in Pahlavi, the language of the Sāsānians.

Other works in Pahlavi include, besides a translation and commentary on the Avesta, the *Bundahishn* ("Primal Creation"), a cosmology. Most Pahlavi books are anonymous, such as *Mēnōk-i Khrat* ("Spirit of Wisdom"), a lucid summary of a doctrine based on reason, and the *Book of Artāy Virāf,* which describes Virāf's descent into the netherworld as well as heaven and hell and the pleasures and pains awaiting the virtuous and the wicked. There are also a few signed works, such as those of the two brothers Zātspram and Mānushchihr, or Mardān-Farrukh's *Shkand-Gumānīk Vichār* ("Final Dispelling of Doubts"), an apology of the Mazdean religion directed against Manichaeism, Christianity, Judaism, and Islām.

Finally, there are Zoroastrian books written in Persian, either in verse or in prose. The latter include the correspondence exchanged between Zoroastrians of Iran and India and the treatise entitled *'Olemā-ye Islām* ("The Doctors of Islām"), with decidedly Zurvanite tendencies.

Mithra

God. Zoroaster's silence on Mithra is not easy to interpret. Since this god was closely associated with Varuṇa in India and with Varuṇa's likely substitute in Iran, Zoroaster can hardly have ignored one-half of this divine pair without a definite purpose. Otherwise, it might be presumed that Mithra was included in the formula "Mazdā and the [other] *ahura*s"; however, Mithra is called in the Later Avesta (non-Gāthic) an *ahura;* so is Apām Napāt, a fire or brightness in the waters, corresponding to the Vedic Apām Napāt. As for Verethraghna (the entity or spirit of victory), it seems that since he took over the function of Indra, who was a *daeva,* he could not be called an *ahura*; but in order to mark his belonging to the world of *ahura*s he was called *ahuradāta*, "created by an *ahura.*"

It is in the framework of the religion of the *ahura*s, hostile to the cult of the *daeva*s, that Zoroaster's message should be understood. He emphasized the central importance of his god, the wise Ahura, by portraying him with an escort of entities, the powers of all the other gods, in an array against the forces of evil.

The moral dualism expressed in the opposition Asha–Druj (truth–falsehood) goes back at least to Indo-Iranian times, for the Veda knows it too, as *ṛta–druh,* although the contrast is not as sharply defined as in the Avesta. Between these two principles, the Twin Spirits made an ominous choice, the Bounteous One becoming in thoughts, words, and deeds a partisan of Asha, *ashavan,* while the other became *dregvant,* partisan of the Druj. After them it was the *daeva*s' turn; they all chose wrongly. Ever since, the *daeva*s have tried to corrupt man's choice also.

To the army of the *ashavan*s, headed by the Bounteous Spirit, was counterposed the host of the *dregvant*s, under the Destructive Spirit, Angra Mainyu. Each combatant faced his exact counterpart: the Good Mind opposing the Bad Mind and Aramaiti being countered by Taromaiti.

In this battle, the whole material universe is, through the entities, potentially enrolled, the Bounteous Spirit being the patron of man, Asha of fire, the Good Mind of the Ox, the Dominion of the metals, Aramaiti of the earth, Integrity and Immortality of the waters and plants. Moreover, since the entities are at once divine and human (because both the spiritual and material qualities of man partake of divine), everyone faithful to the wise Ahura can commune with him.

After Zoroaster, considerable changes occurred in the theology he had professed. The entities were reduced to mere deities, which were even separated into male and female. Never again were their names used to designate human faculties. This is probably a consequence of the resurgence of the ancient gods.

It is not known to what extent Zoroaster's system was meant to be exclusively the cult of Ahura Mazdā. In the Later Avesta all the gods he had ignored emerged again, such as Mithra, Airyaman (whom he had replaced by Sraosha), Anahita, Apām Napāt, Verethraghna, and Vayu. This vast pantheon, still nominally headed by Ahura Mazdā, is similar to the compromise that Darius, according to the interpretation cited above, made between the cult of Auramazda and that of the gods of the nobility.

Not only did Zoroaster's theology thus lose its exclusive position, but an internal change also modified its equilibrium and even threatened its very essence. The Bounteous Spirit was almost completely reabsorbed into Ahura Mazdā. Whereas in a *Yasht* the two Spirits fought each other, in the *Vidēvdāt* Ahura Mazdā and the Destructive Spirit opposed each other by creating, respectively, the good and the bad things. This profoundly affected Zoroaster's system, for Ahura Mazdā could no longer be the father of the Twin Spirits; he now faced, on equal terms so to speak, a sort of antigod. This alteration probably dates back at least to the 4th century BC, for Aristotle said in the *Peri philosophias* ("On Philosophy") that the Magi preached the existence of two principles, Oromasdes and Areimanios.

Cosmogony. In the cosmogony as expounded in the *Bundahishn,* Ormazd (Ahura Mazdā) and Ahriman are separated by the void. They seem to have existed from all eternity, when Ahriman's invidious attack initiates the whole process of creation. The question of their origin is ignored, but it was implied, ever since Ormazd had taken the place of his Bounteous Spirit in the struggle against the Destructive Spirit. Since Ahura Mazdā could no longer be the father of the two adversaries, the question of their origin was inevitable.

Zurvanism

A solution was provided by Zurvanism; it is Zurvān (Time) who is the father of Ormazd and Ahriman. But this solution upset the very essence of Mazdaism and was therefore condemned as heretical. Zurvanism was widely accepted, however, perhaps even prevalent, in Sāsānian times. Traces of it are found in Mazdean orthodoxy, some features of which cannot otherwise be explained.

In Mazdean orthodoxy, when Ormazd created the material world, he first produced from Infinite Light a form of fire, out of which all things were to be born. This form of fire is "bright, white, round, and visible from afar." Gayōmart, the Primal Man, was also conceived as spherical, in the image of the sky. Mānushchihr writes that "Ormazd, the lord of all things, produced from Infinite Light a form of fire whose name was that of Ormazd and whose light was that of fire." This phrase can be accounted for only as a clumsy adaptation of a Zurvanite text that must have said, in effect, that Zurvān created Ormazd.

The Mazdean quaternity can hardly be explained except as an adaptation of the Zurvanite one. The latter is attested in several texts citing, besides Zurvān, three other names given as those of separate gods but that must be hypostases

(essences) of the first one, also called in Manichaeism the god with four faces. Among the various forms under which the Zurvanite quaternity manifested itself, the one associating Zurvān with Light, Power, and Wisdom seems to be the origin of the Mazdean quaternity. Ormazd, in the *Bundahishn,* has three other names, namely Time, Space, and Religion. To obtain this quaternity, it was sufficient to replace Zurvān by Time, Light by Space, Wisdom by Religion, and Power by Ormazd and to put the latter at the end of the series.

The Mazdean quaternity is reflected in the calendar at Nisa in 90 BC. The Zurvanite speculation that preceded it probably dates back to the first centuries of the Arsacid period and thus was born in the wake of Hellenism and in connection with the spread of astrology.

Cosmology. In order to vanquish Ahriman, Ormazd created the world as a battlefield. He knew that this fight would be limited in time—it would last 9,000 years—and he offered Ahriman a pact to that effect. After they had created their respective material creations, Ahriman's first attack was defeated by Ormazd with the help of the Ahuna Vairya prayer (the most sacred Zoroastrian prayer), and he lay prostrate for another period of 3,000 years, the second in a total of four. He was then stirred up by the prostitute (Primal Woman) and went back to the attack, this time in the material universe. He killed the Primal Bull, whose marrow gave birth to the plants and whose semen was collected and purified in the moon, whence it would produce the useful animals. Ahriman then killed Gayōmart, the Primal Man, whose body produced the metals and whose semen was preserved and purified in the sun. A part of it would produce the rhubarb from which the first human couple would be born.

Gayōmart, the Primal Man

The first human couple were perverted by Ahriman, and it is only with the advent of Zoroaster, after 3,000 years, that Ahriman's supremacy came to an end. Ormazd and Ahriman then fight on equal terms until Ormazd, at the end of the last 3,000 years, finally will triumph.

Concepts of man. The idea of man as a microcosm, already illustrated in the cosmogony, is further developed in the *Bundahishn.*

As a result of the aggressor's attack, man is mortal. But he does not die altogether. There are five immortal parts in him: *ahu* ("life"), *daēnā* ("religion"), *baodah* ("knowledge"), *urvan* ("soul"), and *fravashi* ("preexistent souls"). The latter term seems literally to mean "preeminent hero." The conception that caused this term to be applied to the "manes" (spirits) or *pitarah* of Iran is that of a defensive, protective power that continues to emanate from a chief even after death. This originally aristocratic notion seems to have been vulgarized in the same way as, in Greece, any dead person came to be considered a hero, or, in Egypt, an Osiris. Zoroaster ignored the *fravashi,* but he was familiar with the *daēnā.* The latter term meant "religion" in both its objective and subjective senses.

The afterlife

Indian and Iranian beliefs in the afterlife have many features in common, probably dating back to the Indo-Iranian period: a feminine encounter, a bridge with dogs watching it, a heavenly journey. In the ancient Indian texts, the *Upaniṣad*s, the soul is welcomed in heaven by 500 *apsara*s (cloud maidens). In Iran the soul meets his own religion (*daēnā*) in the form of a beautiful damsel if he has lived justly; otherwise, he meets a hideous hag.

Either before this encounter or after, according to the various texts, the soul must cross a bridge. This, with the young girl and the gods, is attested in India in the Yajurveda and the *Upaniṣad*s. In the *Gāthā*s it is called the Bridge of the Requiter. It leads the good souls to paradise, but the bad ones fall into hell.

The soul has also to undergo a judgment; it appears before Mithra and his two companions, Sraosha and Rashnu. Finally it ascends through successive stages representing respectively his good thoughts (the stars), good words (the moon), and good deeds (the sun) to the paradise (of infinite lights). In the Veda it is said only that the sojourn of the good deed is beyond the path of the sun. In paradise the soul is led by Vohu Manah, the Good Mind, to the golden throne of Ormazd.

Hell also has, symmetrically, four levels. And there is, for the souls whose good actions exactly balance their evil ones, an intermediate place.

Eschatology. Zoroaster used to invoke saviours who, like the dawns of new days, would come to the world. He hoped himself to be one of them. After his death, the belief in coming saviours developed. Zarathushtra (Zoroaster) was expected to return, if not personally, at least in the form of his three sons who would be born, at intervals of a thousand years, from his semen. The last of these saviours, Astvat-ereta, or justice incarnate, was also simply called the Saviour (Saoshyans).

Only in the Pahlavi books is this theme systematically developed. It is dominated by the idea of a final return to the initial state of things. The first human couple had at first fed on water, then on plants, on milk, and at last on meat. The people in the last millennia will, at the advent of the three successive saviours, abstain in the reverse order from meat, milk, and plants to keep finally only water. The primeval combatants also have their counterparts at the end of time. The dragon that was killed in order to liberate the imprisoned waters will appear again at the resurrection to be killed by another hero. In the last great struggle, the host of good and the host of evil will vie with each other, and each soldier of Ormazd will defeat and kill his own special adversary. This will restore the state of peace that had prevailed initially. The wicked will then submit to an ordeal of molten metal and fire. Fire and Airyaman will cause the metals of the mountains to melt and to flow down as a river of fire. The whole of resuscitated mankind must traverse it; it will burn only the wicked, whereas to the just it will be as sweet as warm milk. The suffering of the wicked will last only three days, however, after which all mankind will enjoy much happiness. On the flattened earth (for the metal will fill in all the valleys), men and women, henceforth shadowless since they are sinless, will taste the bliss of family life. Hell will be sealed forever, and Ahriman will be either powerless or annihilated.

PRACTICES AND INSTITUTIONS

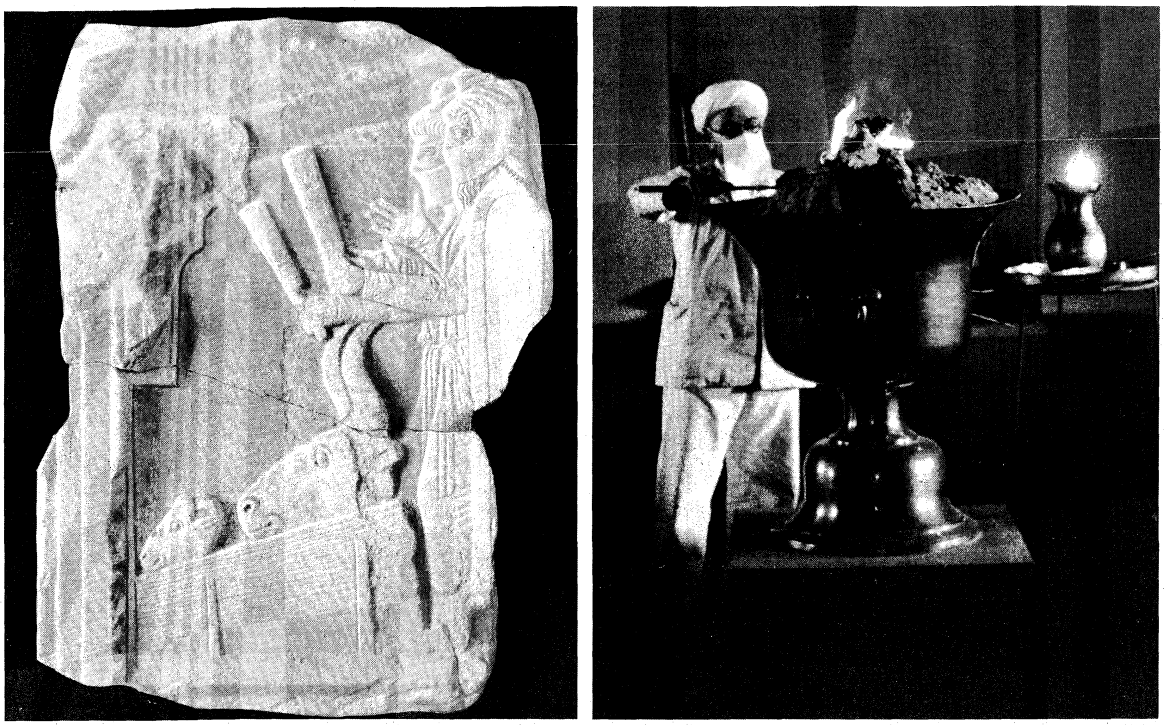
Cultic places. Although Herodotus wrote that the Persians had no temples, some have been found, in the shape of terraces or towers or square rooms. *Chahārtāq*s (sacred buildings with four gates or doors) are scattered over most of Iran. Permanent altars exist from the Sāsānian period and are depicted on coins with a burning fire.

The sacred fire

The Farnbag, Gushnasp, and Burzen-Mihr fires were connected, respectively, with the priests, the warriors, and the farmers. The Farnbag fire was at first in Khwārezm, until in the 6th century BC, according to tradition, Vishtāspa, Zoroaster's protector, transported it to Kabulistan; then Khosrow in the 6th century AD transported it to the ancient sanctuary of Kariyan in Fars. The latter, however, has not yet been identified. The Gushnasp fire, located at Shiz, was the ancient fire of the Magi (in Media), but it came to be the symbol of the monarchic and religious unity. The Burzen-Mihr fire never ranked as high as the other two because the peasants, unlike the kings and the clergy, never possessed any sovereignty. Besides these individual designations, the fires were classified according to two categories: the Adurān, village fires; and the Varhrān, provincial and royal fires.

Magians

Priesthood. The Magians, though not originally Zoroastrian, apparently became acquainted with the prophet's teachings not later than the 4th century BC. They had the monopoly on religion at the Achaemenian court. The term magus was still used in the Arsacid period. Thereafter, under the Sāsānians, a hierarchy developed, with the creation of the *magupat,* or chief of magi, and of its superlative *magupatān magupat* (coined on the model of *shāhanshāh,* "king of kings"). The *ehrpat,* originally a religious teacher, was especially entrusted with the care of the fire. The modern equivalent of the word, *herbad* or *ervad,* designates a priest of the lower degree, who in the more important ceremonies only acts as the assistant priest. Above him is the *mobed.* Ranked above all of these functionaries is the *dastūr,* a kind of bishop, who directs and administers one or more important temples. Priesthood is hereditary, but all priests have to go through one



(Left) Stone relief of the Achaemenid period shows priest wearing mouth cover while presiding at a sacrifice. (Right) Modern Zoroastrian priest wearing similar mouth cover while tending a temple fire.

By courtesy of (left) the Archaeological Museums of Istanbul; photograph (right), Inge Morath—Magnum

or more ceremonies of investiture over and above those practiced by all the faithful.

Ceremonies. All young Parsis must be initiated when they reach the age of seven (in India) or 10 (in Persia). They receive the shirt (*sadre*) and the girdle (*kusti*), which they are to wear their whole life.

There are three types of purification, in order of increasing importance: the *padyab,* or ablution; the *nahn,* or bath; and the *bareshnum,* a complicated ritual performed at special places with the participation of a dog—whose left ear is touched by the candidate and whose gaze puts the evil spirits to flight—and lasting several days.

Penance entails reciting the *patet,* the firm resolve not to sin again, and the confession of sins to a *dastūr* or to an ordinary priest if a *dastūr* is not obtainable.

The chief ceremony, the *Yasna,* essentially a sacrifice of *haoma* (the sacred liquor), is celebrated before the sacred fire with recitation of large parts of the Avesta. There also are offerings of bread and milk and, formerly, of meat or animal fat.

The sacred fire must be kept burning continually and has to be fed at least five times a day. Prayers also are recited five times a day. The founding of a new fire involves a very elaborate ceremony. There are also rites for purification and for regeneration of a fire.

Burial rites. After death, a dog is brought before the corpse; it should preferably be a "four-eyed" dog (*i.e.,* it should have a spot above each eye, as this is said to increase the efficacy of its look). The rite is repeated five times a day. After the first one, fire is brought into the room where it is kept burning until three days after the removal of the corpse to the Tower of Silence. The removal must be done during the daytime.

The interior of the Tower of Silence is built in three concentric circles, one each for men, women, and children. The corpses are exposed there naked. The vultures do not take long—an hour or two at the most—to strip the flesh off the bones, and these, dried by the sun, are later swept into the central well. Formerly the bones were kept in an ossuary, the *astodān,* to preserve them from rain and animals. The morning of the fourth day is marked by the most solemn observance in the death ritual, for it is then that the departed soul reaches the next world and appears before the deities who are to pass judgment over it.

Festivals. Festivals, in which worship is an essential part, are characteristic aspects of Zoroastrianism, a faith that enjoins on man the pleasant duty of being happy. The principal festivals in the Parsi year are the six seasonal festivals, *Gahānbārs,* and the days in memory of the dead at year's end. Also, each day of the month and each of the 12 months of the year is dedicated to a deity. The day named after the month is the great feast day of that particular deity.

The New Year festival, Nōrūz, is the most joyous and beautiful of Zoroastrian feasts, a spring festival in honour of Rapithwin, the personification of noonday and summer. The festival to Mithra, or Mehragān, was traditionally an autumn one, as honoured as the spring feast of Nōrūz.

Ethics. The precepts of Mazdean ethics focus upon the maintenance of life and the fight against evil. In order to maintain life one must earn one's living by means of cattle raising and agriculture, and one must procreate. To fight against evil is to combat the demons and whatever beings, men or animals, belong to them. The two points of view seem to coincide, considering that the forces of evil are the forces of death: good is opposed to evil as light is to darkness, as life is to nonlife. The life precepts can be transposed into fight precepts; for instance, eating and drinking are interpreted by Zātspram as a struggle against the she-demon Āz, "Concupiscence." The two points of view, however, are also contradictory: how can man fight the forces of evil without suppressing certain lives, such as baleful animals? The second viewpoint prevails: Iran ignores, even in theory, the universal respect of life that is preached by Buddhism or that justifies the vegetarian diet of Brahmanic India.

Social reasons (*e.g.,* the desire to maintain family privileges) apparently explain the development of consanguineous marriage, an acute form of endogamy.

Future life should be determined by the balance of the good and evil deeds, words, and thoughts of the whole life. This principle, however, is tempered to allow for human weakness. All faults do not have to be registered or weighed forever on the scales. There are two means of effacing them: confession and the transfer of supererogatory merits (the equivalent of the Roman Catholic "Treasury of Merits" of Christ and the saints). The latter is the justification for the prayers and ceremonies for the departed.

ICONOGRAPHY

There is no Zoroastrian art. Be it in the Achaemenid, Arsacid, or Sāsānian period, Iranian art was predominantly royal. Only one god is represented during the first period: Auramazda, as a winged disk hovering above the king. It is known, however that Artaxerxes II introduced statues of Anahita into her temples, after the Greek fashion. In the Arsacid period, Greek models also served for the representations of Iranian gods ordered by the kings on reliefs or coins. In the Sāsānian period, deities were represented only in the giving of the royal investiture, as is the case with Ormazd and Anahita at Naqsh-e Rostam, or Ormazd and Mithra at Taq-e Bostan. The frequency of the bullman in Achaemenid and Sāsānid iconography may be due to the obviously royal character of this personage: on seals he wears a crown, and the Pahlavi text calls him Gopatshāh, "King of Gopat."

RELATION TO OTHER RELIGIONS

The debt of Israel to its Eastern neighbours in religious matters is easy to demonstrate on a few precise points of minor importance but less so in other more important points, such as dualism, angelology, and eschatology.

The Judeo-Christian tradition

Isaiah 40–48 offers striking parallels with the *Gāthā* 44:3–5, as has been shown by Morton Smith. Besides the common procedure of rhetorical questions, there is the notion of a god who has created the world and, notably, light and darkness. The very idea of a creator god may be common to all of the western part of the Semitic world. But the notion that God created light and darkness appears in both prophets. It is true that Zoroaster associates light and darkness only to waking and sleep and that no Iranian text says that God created good and evil. Nevertheless, the juxtaposition, in Isaiah, of light–darkness with good–evil sounds remarkably Iranian.

After the exile, the traditional hope in a messiah-king of the House of David who would reestablish Israel as an independent nation and make it triumph over all enemies gave way gradually to a concept at once more universal and more moral. The salvation of Israel was still essential, but it had to come about in the framework of a general renewal; the appearance of a saviour would mean the end of this world and the birth of a new creation; his judgment of Israel would become a general judgment, dividing mankind into good and evil. This new concept, at once universal and ethical, recalls Iran so strongly that many scholars attribute it to the influence of that country. John R. Hinnells has seen this influence especially in the saviour's defeat of the demons, his gathering of men for the judgment scene, his raising of the dead, and his administration of the judgment. The occasion of this influence, according to Hinnells, may be found in the contacts between the Jews and the Parthians that were initiated in the 2nd century BC but that reached a climax in the middle of the 1st century BC.

Influence on Greek thought

Although Pythagoras cannot have been a pupil of Zoroaster, there are striking similarities of doctrine between Iran and Greece. Anaximander's world picture corresponds to that of the Avesta. Heracleitus seems to have been impressed, in Ephesus, by the practices of the Magi, if not by their theory on the fiery nature of the soul. This would account for the emergence, in 5th-century Greece, of the belief in the heavenly fate of the soul.

The search for an Iranian background to Gnosticism must be placed in a new perspective if the recent view that Gnosticism is really a Christian heresy is accepted.

CONCLUSION

Zoroastrianism is not the purely ethical religion it may at first seem. In practice, despite the doctrine of free choice, a Zoroastrian is so constantly involved in a meticulous struggle against the contamination of death and the thousand causes of defilement, and against the threat, even in his sleep, of ever-present demons, that he does not often believe that he is leading his life freely and morally.

Apart from this attitude, the belief in the power of destiny sometimes culminates in fatalism. The latter is easily associated with Zurvanism, itself sometimes tainted with materialism. In the *Mēnōk-i Khrat,* it is stated that "though one be armed with the valour and strength of wisdom and knowledge, yet it is not possible to strive against fate."

On the whole, however, as R.C. Zaehner notes, "the theological premises" of Zoroastrianism "are based on an essentially moralistic view of life."

BIBLIOGRAPHY. A full critical bibliography is found in JACQUES DUCHESNE-GUILLEMIN, *La Religion de l'Iran ancien* (1962); a revised version was published in Bombay in English in 1973. GEO WIDENGREN, *Die Religionen Irans* (1965), is comprehensive. JAMES HASTINGS (ed.), *Encyclopaedia of Religion and Ethics*, 12 vol. (1908–27), contains nearly 100 articles on Zoroastrianism. See also S. INSLER, *The Gāthās of Zarathustra* (1975); JACQUES DUCHESNE-GUILLEMIN, "Zoroastrian Religion," *Cambridge History of Iran*, vol. 3(2), ch. 23 (1983); MARY BOYCE, *A History of Zoroastrianism*, 2 vol. (1975–82), and *Zoroastrians: Their Beliefs and Practices* (1979); J.R. HINNELLS, "Zoroastrian Saviour Imagery and Its Influence on the New Testament," *Numen*, 16:161–185 (Dec. 1969); SIMONE PÉTREMENT, "Sur le problème du gnosticisme," *Revue de Métaphysique et de Morale*, 85:145–177 (April–June 1980); P. LECOQ, "Ahura Mazdā ou Khvarnah," *Acta Iranica*, 23:301–326 (1984).

(J.D.-G.)